Tinkering with carbohydrate-active enzymes for novel specificities: Chitin deacetylases and Glycosidases

Sergi Pascual Torrent

UNIVERSITAT RAMON LLULL

DOCTORAL THESIS

Title	Tinkering with carbohydrate-active enzymes for novel specificities: Chitin deacetylases and Glycosidases.
Presented by	Sergi Pascual Torrent
Centre	IQS School of Engineering
Department	Bioengineering
Directed by	Prof. Dr. Antoni Planas Sauter

C. Claravall, 1-3 | 08022 Barcelona | Tel. 93 602 22 00 | Fax 93 602 22 49 | info@url.edu | www.url.edu

Work supported by the European project “European Union’s Seventh Framework Programe for research, technological development and demonstration” under grant agreement number 613931 (NANO3BIO), grants BFU2016-77427-C2-1-R (GLYCO2ZYMES) and PID2019-104350RB-I00 (GLYCODESIGN) from Ministerio de Ciencia e Innovación, Spain, and grant for Consolidated group GQBB (Química Biològica I Biotecnologia) 2017 SGR 727 from AGAUR, Generalitat de Catalunya.

Sergi Pascual acknowledges a predoctoral contract FPI (BES-2017-080026) associated to the GLYCO2ZYMES project from Ministerio de Ciencia e Innovación, Spain.

Per la meva família i pel padrí

''Life is a symbiotic and cooperative union that allows those who associate to succeed''

Lynn Margulis

ACKNOWLEDGEMENTS

Agraïments

Arriba el final d'una llarga etapa que m'ha permès créixer com a persona i com a científic, per això és l'hora d'agrair a tothom que s'hi ha vist implicat.

En primer lloc agrair al Dr. Antoni Planas per deixar-me entrar a formar part del grup de Bioquímica, per confiar en mi des del primer moment, per donar-me la llibertat de fer-me meva aquesta tesi, per ensenyar-me a pensar de forma crítica, per totes les discussions científiques i personals que m'han enriquit de forma incalculable. Moltes gràcies Toni. També gràcies a tota la família per acollir-me com un més durant la meva estada a Connecticut.

Al Dr. Xevi Biarnés, per ser una part important d'aquesta tesi, per les idees i consells però sobretot pels moments viscuts tant al dia a dia com als congressos que m'han permès conèixer un gran científic però també una gran persona.

A la Dra. Magda Faijes, per l'entusiasme i l'alegria, per sempre ser un punt de recolzament i comprensió al llarg d'aquest camí.

A la Dr. Teresa Pellicer, per les conversacions personals i professionals tant enriquidores i per tots els consells que sempre han sigut de gran ajuda.

També donar les gràcies a tots els professors i investigadors del Laboratori de Bioquímica, al Dr. Pau Leivar, al Dr. Marc Carnicer i la Patri, per la vostra ajuda i aportacions, per interessar-vos sempre i aportar de forma constructiva.

I would like to thank Prof. Mark Peczuh from University of Connecticut for allowing me to stay in his group and especially thanks to Dr. Aditya Pote for all the guidance and help in the synthesis of septanoside compounds.

Agradecer también al Dr. Jesus Jiménez-Barbero, Dr. Óscar Millet y Dr. Ganeko Bernardo por todos los experimentos de RMN realizados así como por todas las discusiones científicas que han sido de gran ayuda.

To all members of the Thesis Committee.I hope you enjoy it. Thanks to be part of it.

Han sigut 5 anys compartint laboratori amb molta gent, gràcies a tots per la vostra aportació en aquesta tesi i per fer el camí inoblidable.

Agradecer primero al Dr. Hugo Aragunde, he tenido el placer de poder aprender de alguien como tú, alguien que fue capaz de guiarme y enseñarme como nadie, fuiste un referente desde el principio. Parte de esta tesis también es tuya.

La primera etapa de la tesi no hauria tingut sentit sense les persones que formaven part del laboratori. Gràcies a les doctores Estela Castilla, Cristina Alsina, Laia Grifoll i Núria Orive, per acompanyar-me durant aquest camí però sobretot per tots els moments viscuts dintre i fora del laboratori. Gràcies a totes per ser com sou, per haver-me deixat entrar a les vostres vides i per fer de bioquímica, una família. No podria haver escollit unes millors companyes per compartir aquesta experiència. Espero que puguem celebrar molts èxits juntes! També agrair a la Mireia, per compartir tot el camí junts, de principi a final, per ser algú que sempre et treu un somriure i et fa veure el costat positiu de les coses.

Gràcies també als doctorands de l'etapa final, al Marc, al Bernat, a l'Aitor i a l'Álvaro, heu fet que el final sigui un plaer, hem compartit molts moments que sense dubte fan que el laboratori de bioquímica sigui un lloc que trobaré a faltar. Aquí teniu el relleu!

Gràcies també a tots els estudiants de grau i de màster que han format part d'aquest projecte. Sense la vostra ajuda aquesta tesi no estaria completa.

Al Pol, més de 10 anys creant teories i portant-les a la pràctica, compartint els millors moments de les nostres vides i creant records que seran eterns, tenim la sort que això és el principi. Gràcies per tots els moments viscuts!

A Juan, por ser alguien a quien realmente admiro, empezamos la carrera juntos y no he parado de aprender de ti. Gracias por ser una fuente de inspiración! Aun tenemos muchos viajes pendientes!

Als de Lleida, al Enric, el Jofre, el Kevin i l'Aleix, simplement per estar sempre al meu costat, per recolzar-me sense demanar res a canviï i per ser la definició d'amistat. Perquè poques coses valen més la pena que una bona estona amb els amics.

Als de Sansa, a l'Alex, a la Laia, a l'Artur i l'Andrea, a la Marta, al Lluis, al David, a la Site, al Salva i a la Esti, crec que els millor moments els he viscut amb vosaltres. Gràcies per escoltar-me i estar al meu costat en tot moment. Créixer al vostre costat és un plaer. Ens queden moltes estones per passar junts!

A la Sandra, per aparèixer en el moment menys pensat i entrar de ple a la meva vida. Per cuidar-me i ajudar-me en aquesta etapa final. Gràcies per les il·lusions i aventures junts! Això és el principi!

A la meva família, els meus pares, per sempre voler entendre de que va la tesis, per totes les converses, pels petons i abraçades, per recolzar-me en totes les decisions que he pres i sobretot per estimar-me. Gràcies per amb el vostre exemple, ensenyar-me que amb esforç i treball, es pot aconseguir tot. Perquè tot el que soc és gràcies a vosaltres. Us estimo!

INDEX

1. INDEX

2. SUMMARY ..... - 23 -

3. LIST OF PUBLICATIONS, COMMUNICATIONS AND COLLABORATIONS ..... - 27 -

4. LIST OF FIGURES ..... - 33 -

5. LIST OF TABLES ..... - 45 -

6. LIST OF EQUATIONS ..... - 49 -

7. ABBREVIATIONS ..... - 53 -

8. THESIS OUTLINE ..... - 59 -

9. TOPIC 1: Engineering substrate specificity of chitin deacetylases. INTRODUCTION ..... - 65 -

9.1. Biological implications and applications of chitin and chitosan oligosaccharides. ..... - 65 -

9.1.1. Structural properties of chitin and chitosan. ..... - 65 -

9.1.2. Chitin and chitosan: Biological applications and implications. ..... - 68 -

9.1.3. Production of chitosan and COS. ..... - 72 -

9.2. Nano3bio project and application for paCOS obtention. ..... - 77 -

9.3. CAZy classification, Carbohydrate Esterase family 4 (CE4). ..... - 80 -

9.3.1. Carbohydrate esterase family, de-*N*-acetylation mechanisms and structures. ..... - 81 -

9.3.2. Deacetylases acting on structural polysaccharides: chitin de-*N*-acetylases. ..... - 83 -

9.3.3. Chitin de-*N*-acetylase from *Vibrio cholerae* ..... - 89 -

9.3.4. Determinants of substrate specificity: Subsite Capping Model. ..... - 96 -

9.4. Protein engineering: Rational design and Directed evolution ..... - 101 -

9.4.1. Rational design of enzymes ..... - 101 -

9.4.2. Directed evolution of enzymes ..... - 102 -

10. OBJECTIVES ..... - 113 -

11. CHAPTER 1. Biochemical Characterization of *Vc*CDA ..... - 117 -

11.1. Expression and purification of VcCDA FL wt ..... - 117 -

11.2. Kinetic characterization of the VcCDA FL ..... - 122 -

11.2.1. Reported methods for analysis of deacetylase activity ..... - 122 -
11.2.2. Development and validation of an HPLC-MS method with high capacity for kinetic studies - 123 -

11.3. Protein thermostability characterization ..... - 129 -
11.3.1. Temperature profile of enzyme ..... - 129 -
11.3.2. Melting temperature ..... - 130 -
11.3.3. Thermotolerance ..... - 132 -

12. CHAPTER 2. Screening assay for directed evolution of chitin deacetylases ..... - 137 -

12.1. Introduction ..... - 137 -

12.2. Development of High Throughput Screening (HTS) assay. ..... - 138 -
12.2.1. Evaluation of protein binding to Chitin Magnetic Beads (CMB). ..... - 139 -
12.2.2. Specificity of *Vc*CDA binding to the CMBs. ..... - 142 -
12.2.3. Deacetylase activity of protein bound to CMB. ..... - 144 -
12.2.4. Evaluation of *E.coli* expression hosts. ..... - 146 -
12.2.5. Evaluation of cell lysis method. ..... - 147 -
12.2.6. Development and Validation of final HTS assay. ..... - 149 -
12.2.7. Application to *Vibrio cholera* deacetylase mutant libraries for engineered specificity. ..... - 153 -

13. CHAPTER 3. Engineering loop flexibility to tune enzyme specificity in *Vc*CDA. ..... - 173 -

13.1. Introduction and Project statement. ..... - 173 -

13.2. Loop 5 blocking by a disulfide bond ..... - 176 -
13.2.1. Expression, purification of *Vc*CDA FL P193C / N273C. ..... - 178 -
13.2.2. Evaluation of intramolecular C193-C273 disulfide bridge. ..... - 179 -
13.2.3. Kinetic characterization of *Vc*CDA FL P193C/N273C ..... - 181 -

13.3. Flexibilization of Loop 5: Proline to Glycine variants. ..... - 183 -
13.3.1. Introduction: Loop 5 flexibilization ..... - 183 -
13.3.2. P-G mutant deconvolution: Evaluation of individual and double mutations. ..... - 187 -

13.4. Protein dynamics by NMR. ..... - 199 -
13.4.1. Introduction: NMR techniques for protein dynamics. ..... - 199 -
13.4.2. Protein expression and purification in minimal media for isotopic labelling. ..... - 199 -
13.4.3. Elucidation of loop 5 dynamics: *Vc*CDA CD wt and *Vc*CDA CD PG ..... - 205 -

13.5. Molecular dynamics simulations of VcCDA, targeting loops dynamics. ..... - 219 -

13.6. Closing remarks of VcCDA enzyme loop dynamics ..... - 222 -

14. CHAPTER 4. Engineering *Vc*CDA enzyme towards new deacetylation patterns. ..... - 227 -

14.1. Introduction. Subsite Capping Model and Pattern of deacetylation. ..... - 227 -

14.2. Development and validation of a rapid screening assay for deacetylase activity. ..... - 237 -

14.3. Library screening of VcCDA FL D74X-E75X-N76X. ..... - 241 -

14.4. Closing remarks in engineering VcCDA towards new deacetylation patterns ..... - 251 -

15. TOPIC 2: Engineering a septanoside hydrolase. INTRODUCTION. Glycosidases and glycosynthases. ..... - 257 -

15.1.1. Catalytic mechanism of glycosynthases ..... - 259 -

15.2. β-glycosidases ..... - 261 -

15.2.1. β-glucosidase 3 from *Streptomyces sp.* ..... - 261 -

15.3. Seven-member rings ..... - 263 -

16. OBJECTIVES ..... - 267 -

17. CHAPTER 5. Indolyl Septanoside Synthesis for *In Vivo* Screening of Bacterial Septanoside Hydrolases ..... - 271 -

18. CHAPTER 6. Towards engineering a β-glucosidase to accept seven-member ring septanoside substrates ..... - 295 -

18.1.1. Alanine scanning mutagenesis ..... - 296 -

18.1.2. Kinetic parameters a for wt and W430A mutant ..... - 299 -

18.1.3. In silico analysis of W430A mutant ..... - 300 -

18.2. Closing remarks towards engineering a β-glucosidase to accept seven-member ring septanoside substrates ..... - 306 -

19. CONCLUSIONS ..... - 311 -

19.1. Conclusions from engineering substrate specificity of Chitin deacetylases (Chapters 1 to 4) - 311 -

19.2. Conclusions from β-glucosidase from Streptomyces sp. ..... - 313 -

21. MATERIALS AND METHODS ..... - 317 -

21.1. Development of HTS assay for Chitin de-N-Acetylases. ..... - 317 -
21.1.1. Binding efficiency of *Vc*CDA to the CMBs using different conditioning and binding buffers. - 317 -
21.1.2. Specific activity of CMB-immobilized *Vc*CDA ..... - 318 -
21.1.3. Specificity of *Vc*CDA binding to the CMBs. ..... - 318 -
21.1.4. Specificity of *Vc*CDA binding to the CMBs. ..... - 318 -
21.1.5. Kinetics of *Vc*CDA binding to the CMBs. ..... - 319 -
21.1.6. Deacetylase activity of free and CMB-immobilized *Vc*CDA using DP2 and DP4 as substrates. ..... - 319 -
21.1.7. *E. coli* expression hosts. ..... - 320 -
21.1.8. Cell lysis methods on microplates. ..... - 321 -
21.1.9. HTS protocol of mutant libraries ..... - 322 -
21.1.10. Reproducibility of the HTS assay. ..... - 323 -
21.1.11. DNA libraries by error-prone PCR. ..... - 324 -
21.1.12. Temperature profile of CMB-immobilized wt *Vc*CDA. ..... - 324 -
21.1.13. Thermal stability of wt and mutant VcCDA proteins ..... - 325 -
21.1.14. Thermotolerance of free and immobilized enzymes. ..... - 326 -

21.2. Screening assay to evaluate the deacetylation pattern ..... - 327 -
21.2.1. Rapid screening assay for large libraries of chitin deacetylases ..... - 327 -
21.2.2. Secondary screening assay to evaluate the deacetylation pattern ..... - 328 -

21.3. DNA manipulation ..... - 329 -
21.3.1. *Vc*CDA individual and double mutants ..... - 329 -
21.3.2. Mutagenesis analysis for NMR experiments ..... - 330 -
21.3.3. Alanine scanning mutagenesis for βgl-3 ..... - 331 -

21.4. Expression and purification of proteins ..... - 332 -
21.4.1. Protein expression with autoinduction media ..... - 332 -
21.4.2. Protein expression with IPTG induction ..... - 332 -
21.4.3. Protein expression in minimal media ..... - 333 -
21.4.4. Strep-trap affinity chromatography ..... - 335 -
21.4.5. His-tag column affinity ..... - 336 -
21.4.6. Size Exclusion Chromatography ..... - 336 -

21.5. Kinetics characterization of VcCDA enzymes ..... - 339 -
21.5.1. Deacetylase activity determination by HPLC-MS ..... - 339 -
21.5.2. Specific activity of enzymes ..... - 339 -

21.5.3. HPLC-MS methodology for the estimation of monodeacetylated products: Preparation of standards. ..... - 339 -
21.5.4. Michaelis-Menten reaction ..... - 340 -
21.6. Kinetic characterization of β-glucosidase enzymes ..... - 341 -
21.6.1. Specific velocity assay ..... - 341 -
21.6.2. Michaelis-Menten assay ..... - 341 -
21.7. Modelling of βgl3 enzyme in complex with substrates ..... - 342 -
21.8. NMR methodology ..... - 343 -
21.9. Basic Biochemistry protocols ..... - 345 -
21.9.1. SDS-PAGE electrophoresis ..... - 345 -
21.9.2. Protein quantification by BCA ..... - 346 -
21.10. Basic molecular biology protocols ..... - 347 -
21.10.1. Chemical competent cells preparation and transformation ..... - 347 -
21.10.2. Electrochemical competent cells preparation and transformation ..... - 348 -
21.10.3. DNA obtention and purification ..... - 349 -
21.10.4. DNA quantification ..... - 349 -
21.10.5. DNA electrophoresis gel ..... - 349 -
22. BIBLIOGRAPHY ..... - 353 -

SUMMARY

2. SUMMARY

This thesis is divided in two distinct topics based on the enzymes studied: chitin deacetylases and glycosidases. In both cases the goal is to engineer substrate specificity to design novel biocatalysts.

Chitosans are a family of polysaccharides of N-acetyl-glucosamine and glucosamine. Chitosan oligosaccharides have proven to be bioactive molecules in a broad variety of applications in the fields of biotechnology and biomedicine. One way to overcome the problems of chemical deacetylation of chitin is using the high selectivity in terms of substrate preference and product yields of the enzymatic procedures. The solution of the 3D structure of the *Vibrio cholera* Chitin deacetylase in complex with different substrates was a milestone in the establishment of the Subsite Capping Model to rationalize chitin deacetylases specific deacetylation patterns.

The main objective of the first thesis topic comprises the evaluation of structure-function relationships aiming to support the Subsite Capping Model hypothesis and modify the substrate specificity of the Vibrio cholera Chitin deacetylase towards novel biocatalysts. After designing a high throughput screening (HTS) assay for deacetylase activity on chitooligosaccharides, we prepared a series of mutants involving loop 5 at the positive subsites of the binding cleft based on rational design and directed evolution approaches. Structural and kinetic experiments of designed mutants concluded that modifying the dynamics of enzyme loops affects not only the substrate specificity towards new specificities, but also the binding mechanism, switching the conformational selection mechanism of the wt enzyme to an induced fit mechanism in a triple Proline to Glycine mutant that allowed us to elucidate the molecular determinants that define substate specificity.

The second thesis topic on glycosidase engineering was a test bed to investigate and influence enzyme·ligand interactions with a different class of potential substrates, seven-member ring septanosides, unnatural druggable glycomimetics that are pyranose homologues. The aim is to engineer a *Streptomyces sp.* beta-glucosidase to accept septanoside substrates that may become a novel biorthogonal system with applications in chemical biology. In this work, the synthesis of chromogenic septanosides as potential glycosidase substrates was achieved for screening a panel of glycosidases. A β-glucosidase from *Streptomyces* was identified as a potential candidate for engineering septanoside

hydrolylase activity. Preliminary mutagenesis studies identified a number of active site residues as candidates for the preparation of combinatorial libraries to engineer septanoside specificity. Likewise, docking experiments established a premise for new septanoside configurations to increase the panel of substrates available for further directed evolution strategies.

LIST OF PUBLICATIONS, COMMUNICATIONS AND COLLABORATIONS

3. LIST OF PUBLICATIONS, COMMUNICATIONS AND COLLABORATIONS

Publications

1. Chitin deacetylases: structures, specificities, and biotech applications.

L. Grifoll-Romero, S. Pascual, H. Aragunde, X. Biarnés, A. Planas.

***Polymers* 10, 352 (2018).** DOI: https://doi.org/10.3390/polym10040352

2. Screening Assay for Directed Evolution of Chitin Deacetylases: Application to Vibrio cholerae Deacetylase Mutant Libraries for Engineered Specificity.

Sergi Pascual and Antoni Planas

***Anal. Chem.* 2018, 90, 18, 10654–10658,** https://doi.org/10.1021/acs.analchem.8b02729

3. Carbohydrate de-N-acetylases acting on structural polysaccharides and glycoconjugates.

Sergi Pascual and Antoni Planas

***Curr Opin Chem Biol.* 2021 Apr;61:9-18.** https://doi.org/10.1016/j.cbpa.2020.09.003

4. Indolyl Septanoside Synthesis for *In Vivo* Screening of Bacterial Septanoside Hydrolases.

Aditya R. Pote, Sergi Pascual, Antoni Planas, and Mark W. Peczuh

***Int. J. Mol. Sci.* 2021, 22(9), 4497;** https://doi.org/10.3390/ijms22094497

In preparation

5. An expeditious HPLC-MS method for kinetic characterization of chitooligosaccharide deacetylases

S. Pascual, H. Aragunde, and A. Planas

6. Modulation of loop dynamics alters enzyme specificity and switches substrate binding from conformational selection to induced fit mechanisms

S. Pascual, H.Aragunde, G.Bernardo, J.Jiménez-Barbero, X.Biarnés, O.Millet, A.Planas

Communications on scientific conferences and meetings

1. Poster communication, presented on "19th European Carbohydrate Symposium EUROCARB" (Barcelona, July 3, 2017).

Screening methodology for the evaluation of chitin deacetylase libraries using a medium throughput robotic platform

S. Pascual, H. Aragunde, A. Planas.

2. Oral communication, presented on "NANO3BIO Project Final Dissemination Event" (Hyderabad) (India), September 20, 2017).

Enzyme engineering for the production of chitosans with defined deacetylation pattern.

A. Planas, X. Biarnés, H. Aragunde, C. Alsina, L. Grifoll-Romero, S. Pascual, A. Aranda-Martínez.

3. Poster communication, presented on "NANO3BIO Project Final Dissemination Event" (Hyderabad (India), September 21, 2017).

Screening methodology for the evaluation of chitin deacetylase libraries using a medium throughput robotic platform

S. Pascual, H. Aragunde, A. Planas.

4. Poster communication, presented on "2nd New England Glycochemistry Meeting" (Northeastern University, Boston, USA, June 9, 2018).

Screening assay for directed evolution of chitin deacetylases. Application to Vibrio cholera deacetylase mutant libraries for engineered specificity.

S. Pascual, H. Aragunde, A. Planas.

5. Poster communication, presented on "29th International Carbohydrate Symposium" (Lisbon, July 18, 2018).

Screening methodology for the evaluation of chitin deacetylase libraries using a medium throughput robotic platform

S. Pascual, H. Aragunde, A. Planas.

6. Poster communication, presented on "International workshop on biomedical glycosciences" (San Sebastian, June 4, 2019).

Directed evolution of chitin deacetylases

S. Pascual, A. Planas.

7. Oral communication, presented on "International workshop on biomedical glycosciences" (San Sebastian, June 4, 2019).

Directed evolution of chitin deacetylases

S. Pascual, A. Planas.

8. Poster communication, presented on "11a trobada de joves investigadors dels països catalans" (Sitges, January 29, 2020).

Directed evolution of chitin deacetylases

S. Pascual, A. Planas.

Collaborations

Parts of this work has been done in collaboration with:

Dr. Oscar Millet, Dr. Ganeko Bernardo and Dr. Jesús Jiménez Barbero, CICbioGUNE, Bilbao, for Relaxation Dispersion NMR studies of Vibrio cholerae chitin deacetylase. Chapter 3, NMR studies. NMR experiments were performed by Dr. Ganeko Bernardo at CICBioGUNE.

Dr. Mark Peczuh and Dr. Aditya Pote (former PhD student), Department of Chemistry at University of Connecticut (UConn), Storrs, CT, USA, for synthesis of septanoside substrates, where I did a research stay March 1st to May 15th (experimental) plus two month virtual work (computational), Chapter 4.

LIST OF FIGURES

4. LIST OF FIGURES

Figure 9.1. Scheme of the key factors that describe chitosan oligomers: degree of polymerization (DP), degree of acetylation (DA) and pattern of acetylation (PA). - 67 -

Figure 9.2. Physico-chemical and biological properties of chitosan or COS. - 67 -

Figure 9.3. Extraction and preparation of chitin and chitosan from natural sources. Adapted from [13]. - 74 -

Figure 9.4. Scheme of the biotechnological production of paCOS in Nano3Bio project. - 78 -

Figure 9.5. Scheme of the production routes of all possible chitin and chitosan tetramers using 4 different CDAs to specifically deacetylate or *N*-acetylate COS. A: GlcNAc; D: GlcNH2. Blue arrows: deacetylation reactions; red arrows: *N*-acetylation reactions in the presence of excess acetate. Adapted from [85]. - 79 -

Figure 9.6. Carbohydrate de-*N*-acetylases showing the substrate specificities of each CE family. NH_2- in red is the deacetylated–NAc moiety in the corresponding product. Adapted from [96] - 82 -

Figure 9.7. Catalytic mechanism of carbohydrate de-*N*-acetylases. (A) Metal-assisted acid/base mechanism of carbohydrate de-N-acetylases. Conserved Asp-His-His metal-binding triad, general acid (conserved His) and general base (conserved Asp). (B) Metal binding sites in de-N- acetylases, B.1) mononuclear metal coordination (VcCDA, PDB 4OUI, mutant D39S at the general base) and B.2) binuclear metal coordination (MsNagA, PDB 6FV3) found in some CE9 enzymes. - 83 -

Figure 9.8. Conserved motifs MT1–MT5 defining the active site of CE4 enzymes. Motif 1 (TFDD) includes the general base aspartate (first D) and the metal-binding aspartate (second D) that, together with two His of motif 2 (H(S/T)xxH), form the characteristic Asp-His-His 3D structures of representative and novel chitin deacetylases (CDAs). - 84 -

Figure 9.9. Multiple sequence alignment of CE4 enzymes. Loops are highlighted with coloured boxes according to [94]. Conserved catalytic motifs are labelled MT1-5. “His-His-Asp” metal binding triad (▼), catalytic base (❋), and catalytic acid (⧫) are labelled. The alignment includes sequences of the enzymes, except for BmCDA1 and BmCDA8 from the insect Bombyx more, which were added later to the list and show some differential structural features to other CE4 members. - 85 -

Figure 9.10. Modes of enzymatic action patterns for CE4 enzymes. A) Multiple-attack or processive mechanism; B) Multiple-chain or distributive mechanism and C) single-site mechanism. - 87 -

Figure 9.11. Different X-ray structures of carbohydrates de-*N*-acetylases and deacetylation patterns. (a) X-ray structures of Vibrio cholera CDA (*Vc*CDA, long loops, deep and narrow cleft) and Arthrobacter CDA (ArCE4, short loops, open and shallow cleft) in complex with chitooligosaccharide substrates. Loops 1 to 6 (color-labelled) illustrate their function according to the “subsite capping model”. (b) Reported CDAs with known deacetylation pattern on chitooligosaccharides. PA: pattern of acetylation of end products from COS. A: GlcNAc; D: GlcN. In bold, characterized enzymes since 2018. (c) Relevant new 3D structures of CE enzymes since 2018 (left) insect CDA, loops 1 to 6 characteristic of CE4 enzymes labelled as in A), new loop insertion labelled in magenta, (middle) BsPdaC has dual MurNAc DA activity on peptidoglycan and GlcNAc DA activity on chitooligosaccharides (right), AfAgd3, GalNAc exopolysaccharide deacetylase, N-terminus domain (yellow) CBM87, C-terminus (pale green), founding member of family CE18. - 88 -

Figure 9.12. Structure of chitin de-*N*-acetylases from *vibrio cholerae*. (a) Modules present within the *Vc*CDA enzyme sequence. From the N end to the C terminal is the CE4 domain (26-338) and 2 CBMs of family 5 (336-382 and 385-431). (b) Schematic representation showing the general crystallized structure of *Vc*CDA as a homodimer. One monomer is represented in orange while the other domains that compose it are coloured in the other monomer: the deacetylase domain in red and yellow and the CBMs in blue and green. (c) Topological diagram of the *Vc*CDA monomer. The upper part shows the deacetylase domain while the lower part corresponds to the two CBMs. - 91 -

Figure 9.13. Loops of *Vc*CDA. Representation of the loops that decorate the periphery of the active site of *Vc*CDA. - 92 -

Figure 9.14. Chromatogram resulting from the preparative GFC of *Vc*CDA expressed in E. coli after being purified by affinity chromatography (Strep-Trap). - Absorbance at 280nm (···) enzymatic activity using AcOMU as substrate. Upper left graph: molecular weight of each of the fractions estimated by HPSEC; Upper right graph: MALDI-TOF MS analysis of fraction 5 (monomeric). [94] - 93 -

Figure 9.15. Determination of the optimal pH of *Vc*CDA. Dependence of the enzymatic activity of VcCDA as a function of pH using chitobiose (DP2 4 mM). ..... - 93 -

Figure 9.16. Characterization of the kinetic parameters using the monomeric fraction of *Vc*CDA FL and the chitooligosaccharides DP2 to DP4 in 50 mM phosphate buffer pH 8.5 NaCl 300 mM and at a temperature of 37ºC. ..... - 94 -

Figure 9.17. Crystallographic structures of VcCDA in the unligated form, in complex with $(GlcNAc)_2$ and in complex with (GlcNAc)3 [94]. A) Superimposition of the three structures. Left: loop 4 (orange) has different conformations; right: magnification of the active site with loop 4 in the unligated form (blue), and in enzyme-substrate complexes with $(GlcNAc)_2$ (yellow) and $(GlcNAc)_3$ (red) ligands. Only DP2 ligand is shown. B) Overall structures. *Vc*CDA:unligated, *Vc*CDA-DP2:in complex with DP2, *Vc*CDA-DP3:in complex with DP3. Loops 1 to 6 are coloured as in Figure 9.13. ..... - 95 -

Figure 9.18. CDAs and CODs with known pattern of deacetylation on chitooligoshacarides. ..... - 97 -

Figure 9.19. 3D structures of enzyme·substrate complex. (a) VcCDA with $(GlcNAc)_3$ substrate. (b) ArCE4 with $(GlcNAc)_2$ substrate. Loops 1 to 6 are colored as in Figure 9.13. ..... - 99 -

Figure 9.20. General steps of a directed evolution protocol. The starting gene encoding the protein that will be evolved is subjected to a gene diversification strategy; the quality and size of the resulting library will depend on the criteria chosen for its construction. The individual variants of the genetic library are expressed in an appropriate host or by cell-free *in vitro* transcription and translation (IVTT). Using an appropriate assay, a screening will then be implemented to identify one or more variants of interest, typically showing a desired activity, thermostability, or selectivity. These variants can be subjected to a new round of directed evolution, until the improvements are satisfactory enough to enable the biocatalytic process. ..... - 103 -

Figure 9.21. Relationship between catalytic promiscuity and the evolution of a new function. The aim of a DE experiment is to traverse a fitness landscape towards a maximum; ideally the global maximum. Here, the library starting point is represented by the red ball, intermediates by grey balls and local and global maximums represented by amber and green balls respectively. Adapted from [148]. ..... - 105 -

Figure 9.22. Three commonly used library generation techniques are illustrated. (a). An epPCR schematic is shown where a gene of interest is amplified by a low-fidelity polymerase producing errors in amplification (coloured circles) in the final product. (b). A protein region of interest is designated, in cyan, for optimisation via saturation mutagenesis. (c). Two phenotypically equivalent genes are identified for a gene shuffling experiment. This involves restriction and ligation to create a library random homologous genes. ..... - 107 -

Figure 11.1. Strep-Tag II affinity purification process. Outline of the Strep-Trap affinity chromatography protocol: 2. The protein with Strep-Tag II is retained by its affinity with Strep-tactin, 3. *E. coli* proteins are washed with buffer, 4. Desthiobiotin displaces recombinant enzyme and eluted protein of interest, 4. HABA displaces desthiobiotin to regenerate column, 5. HABA is washed away and column regenerated. ..... - 119 -

Figure 11.2. Strep-Trap purification chromatogram of *Vc*CDA FL Chromatogram resulting from affinity chromatography purification of *Vc*CDA in which three column loads were performed. The numbers indicate the stages of a charge / elution cycle: 1. Charge, 2. Wash, 3. Elution, 4. Wash, 5. Column Regeneration. ..... - 120 -

Figure 11.3. SEC chromatogram of *Vc*CDA FL Chromatogram obtained in filtration gel chromatography using a Superdex 200 column (16/600). Five multimeric fractions are shown. F5 corresponds to the monomer. ..... - 121 -

Figure 11.4. 14% SDS-PAGE gel of the entire purification process of *Vc*CDA FL. M: Marker Low Range, P: Pellet from the centrifugation after the lysis, C: crude, soluble fraction of the lysate, FT: Flow through the column, ST: Protein purified by the Strep-Trap, F1-F5: SEC fractions 1 to 5. ..... - 121 -

Figure 11.5. Structures of chitin oligosaccharide DP2 and DP4 and their respective chitosan homologs after being monodeacetylated by the *Vc*CDA. This enzyme deacetylates the second position from the non-reducing end. ..... - 125 -

Figure 11.6. A) Ion scan spectra for the DP2(N) products after deacetylated by *Vc*CDA and purified. m/z 383.3=$[DP2(N)+H]^+$; m/z 405.4= $[DP2(N)+Na]^+$; m/z 765.4= $[2{\cdot}DP2(N)+H]^+$; m/z 787.4= $[2{\cdot}DP2(N)+Na]^+$. B) Ion scan spectra for the DP4(N) products after deacetylated by VcCDA and purified. m/z 789.4=$[DP4(N)+H]^+$; m/z 806.4= $[DP4(N)+Na]^+$. ..... - 126 -

Figure 11.7. *Vc*CDA reaction progress curves at pH 8.5 and 37ºC of the chitobiose and chitotetraose substrate. Concentrations are calculated using a response factor for each compound: DP2 or DP4 as substrate and DP2(N) or DP4(N) as product. ..... - 127 -

Figure 11.8. Michaelis-Menten curves for the purified monomeric *Vc*CDA deacetylation reaction at substrate concentrations ranging from 4 mM to 0.025 mM in Tris 50 mM pH 8.5 and NaCl 300 mM. Temperature was kept constant at 37ºC and a final volume of 100 µl. Substrates used were DP2 and DP4. ..... - 128 -

Figure 11.9. The effect of temperature on the specific activity of *Vc*CDA FL wt. Reaction conditions: Buffer: 50 mM phosphate buffer pH 8.5, 300 mM NaCl, 37 °C. ..... - 130 -

Figure 11.10. Mechanism of action and calculation of Tm using Sypro Orange in a TSA test. Scheme of operation of the Sypro Orange reagent used in a TSA protocol in which the increase in its fluorescence is monitored by binding to the hydrophobic areas of the proteins exposed to the solvent in the denaturation. Example of the crude result obtained in a TSA experiment in which, once a maximum point is reached, the fluorescence is reduced by the formation of protein aggregates. Adapted from [184]. ..... - 132 -

Figure 11.11. Thermal stability of *Vc*CDA wt in presence or absence of Zn. Green: Tm of VcCDA FL wt without Zn^{2+}: 56.0 ± 0.17, Blue: Tm of VcCDA FL wt with Zn^{2+} (0,1 mM $ZnCl_2$): 58.2 ± 0.2. ..... - 132 -

Figure 11.12. The thermotolerance of *Vc*CDA FL WT was determined with free enzymes at 50°C. Exponential decay equation has been adjusted (Equation 21.1). ..... - 133 -

Figure 12.1. Principle of the HTS assay for deacetylase activity on COS. ..... - 138 -

Figure 12.2. Percentage of bound protein to 25 µL CMB suspension at increasing purified protein loads. The binding assay was performed at 4°C, 650 rpm for 1 hour using different buffers. Buffers: A (red column). PBS (50 mM phosphate, 300 mM NaCl, pH 8.5) as conditioning and binding buffer. B (green column). PBS+Triton X-100 (0.05%) as conditioning buffer and PBS as binding buffer. C (blue column). CBB (20 mM Tris·HCl, 500 mM NaCl, 1 mM EDTA, 0.05% Triton X-100, pH 8.0) as conditioning buffer, and PBS as binding buffer. D (black column). CBB as both conditioning and binding buffer. ..... - 140 -

Figure 12.3. Binding efficiency of VcCDA to the CMBs using buffer A: PBS (50 mM phosphate, 300 mM NaCl, pH 8.5) as conditioning and binding buffer. M: molecular weight marker. ..... - 141 -

Figure 12.4. Binding efficiency of *Vc*CDA to the CMBs using different conditioning and binding buffers. Deacetylase activity (60 µg *Vc*CDA incubated with 25 µL CMB suspension with different buffers as conditioning and binding step at 4°C for 1 h) was determined using tetraacetylchitotetraose (DP4) as a substrate (2 mM). Buffers: A (red column). PBS (50 mM phosphate, 300 mM NaCl, pH 8.5) as conditioning and binding buffer. B (green column). PBS+Triton X-100 (0.05%) as conditioning buffer and PBS as binding buffer. C (blue column). CBB (20 mM Tris·HCl, 500 mM NaCl, 1 mM EDTA, 0.05% Triton X-100, pH 8.0) as conditioning buffer, and PBS as binding buffer. D (black column). CBB as both conditioning and binding buffer. SA= Specific activity. ..... - 141 -

Figure 12.5. Specificity of *Vc*CDA binding to the CMBs. Upper gel: CBM-immobilized proteins, Lower gel: supernatants. ..... - 142 -

Figure 12.6. Specificity of *Vc*CDA binding to the CMBs in presence of BSA. M: molecular weight marker; C: no BSA nor VcCDA. In BSA lane, 10 µg of BSA without *Vc*CDA were added. ..... - 143 -

Figure 12.7. Kinetics of *Vc*CDA binding to the CMBs. A) SDS-PAGE of protein-bound and unbound to CMB after different incubation times; B) quantification of bound protein (by quantification of the unbound protein in the supernatant (S) by the BCA assay [200]). ..... - 144 -

Figure 12.8. Deacetylase activity of free and CMB-immobilized *Vc*CDA using DP2 and DP4 as substrates. A) Activity with DP2 substrate monitored by HPLC-MS (protocol). A1) Purified enzyme: Free enzyme (grey circles), CMB-immobilized enzyme (black squares). Inset graph shows specific activity. A2) Cell-free extracts. 150 µl (black cicles) of a cell-free extract containing *Vc*CDA. B) Activity with tetraacetylchitotetraose (DP4) substrate monitored by HPLC-MS. B1) Purified enzyme. 30 µg of purified protein in free form (grey circles) and 60 µg of protein bound to CMB (black squares). B2) Cell-free extract. Deacetylase activity using DP4 was performed using the same protocol described above (A2) working with 100 µl of cell free extract. ..... - 145 -

Figure 12.9. *E.coli expression hos*ts. A) Protein expression: 0-150 µl of cell-free extract from each strain were mixed with CMB in PBS pH 8.5 buffer, and the suspensions were incubated at 4°C for 1 h with shaking (650 rpm). After the binding step, the beads were retained with a magnet and the supernatants transferred to new tubes (S) analyzed by SDS-PAGE. B) Activity of the immobilized enzymes: The beads (CBM-immobilized proteins) coming from 75 µl of cell-free extract from both strains as above were kept at 37°C for 5 min. Reactions were initiated by the addition of 100 µl DP4 (2 mM). After 60 min, the beads were retained with a magnet, and 10 µl of supernatant were withdrawn and added to 90 µl of water/1-propanol (1:1) to stop the reaction and quantify product formation by HPLC-MS. ..... - 147 -

Figure 12.10. Study of cell lysis method. a) Bugbuster Reagent®: the cell pellet was resuspended with 250 µl of Bugbuster reagent with shaking at 650 rpm for 20 min at rt. The lysate was centrifuged at 3500 rpm for 1 h and the pellet discarded. b) Freeze-thaw cycles: the cell pellet was frozen by submerging the microplate in a dry-ice/ethanol bath for 2 min. Then it was thawed by transfer to an ice/water bath for 8 minutes. The cycle was repeated 10 times. The lysate were resupended in 250 µL PBS pH 8.5, centrifuged at 3500 rpm for 1 h, and the pellet discarded. c) Sonication: the cell pellet was resuspended in 250 µL PBS pH 8.5 and lysed by sonication during 7 minutes (50% amplitude, 5 s ON, 15 s OFF, 2 mm diameter probe). The lysate was centrifuged at 3500 rpm for 1 h and the pellet discarded. For the three method, the protein in the clarified cell-free extracts was quantified using the BCA assay [200]. All the procedure was done in triplicate. ..... - 148 -

Figure 12.11. Reaction between fluorescamine and primary amine. ..... - 150 -

Figure 12.12. Deacetylase activity by the fluorescamine assay using the HTS protocol. A) Kinetics of deacetylated product (DP4(N)) formation). B) DP4(N) formation at 60 min reaction, varying the cell-free extract (5-75 µL) at fixed CMB (10 µL). C) DP4(N) formation at 60 min reaction, varying CMBs (1-15 µL) at fixed cell-free extract (25 µL). D) Z'-factor calculation. DP4(N) formation at 60 min reaction using final conditions (10 µL CMB, 25 µL cell-free extract) for 75 repetitions of wt and inactive D39S mutant. Solid lines, average value; dotted lines, ± 3 SD (standard deviation). All deacetylation reactions were done at 37ºC, PBS pH 8.5. ..... - 150 -

Figure 12.13. HTS protocol of mutant libraries. ..... - 152 -

Figure 12.14. Reproducibility of HTS assay applied in 6 cell cultures. ..... - 153 -

Figure 12.15. Plasmid map of VcCDA FL. Map of the original vector in which the complete VcCDA proteins with the Strep-TagII sequence at the C-terminal end were cloned. Library of mutants were clone in a vector containing the CBM. ..... - 154 -

Figure 12.16. CPEC technique. Linearized vector and inserts (with overlapping sequences) are added to a single reaction tube where PCR occurs with the help of a high-fidelity polymerase. The vector and insert use each other as templates to give cloned products. Adapted from [202]. ..... - 154 -

Figure 12.17. Linker between CBMs and catalytic domain of *Vc*CDA. Scheme of the WT FL structure in which the T336 cleavage site is indicated. ..... - 155 -

Figure 12.18. Temperature profile of CMB-immobilized wt *Vc*CDA. ..... - 156 -

Figure 12.19. Random libraries by epPCR. A) First generation, template: wt enzyme, screening for activity at 50°C. G1= *Vc*CDA K275E mutant B) Second generation, template: G1 mutant, screening for activity at 60°C. G2= VcCDA K275E-H127R. ..... - 157 -

Figure 12.20. Chromatogram of Strep-trap purification of *Vc*CDA K275E mutant. A) loading and washing, B) elution; C) regeneration with HABA and washing. mAU: 280nm. ..... - 158 -

Figure 12.21. Chromatogram of SEC purification of *Vc*CDA K275E. Chromatogram obtained in the gel filtration chromatography using a Superdex 200 (16/600). Five multimeric fractions are shown, from which fraction 4 (F4) corresponds to the dimer and fraction 5 (F5) corresponds to the monomer of the protein. ..... - 159 -

Figure 12.22. SDS-PAGE 14% acrylamide of *Vc*CDA FL K27E purification. M: Marker Low Range (Biorad); P: cell lysate, S: soluble fraction; F: flow through during column loading and washing; St: elution of protein purified by Strep-trap column; F1: fraction 1, F2: fraction 2, F3: fraction 3, F4: Fraction 4, F5: fraction 5 from GFC. ..... - 160 -

Figure 12.23. Determination of Michaelis-Menten parameters. Kinetics were performed with 3 different substrates, DP2, DP4, and DP5, according to Protocol 4. A) *Vc*CDA WT; B) K275E mutant; C) K275E-H127R mutant. Conditions: 100 µM to 4 mM substrate, 0.05-0.8 µM enzyme, 50 mM phosphate buffer pH 8.5, 300 mM NaCl, 37°C. Initial rates vs. substrate concentration were fitted by non-linear regression to a cooperative Hill model for the DP2 substrate ($v/[E] = k_{cat}\cdot[S]^h / (K_M^h + [S]^h)$), and to a Michaelis-Menten model for DP4 and DP5 substrates. Hill slope value was 1.6 for all the enzymes with DP2 substrate. Data were fitted using Graphpad software (Prism). Kinetic parameters are given in Table 12.1. ..... - 161 -

Figure 12.24. k_{cat} values of wt and mutants on DP2, DP4 and DP5 substrates. Activity determined by HPLC-MS method. Orange: DP2, Blue: DP4 and Grey: DP5. Buffer: 50 mM phosphate buffer pH 8.5, 300 mM NaCl, 37°C. ..... - 162 -

Figure 12.25. Deacetylation specificity of K275E and K275E-H127R mutants with DP4 substrate. A.1) DP4 substrate (A4, tetraacetyl-chitotetraose). A.2) DP4 incubated with WT VcCDA for 48h, giving the monodeacetylated product A3D (ADAA). B.1) K275E + DP4 reaction

for 24h. B.2) Reaction B.1 incubated with wt VcCDA for 48h. C.1) K275E/H127R + DP4 reaction for 24h. C.2) Reaction C.1 incubated with wt VcCDA for 48h. Reaction conditions: 8 mM substrate, 1 μM enzyme, PBS buffer (50 mM K_2HPO_4, 300 mM NaCl, pH 8.5), 37 °C. ..... - 164 -

Figure 12.26. Thermal stability of wt and mutant *Vc*CDA proteins in presence or absence of Zn. A) Melting temperature: *Vc*CDA wt (green): 56.0 ± 0.17, *Vc*CDA K275E (red), *Vc*CDA K275E-H127R (blue). B) Melting temperature in presence of $ZnCl_2$:*Vc*CDA wt (green),*Vc*CDA K275E (red), *Vc*CDA K275E-H127R (blue) ..... - 165 -

Figure 12.27. Temperature profiles with DP4 substrate. Activity determined by HPLC-MS method. Orange: VcCDA wt, Blue: VcCDA K275E, Grey: VcCDA K275E-H127R. Buffer: 50 mM phosphate buffer pH 8.5, 300 mM NaCl, 37°C. Kinetic parameters are given in Table 4.3. . - 166 -

Figure 12.28. Thermotolerance of free and immobilized enzymes. The thermotolerance of *Vc*CDA WT (green dots), K275E (blue squares), and K275-H127R (red diamonds) was determined with free enzymes at 50°C (A), and with CMB-immobilized enzymes at 50ºC (B) and 60ºC (C). Exponential decay equation has been adjusted (Equation 21.1). ..... - 167 -

Figure 12.29. Mutated residues in the 3D structure of *Vc*CDA. ..... - 169 -

Figure 13.1. Bioinformatic model showing the possible dynamic behaviour of DP5 to allow accommodation of substrates longer than DP3. (Loop 5 in green). ..... - 175 -

Figure 13.2. Comparison of crystallographic structure (A) and computational model (B). Comparative diagram of the crystallographic structure obtained in complex with DP2 and the computational model in which loop 5 has been moved and DP4 was docked in the active site. The exposure of two additional positive subsites was observed. Shown as loop 1 and 6 blocking additional negative subsites. ..... - 176 -

Figure 13.3. Loop 5 blockage location by disulfide bridge. Structural study for the establishment of a disulfide bridge for the loop 5 block. Detail of specific amino acids and distance between them was highlighted. ..... - 178 -

Figure 13.4. GFC chromatogram of VcCDA FL P193C / N273C. Chromatogram obtained on filtration gel chromatography using a Superdex 200 column (16/600). Five multimeric fractions are shown which F5 corresponds to the monomer. ..... - 179 -

Figure 13.5. Mechanism of Ellman's reaction: DTNB. The disulfide bond reacts in the presence of a thiol releasing 2-nitro-5-thiobenzoate (TNB) which, when ionized in water, acquires a yellow color. Adapted for [205]. ..... - 180 -

Figure 13.6. Standard curve obtained by the Ellman method for the titration of free cysteines in protein solutions. ..... - 180 -

Figure 13.7. Kinetic characterization of *Vc*CDA FL P193C/N273C compared to WT FL. A) Deacetylase activity for DP2 substrate. B) Deacetylase activity for DP4 substrate. All assays were done in 50mM Tris buffer pH 8.5 300mM NaCl and at 37°C. ..... - 181 -

Figure 13.8. β factor in the structure of the *Vc*CDA. Structure of the *Vc*CDA represented based on the different values of β factors throughout its sequence. Gradual scale colouring: blue, minimum values - Red maximum values. Access code PDB: 4NY2. ..... - 184 -

Figure 13.9. Increased flexibility by prolines to glycine's mutant. Increasing the flexibility of loop 5. The three prolines are mutated by glycines, a type of amino acid that provides more flexibility to the backbone of a protein. ..... - 185 -

Figure 13.10. Size exclusion chromatography of VcCDA CD PG. Superdex 200. Different fractions were assigned. F4: Dimeric fraction, F5: Monomeric fraction. ..... - 185 -

Figure 13.11. Evolution of activities (k cat) depending on the DP of the substrate. All assays were done in 50mM Tris buffer pH8.5 300mM NaCl and at 37°C. ..... - 186 -

Figure 13.12. Comparison of the monomeric and dimeric fraction of the PG mutant. Graphs showing the comparison of the specific rates for the monomeric (B) (F5) and dimeric fraction (A) of the PG for DP2 and DP4. All assays were done in 50mM Tris buffer pH 8.5 300 mM NaCl and at 37°C. ..... - 188 -

Figure 13.13. Strep-Trap purification chromatogram of individual mutants. A) *Vc*CDA CD (catalytic domain) wt B) *Vc*CDA CD PG (contains three prolines mutated to glycine; P266G, P271G, P280G. C) *Vc*CDA CD P266G D) *Vc*CDA CD P271G E) *Vc*CDA CD P280G. 3 and 4 column loads were carried out in which three column loads were carried out and the usual protocol of loads washes and elutions was followed. The numbers indicate the stages of a charge/elution cycle: 1. Charge, 2. Wash, 3. Elution, 4. Wash, 5. Regeneration ..... - 189 -

Figure 13.14. Chromatogram obtained in gel chromatography filtration using SuperDex 200 column (16/600). A) *Vc*CDA CD (catalytic domain) wt B) *Vc*CDA CD PG (contains three prolines mutated to glycine; P266G, P271G, P280G. C) *Vc*CDA CD P266G D) *Vc*CDA CD P271G E) *Vc*CDA CD P280G. All purificationswere carried out in 50 mM Tris buffer pH 8.5, 300 mM NaCl and at a temperature of 37ºC. ..... - 190 -

Figure 13.15. SDS-PAGE of the *Vc*CDA CD P271G individual mutant as example. M: Marker Low Range, P: Pellet from the centrifugation after the lysate, C: crude, soluble fraction of the lysate, FT: Flow through the column, ST: Protein purified by the Strep-Trap, F1-F5: fractions of the 1 to 5..... - 191 -

Figure 13.16. Kinetic characterization of CD WT, PG and individual mutants. A) *Vc*CDA CD wt, B) *Vc*CDA CD PG (triple mutant), C) *Vc*CDA CD P266G, D) *Vc*CDA CD P271G, E) *Vc*CDA CD P280G. DP4 substrate (red) and DP2 substrate (green). All the assays were carried out in 50 mM Tris buffer pH 8.5, 300 mM NaCl and at a temperature of 37ºC ..... - 193 -

Figure 13.17. Chromatogram obtained in gel chromatography filtration using SuperDex 200 column (16/600) for *Vc*CDA CD P266G-P271G... - 195 -

Figure 13.18. Kinetic characterization of VcCDA CD P266G+P271G. DP4 substrate (red) and DP2 substrate (green). All the assays were carried out in 50 mM Tris buffer pH 8.5, 300 mM NaCl and at a temperature of 37ºC..... - 196 -

Figure 13.19. Evolution of activities in terms of k_{cat} depending on the DP of the substrate for CD wt , PG triple mutant and individual mutants (P280G, P266G and P280G) and double mutant P266G-P271G. ..... - 196 -

Figure 13.20. Evolution of substrate affinities in terms of K_M depending on the DP of the substrate for CD wt, PG triple mutant and individual mutants (P280G, P266G and P280G) and double mutant P266G-P271G. ..... - 197 -

Figure 13.21. Increased flexibility by prolines to glycine's mutant. P266, P271 and P280 positions were highlighted. PDB: 3OUI: *Vc*CDA wt in complex with DP3..... - 198 -

Figure 13.22. VcCDA CD D39S protein expression using glycerol as carbon source (40 g/L) following the protocol . Expected MW: 35 kDa. ... - 201 -

Figure 13.23. SDS-PAGE 14% acrylamide of expression tests of *Vc*CDA. Molecular weight of *Vc*CDA CD wt and PG is 35 kDa. M: Marker, SN: Soluble fraction, P: Pellet. A) Protein expression of *Vc*CDA CD D39S at different times. B) Protein expression of *Vc*CDA PG D39S at different times. C) Protein expression after 24h at 24ºC and 250 rpms. CD: *Vc*CDA CD D39S, PG: *Vc*CDA PG D39S..... - 202 -

Figure 13.24. Study of protein expression at different intervals of time at 20ºC using *Vc*CDA CD D39S enzyme. M: Marker, SN: Soluble fraction, P: Pellet. ..... - 203 -

Figure 13.25. Scheme of final protocol for protein expression in minimal media. M9: minimal media (see section 21.4.3) ..... - 203 -

Figure 13.26. A) Chromatogram of Strep-trap purification of *Vc*CDA CD D39S. Resulting chromatogram from affinity chromatography in which three cycles of loading and elution were performed. 1) loading and washing, 2) elution; 3) regeneration with HABA and washing. mAU: 280nm. B) SEC chromatogram of VcCDA CD D39S. Chromatogram obtained on filtration gel chromatography using a Superdex 200 column (16/600). Five multimeric fractions are shown which F5 corresponds to the monomer. ..... - 204 -

Figure 13.27. A) Basic RD experimental scheme. The total echo time is τCP. B) The chemical exchange is sensittive to a variation in the number of 180º pulses allocated during τCP..... - 207 -

Figure 13.28. $^{1}H,^{15}N$-HSQC spectrum of *Vc*CDA (800 MHz, 298K). Only the signals from flexible regions are observed. The heterogeneous intensity patter is compatible with the protein undergoing pervasive motions..... - 208 -

Figure 13.29. Methyl TROSY spectra of ^{13}C-Met Met/Ileδ (800 MHz, 298K) for CD wt (blue) and CD PG (red). CD wt shows a second set of signals (marked in a box). ..... - 210 -

Figure 13.30. A-B) Methyl TROSY spectra of ^{13}C-Met Met/Ileδ (800 MHz, 298K) for CD wt (blue) and CD PG red). CD wt shows a second set of signals. The regions for the δ-Ile and Met residues are shown in panels A and B respectively. Peaks assigned to the minor species are identified with an asterisk. The bound signals follow the same color than the regions, while peaks from CD wt and CD PG are colored in blue and red respectively. ..... - 211 -

Figure 13.31. NOESY/EXSY spectrum of CD wt. The square-highlighted peaks connect minor and major species peaks and, therefore, are exchange peaks. 1st species corresponds to major species and 2nd to minor species. ..... - 212 -

Figure 13.32. Methyl TROSY spectra of ^{13}C-Met Met/Ileδ (800 MHz, 298K) for CD wt (red) and CD wt + DP4 (red) in presence of DP4 substrate. Notice that the shifts induced upon substrate binding (black arrows) are mostly produced in the minor species. ..... - 213 -

Figure 13.33. A-B) Methyl TROSY spectra of ^{13}C-Met Met/Ileδ (800 MHz, 298K) for CD wt (blue) and CD PG (red). CD wt shows a second set of signals. The regions for the δ-Ile and Met residues are shown in panels A and B respectively. Peaks assigned to the minor species are identified with an asterisk. The bound signals follow the same color than the regions, while peaks from CD wt and CD PG are colored in blue and red respectively. C) Selected regions to show the overlap between apo and DP4 bound CD wt (left) or CD PG (right). Blue color peaks indicates apo enzyme and coloured peaks indicated enzyme-DP4 complex. The bound signals follow the same color than the regions, while peaks from CD wt and CD PGare colored in blue and red respectively. ..... - 214 -

Figure 13.34. a) In induced-fit binding, the change between the conformations P_1 and P_2 of the protein occurs after binding of the ligand L. The intermediate state P_1L relaxes into the bound ground state P_2L with rate k_r, and is excited from the ground state with rate k_e. (b) In conformational-selection binding, the conformational change of the protein occurs prior ligand binding. The intermediate state P_2 is excited from the unbound ground state P_1 with rate k_e, and relaxes back into the ground state with rate k_r. ..... - 215 -

Figure 13.35. Methyl TROSY spectra of 13C-Met Met/Ileδ (800 MHz, 298K) for mutants M242A (left, red) and I240L (right, red) of CD VcCDA and compared to the apo form (blue). Peak numbers (1-M242 and 15-I240) correspond to peaks in Figure 13.36. ..... - 216 -

Figure 13.36. Spectrum of CD wt with all the peaks investigated by RD. The peaks M242 and I240 correspond to peaks 1 and 15 respectively. ..... - 217 -

Figure 13.37. RD to investigate dynamics in a faster timescale (μs-ms). Examples of RD profile for Met242 (Met1). ..... - 217 -

Figure 13.38. Per-residue averaged root-mean-squared fluctuations of catalytic domain of VcCDA along free MD simulations. Top panel: absolute RMSF values for both wild-type and triple mutant P266G/P271G/P280G variant. Bottom panel: differences of RMSF values between both variants. Color bars in the horizontal axes indicate the location of the six loops surrounding the active site of VcCDA: LOOP1 (yellow), LOOP2 (blue), LOOP3 (red), LOOP4 (orange), LOOP5 (green) and LOOP6 (black). The negative peak at position 162 corresponds to an important side-chain rotation of a solvent exposed Tyrosine in thisposition. Tyr162 is located in the opposite site of the catalytic center, thus this movement is not considered relevant for this discussion. ..... - 220 -

Figure 13.39. *Vc*CDA structure in cartoon representation colored according to differences of RMSF values between both variants: red, high flexibility; blue, low flexibility. ..... - 221 -

Figure 14.1. Binding subsites identification in *Colletotrichum lindemuthianum* CDA. A) model of the complex between the crystal structure of the apo form and chitotriose. B) model of the complex with chitotetraose. Surrounding loops defining the binding cavity are represented as topological surface. The rest of the protein is represented as cartoons and colored in white. The active metal cation is represented as a cyan sphere in subsite 0. The substrate is represented as thick lines. ..... - 228 -

Figure 14.2. 3D structure of *Vc*CDA in complex with DP3 highlighting loop 1 (yellow), loop 2 (blue) and loop 6 (black) and residues involved in the extensive network present in negative subsites. PDB:4OUI. ..... - 229 -

Figure 14.3. Main interactions between loop 1 (yellow), loop 2 (blue) and loop 6 (black). ..... - 231 -

Figure 14.4. Schematic illustration of iterative saturation mutagenesis involving (as an example) four randomization sites A, B, C and D: confined protein sequence space for evolutionary enzyme optimization (redundancy in some cases is expected). Adapted from [227]. ... - 233 -

Figure 14.5. Schematic diagram of ISM strategy for directed evolution approach containing four sites. Site I: D74X-E75X-N76X, Site II: N119X, Site III: E300X, Site IV: K303X-R304X. ..... - 235 -

Figure 14.6. Scheme of secondary screening adapted from protocol presented in section 12.2.6. ..... - 238 -

Figure 14.7. Z-factor calculation for adapted secondary screening assay. Equation 12.1. C+: Positive control (VcCDA P193C-N273C), C-: Negative control (empty plasmid, pET22b plasmid without gene cloned), 3SD: 3 times standard deviation. ..... - 239 -

Figure 14.8. Comparation of secondary screening assay for VcCDA wt, VcCDA P193C-N273C and VcCDA D39S using DP2 (A) and DP4 (B) substrates. ..... - 240 -

Figure 14.9. Comparation of secondary screening assay for VcCDA wt, VcCDA P193C-N273C and VcCDA D39S using DP4 as a substrate in presence of (A) oxidation agent (10 mM H_2O_2) and (B) reducing agent (10 mM DTT). ..... - 241 -

Figure 14.10. 96-well microplate distribution for picking cells. ..... - 242 -

Figure 14.11. Screening of 18400 clones of the *Vc*CDA D74X-E75X-N76X library of mutants. X-axes contains all the mutants from each plate sorted by rows (A1, B1, C1, D1..). Each symbol and colour belongs to the same plate. ..... - 243 -

Figure 14.12. Mean of the positive controls (VcCDA P193C-N273C: S-S) and negative controls (plasmid empty: noP) of all plates screened... - 244 -

Figure 14.13. 275 clones selected as a positives hits from 18400 clones screened. ..... - 245 -

Figure 14.14. Example of a distribution for secondary screening assay. The colour of each well indicates the plate of source of each mutant (for example, mutants selected from P001 have a grey colour). Dark green (E4 to E6): positives control (*Vc*CDA wt), Pale green (E7 to E9): positives control (*Vc*CDA P193C/N273C: PdiS), red (E10 to E12): negative control (*Vc*CDA D39S). ..... - 245 -

Figure 14.15. Results of secondary screening assay for 4 plates containing the positives hits from primary screening assay. % of DP4(N) regarding the total amount of DP4 (2mM). Green triangle: Activity with DP4 using HPLC-MS after 24h. Red circle: DP4(N) detected using fluorescamine at 1 hour reaction. Blue square: DP2(N) detected using fluorescamine at 1 hour reaction. ..... - 247 -

Figure 14.16. Results of secondary screening assay for 14 mutants selected as most promising ones. ..... - 248 -

Figure 14.17. SDS-PAGE 14% of acrylamide of purification using Streptag spin column. M) molecular weight marker III. Wt , P193C-N273C (S-S) and different mutants were analyzed. ..... - 249 -

Figure 15.1. Glycosidase mechanism. A: Inverting glycosidase. B: Retaining glycosidase with enzyme nucleophile, via a glycosyl–enzyme intermediate in the two-step displacement mechanism. C: Retaining glycosidase by substrate-assisted catalysis, via an oxazolinium ion intermediate. Adapted from [232] ..... - 258 -

Figure 15.2. Mechanism of double displacement of a glycosidase with retention of configuration, in the lower branch of the scheme, the transglycosylation reaction with a glycosidic acceptor, is shown. Adapted from ..... - 260 -

Figure 15.3. Mechanism of transglycosylation of a glycosynthase by mutation of the nucleophilic residue. ..... - 260 -

Figure 15.4. 3D structure of β-glucosidase from *Streptomyces sp.* PDB code: 1GON. ..... - 262 -

Figure 15.5. pNP-Septanoside molecule. ..... - 264 -

Figure 17.1. (a) Portulasoid 1 and 20-hydroxy-ecdysone septanoside 2 from Atriplex portulacoides roots; D-glycero-D-manno-heptose in its pyranose (3) and septanose (4) ring forms. (b) Hydrolysis of X-Gal 5, giving rise to D-galactose and 3-hydroxyindole 6. (c) Oxidative dimerization of 3-hydroxyindole 6 to indigo dye 7. ..... - 273 -

Figure 17.2. (a) Indolyl septanosides 8 and 9. (b) Synthetic routes, including yields, used to prepare indolyl septanosides 8 and 9. ..... - 276 -

Figure 17.3. Blue-white colony screen of E. coli BL21(DE3) Star cells with indolyl septanoside 8. BL21(DE3) Star cells harboring a plasmid expressing E. coli β-galactosidase: (Left) with X-Sept (8) substrate; and (Right) with substrate X-Gal (5). ..... - 278 -

Figure 17.4. Kinetics of *Streptomyces sp.* β-galactosidase (Bgl3) with pNP-Sept (19) and pNP-Glc substrates. Conditions: 50 mM phosphate buffer, pH 6.5, and 50 °C. ..... - 279 -

Figure 17.5. Synthesis of BODIPY-labeled septanoside 23. ..... - 280 -

Figure 17.6. Confocal fluorescence microscopy of E. coli MG1655 cells with labeling stains. DIC, optical microscopy image; DAPI, DNA staining (bacterial chromosome); FM 4-64, membrane staining; Sept-BODIPY, BODIPY-labeled septanoside (23). Bar scale (5 μm) is the same in all images. ..... - 281 -

Figure 18.1. Set of mutants list highlighted in the manual docking structure between βgl3 and septanose substrate ..... - 296 -

Figure 18.2. Interaction map between WT βgl3 and Glc-pNP. Catalytic residues (E178, E383). In blue, second shell residues. ..... - 298 -

Figure 18.3. WT and W430A with Glc-pNP Michaelis Menten plots. Kinetics assays were done using PB 50 mM pH 7.2 NaCl 0.3M as a buffer. ..... - 299 -

Figure 18.4. Chemical representation of 1st generation of septanoside Me β-2 and 2nd generation Me β-3. ..... - 300 -

Figure 18.5. Snapshot of docking of W430A mutant using septanose 1st (Me β-2) as a substrate. ..... - 301 -

Figure 18.6. W430A Septanose docking over WT Glc-pNP. Superposition is repeated highlighting one and another, in each case, for the opaque (not transparent) molecule carbons are assigned. ..... - 302 -

Figure 18.7. Conformational map for septanosides showing lowest energy conformers of Me β-2 (brown) and Me β-3 (blue); conformers within 3-5 kcal/mol are labelled with dots. ..... - 303 -

Figure 18.8.) Docking clusters from first ($^{5,6}TC_{3,4}$ - blue) and second ($^{3,4}TC_{1,2}$ – red) Me β-3 conformers with βgl3 W430A. Cluster 1: “productive” binding at donor subsite; Cluster 2: binding into the W-hole; Cluster 3: binding into the acceptor pocket. ..... - 305 -

Figure 18.9. Active site of βgl W430A with Me β-3 (blue, cluster 1) and pNP β-1 (light grey). ..... - 305 -

Figure 21.1. Standard curve of the Superdex 200 column calibration. "Y = -1.399*X + 4.118" ..... - 338 -

LIST OF TABLES

5. LIST OF TABLES

Table 9.1. Different biological activities depending on the chitooligosaccharides sequence. ..... - 72 -

Table 9.2. IUPAC nucleotide code. ..... - 106 -

Table 11.1. Kinetic parameters of wt enzyme. Conditions: 25 μM to 4 mM substrate, 0.05-0.8 μM enzyme, 50 mM phosphate buffer pH 8.5, 300 mM NaCl, 37ºC. ..... - 128 -

Table 11.2. Specific activity (s^{-1}) for wt at different temperatures. Reaction conditions: Buffer: 50 mM phosphate buffer pH 8.5, 300 mM NaCl, 37 °C. ..... - 130 -

Table 11.3. Thermotolerance of *Vc*CDA FL wt free enzyme at 50°C. Half-live (t_{50} min) monitoring residual activity with DP4 substrate (HPLC-MS). ..... - 133 -

Table 12.1. Kinetic parameters of wt and mutant *Vc*CDA enzymes selected from the HTS assay. Conditions: 100 μM to 4 mM substrate, 0.05-0.8 μM enzyme, 50 mM phosphate buffer pH 8.5, 300 mM NaCl, 37ºC. ..... - 162 -

Table 12.2. Thermal stability of wt and mutant *Vc*CDA proteins in presence or absence of Zn^{2+}. ..... - 165 -

Table 12.3. Specific activity (s^{-1}) for different mutants at different temperatures. Reaction conditions: Buffer: 50 mM phosphate buffer pH 8.5, 300 mM NaCl, 37°C. ..... - 166 -

Table 12.4. Thermotolerance of free enzymes at 50°C and complex CMB-Enzyme at 50ºC and 60ºC. inactivation (K_i) constant and half-live (t_{50} min) monitoring residual activity with DP4 substrate (HPLC-MS). ..... - 168 -

Table 13.1. Results of cysteine titration for VcCDA FL and P193C/N273C using Ellman's method. ..... - 181 -

Table 13.2. Kinetic parameters of the characterization of *Vc*CDA FL P193C/N273C compared to those of *Vc*CDA WT FL parameters. ... - 182 -

Table 13.3. Kinetic parameters for *Vc*CDA CD and PG. All assays were done in 50 mM Tris buffer pH 8.5, 300 mM NaCl and at 37°C. .. - 186 -

Table 13.4. Specific activities of dimeric (F4) and monomeric (F5) form using DP2 and DP4 substrates (data obtained from Figure 13.12.). .....- 188 -

Table 13.5. Purification yield of *Vc*CDA CD wt, PG triple mutants and individual mutants. ..... - 191 -

Table 13.6. Kinetic parameters of CD WT, PG and individual mutants. As cooperative behavior was not observed, the Michaelis Menten equation was used in all cases. All the assays were carried out in 50 mM Tris buffer pH 8.5, 300 mM NaCl and at a temperature of 37ºC. - 194 -

Table 13.7. Kinetic parameters of VcCDA CD P266G+P271G. k_{cat} and K_M. All the assays were carried out in 50 mM Tris buffer pH 8.5, 300 mM NaCl and at a temperature of 37ºC. ..... - 196 -

Table 13.8. Specific activity of *Vc*CDA Cd from different expressions systems using DP4 as a substrate. Buffer: Tris 50 mM pH 7 5, NaCl 0.3 M. ..... - 205 -

Table 13.9. Methionine's and isoleucine's in CD wt and CD PG. Nt: N-terminal methionine. ..... - 209 -

Table 13.10. Exchange analysis of M240. ..... - 218 -

Table 14.1. Distance between main interactions of loop 1, loop 2 and loop 6. ..... - 231 -

Table 14.2. Biochemical characterization of proposed mutants. Protein expression and deacetylase activity using DP2 and DP4 was compared to *Vc*CDA wt enzyme. ..... - 232 -

Table 14.3. Statistical analysis of codon usage. [a]Number of clones to be screened for 95% coverage (over-sampling) when two or three amino acid positions at a given site are randomized using a specific degenerate codon. N: A/C/G/T; D: A/G/T; T ; B; not A; K:T/G; D; not G; R; A/G. Adapted from [227]. ..... - 237 -

Table 14.4. Best sequences selected from sequencing promising mutants obtained from secondary screening assay. (wt seq: D74-E75-N76). % DP4(N) HPLC activity was previously shown in Figure 14.15. - 248 -

Table 14.5. List of DA (-wt: nefore addition VcCDA wt, +wt: after addition VcCDA wt) and PA analysis of all 11 purified mutants. - 250 -

Table 15.1. Michaelis-Menten Parameters of *Streptomyces sp.* β-Glucosidases (wt and Mutants) with Cellobiose and p-Nitrophenyl β-Glycoside Substrates. Adapted from [238]. - 262 -

Table 17.1. *In vitro* screen of selected glycosidases. - 279 -

Table 18.1. Specific activity and percentage of remaining activity for every mutant with Glc-pNP and Gal-pNP. Kinetics assays were done using PB 50 mM pH 7.2 NaCl 0.3M as a buffer. - 297 -

Table 18.2. Michaelis Menten parameters for WT and W430A enzymes with Glc-pNP as substrate. - 299 -

Table 18.3. Summary of Docking experiments for septanose ligands with wt βgl3 and single mutants W430A and W430L. Key: + = populated, but low counts; ++ = highly populated cluster; flip = ligand orientation is flipped relative to pyranose; ⊥ = ligand is bound perpendicular to the productive pose orientation; W-hole = ligand occupies space created by removing tryptophan residue. - 304 -

Table 21.1. Flanking primers designed for error prone PCR - 324 -

Table 21.2. PCR reaction mixture - 324 -

Table 21.3. PCR reaction conditions - 324 -

Table 21.4. Cloning primers for *Vc*CDA individual and double mutants. - 329 -

Table 21.5. PCR reaction mixture - 329 -

Table 21.6. PCR reaction conditions - 330 -

Table 21.7. Cloning primers for *Vc*CDA M242A and I240L mutants for NMR studies. - 330 -

Table 21.8. Cloning primers for alanine scanning mutants. - 331 -

Table 21.9. M9 media composition, stock and final concentration. All solutions except salts (autoclaved) are prepared by filter sterilization. - 334 -

Table 21.10. M9 media composition, 100 mL and 600 mL final volumes. - 334 -

Table 21.11. Elution volumes for calibration of Superdex 200 column. Elution times of the proteins and compounds used for the calibration of the gel filtration column and data transformation to obtain the standard curve. - 337 -

Table 21.12. NMR Experiments conducted in the present project. - 343 -

LIST OF EQUATIONS

6. LIST OF EQUATIONS

Equation 9.1. (a) Michaelis-Menten equation (b) Hill equation. Allosteric sigmoidal equation. ..... - 94 -

Equation 11.1. Boltzman Sigmoidal equation. ..... - 131 -

Equation 12.1. Z-factor equation. Equation is defined by: the means (μ), and standard deviation (σ), of both the positive (p) and negative (n) controls. ..... - 152 -

Equation 14.1. Degeneracy of a library. D_{max}: maximum degeneracy of a library (assuming infinite number of transformants), T: number of transformants. ..... - 235 -

Equation 14.2. Number of variants V_H of a protein with aminoacids length L. L: number of amino acids in a protein sequence. M: aa exchanges per molecule. ..... - 236 -

Equation 14.3. Product of probabilities of each library is present in the library. P_i: probability that a particular sequence I is in the library. - 236 -

Equation 14.4. Probability of occurrence of each variant of a library. F: frequency that a particular sequence i is present in a library. D_{max}: maximum degeneracy of a library (assuming infinite number of transformants), ... - 236 -

Equation 14.5. Probability of occurrence of each variant of a library. F: frequency that a particular sequence i is present in a library. P_c: probability that the library is complete. D: degeneracy of a library. ..... - 236 -

Equation 17.1. M&M equation with substrate inhibition ..... - 287 -

Equation 17.2. Sigmoidal equation ..... - 287 -

Equation 21.1. One phase exponential decay equation. ..... - 326 -

Equation 21.2. Calculate the apparent Mw of the proteins ..... - 338 -

ABBREVIATIONS

7. ABBREVIATIONS

[X]	*Concentration of compound X*
3D	*Three dimensional*
A	*GlcNAc, in oligosaccharide sequences of chitin and chitosan*
AcOMU	*4-methylumbelliferyl acetate*
DNA	*Deoxyribonucleic acid*
BCA	*Bicinchoninic acid*
BSA	*Bovine serum albumin*
CAZy	*Carbohydrate Active EnZymes*
CBB	*Chitin Binding Buffer*
CBM	*Carbohydrate Binding Module or Chitin Binding Module*
CCME	*Chitin and Chitosan Modifying Enzyme*
CD	*Catalytic domain*
CDA	*Chitin deacetylase*
CE	*Carbohydrate esterases*
CMB	*Chitin Magnetic Beads*
COD	*Chitooligosaccharide deacetylase*
COS	*Chitosan oligosaccharides*
CPEC	*Circular Polymerase Extension Cloning*
D	*GlcN, in oligosaccharide sequences of chitin and chitosan*
DA	*Degree of acetylation*
DMF	*Dimethylformamide*
DP	*Degree of Polymerization*
DSF	*Differential Scanning Fluorimetry*
EDTA	*Ethylenediaminetetraacetic acid*

epPCR *Error-prone Polymerase Chain Reaction*

ESI *Electro Spray Ionization*

FL *Full-length*

FPLC *Fast Protein Liquid Chromatography*

GH *Glycosil Hydrolase*

GlcN, GlcNH2 *Glucosamine*

GlcNAc *N-acetyl glucosamine*

GT *Glycosil transferase*

HABA *2-Hydroxy-4-aminobenzoic acid*

HILIC *Hydrophilic Interaction Liquid Chromatography*

HPLC *High Performance Liquid Chromatography*

HPSEC *High Performance Size Exclusion Chromatography*

IL *Interleukin*

IMAC *Ionic Metal Affinity Chromatogtraphy*

IPTG *Isopropyl β-D-1 thiogalactopyranoside*

LB *Luria-Bertani culture medium*

MALDI-TOF Matrix Assisted Laser Desorption/Ionization - Time Of Flight

M&M *Michaelis&Menten*

MCS *Multiple Cloning Site*

MS *Mass Spectrometry*

MSD *Mass Spectrum Detector*

MT *Motif*

MU *4-methylumbelliferone*

MurNAc *N-acetyl muramic acid*

Mw *Molecular weight*

MWCO *Molecular weight cutoff*

NMR *Nuclear Magnetic Resonance*

OD *Optical density*

OD600 *Optical density at 600 nm*

o/n *Over night*

PA *Pattern of Acetylation*

paCOS *Partially Actylated Chitosan Oligosaccharides*

PAMP *Pathogen Associated Molecular Pattern*

PB *Phosphate Buffer*

PBS *Phosphate Buffer Saline*

PCR *Polymerase Chain Reaction*

PDB *Protein Data Bank*

PGN *Peptidoglycan*

PMSF *Phenylmethylsulfonyl fluoride*

QC-PCR *Quick Change Polymerase Chain Reaction*

RMSD *Root Mean Square Deviation*

RNA *Ribonucleic acid*

rpm *Revolutions per minute*

RT *Room temperature*

RT-PCR *Real-time PCR*

SDS *Sodium Dodecyl Sulfate*

SDS-PAGE *Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis*

SIM *Single Ion Monitoring*

SEC *Size Exclusion Chromatography*

SD *Standard Deviation*

SMD	*Site directed mutagenesis*
TG	*Transglycosidase*
Tm	*Melting temperature*
TLR	*Toll-like receptor*
TSA	*Thermal Shift Assay*
UHPLC	*Ultra High-Performance Liquid Chromatography*
UPLC	*Ultra Performance Liquid Chromatography*
UV	*Ultraviolet*
v_0	*Initial velocity*
V_0	*Dead volume*
V_e	*Elution volume*
VMD	*Visual Molecular Dynamics*
VWD	*Variable Wavelength Detector*
wt	*wild type*

THESIS OUTLINE

8. THESIS OUTLINE

The work carried out in this doctoral thesis revolves around the evaluation of the Carbohydrate Esterase Family 4 (CE4) and the family of Glycosyl Hydrolases 1 (GH1) for the acceptance of septanose molecule as substrate which will be unveiled in Chapter 5.

TOPIC 1: Engineering substrate specificity of chitin deacetylases

Regarding CE4 family, the main objective of the project is to redesign the chitin deacetylase of *Vibrio cholerae* to increase the activity towards longer substrates using directed evolution and rational design approaches.

The second main objective is to understand the molecular basis of loops dynamics in *Vc*CDA enzyme which dictate the enzyme specificity and permit us to experimentally demonstrate the subsite capping model approach. Finally, the last goal is the attempt to engineer and redesign the *Vc*CDA enzyme using a semirational approach towards new a deacetylation pattern.

This knowledge will contribute to exploit chitin deacetylases both as biocatalysts and as therapeutic targets against fungal infections. In this context, the specific chapters are:

CHAPTER 1. **Biochemical characterization of chitin deacetylase from *Vibrio Cholerae***

CHAPTER 2. **Screening assay for directed evolution of chitin deacetylases.**

CHAPTER 3. **Engineering loop flexibility to tune enzyme specificity in *Vc*CDA.**

CHAPTER 4. **Engineering *Vc*CDA enzyme towards new deacetylation patterns**

TOPIC 2: Engineering a septanoside hydrolase

The last topic of this thesis addressed a different protein engineering objective on glycosyl hydrolases, our goal being the identification of a naturally or mutated enzyme capable of hydrolyzing ‘’septanoside’’ substrates. This is the beginning of a new project to develop as brand-new enzymatic tool in chemical biology: Engineering septanoside hydrolases and glycosynthases.

CHAPTER 5. Indolyl Septanoside Synthesis for In Vivo Screening of Bacterial Septanoside Hydrolases

CHAPTER 6. **Towards engineering a β-glucosidase to accept seven-member ring septanoside substrates.**

TOPIC 1: Engineering substrate specificity of chitin deacetylases

INTRODUCTION

Part of this Chapter has been published in:

Chitin deacetylases: structures, specificities, and biotech applications.
L. Grifoll-Romero, S. Pascual, H. Aragunde, X. Biarnés, A. Planas.
Polymers 10, 352 (2018).
DOI: https://doi.org/10.3390/polym10040352

Carbohydrate de-*N*-acetylases acting on structural polysaccharides and glycoconjugates.
Sergi Pascual and Antoni Planas
Curr Opin Chem Biol. 2021 Apr;61:9-18.
DOI: https://doi.org/10.1016/j.cbpa.2020.09.003

9. TOPIC 1: Engineering substrate specificity of chitin deacetylases. INTRODUCTION

9.1. Biological implications and applications of chitin and chitosan oligosaccharides.

9.1.1. Structural properties of chitin and chitosan.

Chitin was first named fongine by Braconnot and then chitine by Odier. Children revealed the nitrogenous nature of chitin in 1824. The history of chitosan, the main derivative of chitin, dates back to 1859 with the work of Rouget [1]. The name of chitosan was, however, introduced in 1894 by Hoppe-Seyler [2]. In 1876, Ledderhose hydrolyzed arthropod chitin and discovered glykosamin, the first derivative of chitin. As the precursor of chitosan, chitin is the most widely occurring homopolymer in nature after cellulose; it can be found in a range of eukaryotic species such as crustacea, insects and fungi. Its main function is to act as a structural component of different tissues: arthropod exoskeletons (insects and crustaceans), mollusc endoskeletons (squid, cuttlefish, octopus, etc), cell walls of fungi and diatoms [3] [4]. Moreover, chitin has recently been annotated in deuterostomes and vertebrates as well [5] [6].

Chitin is defined as a linear polysaccharide composed of *N*-acetyl-D-glucosamine monomers linked by β-(1-4) bonds [7]. Chitin appears in nature as macrofibrils, depending on the organization of the polymers within these bundles, 3 types of chitin were identified which can be differentiated by infrared and solid-state NMR spectroscopy together with X-ray diffraction [4]:

- **α-chitin**: It is composed of antiparallel orientated chains with inter- and intra-sheet hydrogen bonds and is mainly found in crustaceans. These α-chitins have proved particularly interesting for structural studies since some of them remarkably present high crystallinity together with high purity (they are synthesized in the absence of pigment, protein, or calcite).
- **β-chitin**: It is composed of parallel chains with intra-sheet hydrogen bonds and is found mainly in the skeleton of squids. As opposed to α-chitin, β-chitin sheets cannot be penetrated by water or alcohol, being a more stable form in natural conditions and,

hence, the most abundant polymorphic form found in nature [4] [6]. A particularly pure form of β-chitin is found in the monocrystalline spines excreted by the diatom *Thalassiosira fluviatilis.*

- **γ-chitin**: It is composed of a mixture of parallel and antiparallel chains. It is found in some fungi.

Natural chitin occurs partially deacetylated (with a low content of glucosamine units), depending on the source; nevertheless, both α and β forms are insoluble in all the usual solvents despite natural variations in crystallinity. The insolubility is a major problem that confronts the development of processing and uses of chitin. The processing of chitin mainly through depolymerization and de-*N*-acetylation reactions generates a series of derivatives that include chitosan and the oligosaccharides of chitosan (COS). Nowadays, it is known that several fungi species are chitosan producers [8]. These polymers with a lower degree of polymerization have remarkable biological functions .

The natural process of deacetylation of chitin is not very common, in fact only some fungi of the phylum *Zygomycota, Basidiomycota* and *Ascomycota* have been described as capable of synthesizing chitosan polymers [9]. Regardless of the processing method, when chitin is partially deacetylated it is called chitosan. This deacetylation makes the new polymer more soluble and acquires new properties that original chitin did not have due to the positive charge of the amino group formed.

The functions of chitosans are very diverse depending on [10] (Figure 9.1):

- **Degree of polymerization (DP)** that is the number of monomeric units that make up a macromolecule, polymer or oligomer.
- **Degree of acetylation (DA)** that is the relative abundance of acetylated monomers in a polymer or oligomer and ranges from 0 in fully deacetylated chitosans to 60% in chitosans that are barely soluble in water. Variations in this parameter affect both its physical-chemical properties (solubility, charge, reactivity, flexibility, polymer conformation, viscosity, crystallinity, high surface area, porosity, tensile strength, conductivity and photoluminescence) and biological properties (biodegradability, biocompatibility, antimicrobial..).
- **Pattern of acetylation (PA)** is the distribution of acetylated residues in a polymer or oligomer in which there is also the presence of non-acetylated residues. Depending

on their distribution, it can be random, blockwise or regular. This distribution of GlcNAc and GlcN in the chitosan oligomers defines their biological activities [11] [12].

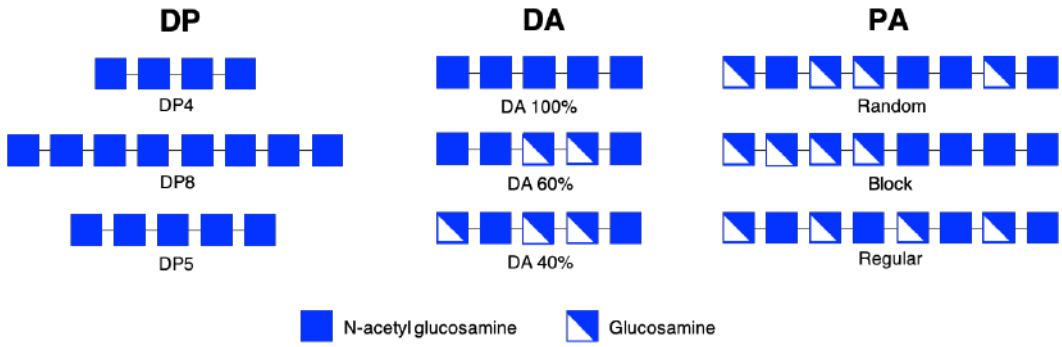


Figure 9.1. Scheme of the key factors that describe chitosan oligomers: degree of polymerization (DP), degree of acetylation (DA) and pattern of acetylation (PA).

Chitosan possesses different chemical properties highlighted hereunder [13]: it is a linear polyamine, contains reactive amino groups and hydroxyl groups that offer the opportunity to chemically or enzymatically modify their structures, and it has the ability to chelate a wide range of transition metal ions. Moreover, the free amino groups of the deacetylated units in the polymer are protonated at physiological pH, thus making chitosans or chitooligosaccharides (COS) the only known natural polycationic polysaccharides. On the other hand, a wide variety of different biological properties of chitosans as a material in terms of its interaction with organisms were described: they are biocompatible natural polymers, its biodegradable and are safe and non-toxic polymers [14] (Figure 9.2).

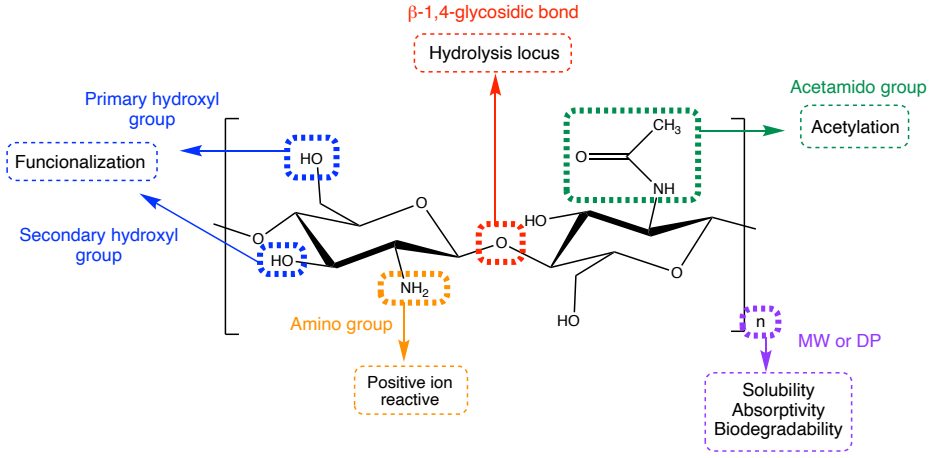


Figure 9.2. Physico-chemical and biological properties of chitosan or COS.

The ubiquity of chitin and chitosan in nature in combination with this set of characteristics and properties, make these polymers highly interesting for several industries and suitable for an extensive number of applications. As polycations they have the ability to interact with several polyanionic biomolecules such as DNA, polysaccharides, enzymes, synthetic anionic polymers, proteins and phospholipids of biological membranes among others [15] [16]. All this set of characteristics (which differentiate them from other natural polymers such as cellulose, agar or pectins, with a more acidic component) have attracted the attention of the industry for their applications in areas as distant as biomedicine, biotechnology and agriculture.

9.1.2. Chitin and chitosan: Biological applications and implications.

Chitosan polymers exhibit a variety of interesting physicochemical and biological properties that make them greatly different from chitin. Chitosans differ in their degree of polymerization (DP) and in their fraction of acetylation (F_A), and both parameters deeply influence the physico-chemical properties and the biological activities of the polymers. These compounds are considered excellent starting products for a large number of applications in areas such as agriculture, cosmetics, water treatment, medicine and the food industry among others. Although chitosan polymers by themselves are already interesting on an industrial level, oligomers (COS, Chitin Oligosaccharides and Partially Acetylated Chitosan Oligosaccharides (paCOS)) have also proven to be as or more relevant due to their current and potential applications within different areas of industrial production of economic relevance [17] [18]. In comparison with chitosan polymers, chitosan oligomers, COS or paCOS (in general, DP less than 20 and an average MW less than 3.9 kDa) have greater solubility, lower viscosity, are soluble in water, easily absorbed through the intestine and rapidly enter into the bloodstream [17]. These properties means that oligomers have a range of different or additional applications to those of polymers [18].

Among its applications within the medical-pharmaceutical industry, the following stand out for its applications:

- **Antioxidant activity** – Oxidative stress is considered as a many cause of different metabolic syndrome, cancer, neuroinflammation and other chronic diseases. COS (low MW and low DA) are considered natural antioxidants with therapeutic effect. They can scavenge free radicals to decrease DNA oxidation and membrane proteins [19][20]. Moreover, different studies confirm that COS inhibit ROS (Reactive Oxygen

Species) and promote the expression of antioxidant genes in cell damage caused by oxidative stress [21] [22].

- **Anti-inflammatory activity** – paCOS (low MW and low DA) are able to inhibit the inflammatory response induced by lipopolysaccharides due to the reduction of the high expression of the inflammatory factors ((IL-1β, IL-17A and IFN-γ) being helpful in the excessive immune response caused by infection or non-pathological condition [23] [20]. These types of pathologies are related to septic shock, inflammatory bowel disease, arteriosclerosis, vascular pathologies, etc [24]. In addition, paCOS also inhibit the activation of basophils, neutrophils, and lymphocytes [25].
- **Immunostimulant activity** - COS have shown promising results in boosting the functions of the innate and adaptive immune system [26]. COS are capable to activate TLR4, TNF-α and IL-1β from macrophages inducing the immune-stimulating activity [27] [28]. Immunostimulatory properties are considered to be one possible explanation for the antitumor activity of COSs.
- **Anti-obesity effects** – COS have the ability to bind fat and bile acids which results in decrease of intestinal absorption of fat and increase of fat excretion leading to a reduction in weight through a mechanism of satiety and the promotion of intestinal motility [29] [30]. In addition, COS have a lipid-lowering effect due to the ability of bind plasma cholesterol leading to a decrease in serum cholesterol levels [31] [32].
- **Antitumor activity** - Several lines of research have shown that COS are capable of interrupting the progression of cancer cells at multiple stages of development: growth, invasion or metastasis. Different studies demonstrate that these oligomers can induce cell death in a large number of tumors of different origin: liver cancer, kidney cancer, leukemia, prostate cancer, etc [33] [18]. Electrostatic interactions contribute both to the binding of COSs to tumor cell membranes and to alterations in the ionic environment, which result in changes in the permeability of tumor cells. On the other hand, there is a mutual repulsion between COS and healthy cells that results in the selective targeting of cancer cells. The MW, DP and acetyl groups also play a major role in antitumor activity [34] [35] [36]. Immunostimulatory properties are considered to be one possible explanation for the antitumor activity of COSs. COS induces the production of factors related to the immune response, which results in increased immunity as well as inhibition of the growth of tumor cells [37].

- **Antibacterial activity** – COS (low MW) exert antibacterial effects on bacteria and fungi. Polycationic COS interact with the negatively charged cell surface of the bacterium producing crosslinking as a result of strong hydrogen bonds that induce their penetration into the bacterial cytoplasm and the prevention of RNA transcription causing bacterial death [38] [39] [40] [41]. Higher MW-COS have strong antibacterial effect on gram-positive bacterial and low MW-COS on gram-negative bacteria.
- **Regenerative activity** - COS have been shown to be able to promote skin and nerve regeneration following tissue injuries. COS are capable of acting as analogues of the cell matrix creating an environment conductive to this regeneration and accelerating wound healing [42] [26]. They have been applied in bandages and in other types of related applications
- **Drug and DNA delivery enhancer** - COS are able to increase the efficiency of DNA/drug delivery to cellular or tissue targets due to their ability to promote permeation through different physiological barriers [43] [44].

In addition to the mentioned biological activities of COS presented above, COS have become a research hotspot in other biological issues like anti-diabetes [45], anti-hypertensive [46], anti-allergy potential [47] and anti-Alzheimer disease [48].

Chitosan oligomers also have very interesting applications within the agricultural sector:

- **Antimicrobial and antifungal**: Various studies using these types of compounds have shown that they are capable to inhibit fungal and bacterial growth in plants in order to control the effect of phytopathogenic species in production crops [49]. It has been reported that the last is increased with higher DP and lower DA [50].
- **Eliciting activities: plant's immune system promotion**: In addition to specific activities against typical plant infections, COS can promote the innate immune response of plant organisms. Chitosans with high DP and also COS with DP6 show resistance related activities in plant cells. They are recognized as PAMP (pathogen-associated molecular pattern) by PRRs (pattern recognition receptors) inducing signalling cascades, stimulating the innate immunity in plants and, consequently, increasing their resistance against pathogens [11] [50]. Regarding the importance of PA role in bioactivities of paCOS, and a very recent publication unequivocally proved for the first time that the α-mono-acetylated COS tetramer with the GlcNAc (A) unit at its non-reducing end had strong disease resistance priming activities in rice cells,

while the priming activities sequentially decreased in the other three isomeric paCOS, and the ω-mono-acetylated one with the GlcNAc unit at the reducing end was completely inactive. Basically, tetramer prepares the plant cells to even low concentrations of resistance reaction-inducing elicitor that would not trigger a naïve cell without activate the cell destructive reaction [51,52].

- **Vegetal growth promotion**: Chitosan oligomers have also been shown to promote growth at different stages from seed germination to fruit development [4] [53].

Some properties already mentioned, such as its antimicrobial capacity, are also interesting for the food and beverage industry. However, there are also other specific applications in this area such as additive, dietary fiber and satiating and, astringent agents [26].

It should be stressed that the precise biological activity of COS is determined by the oligosaccharide chain length (DP) and also the relative abundance of GlcNAc residues (DA) [11] (Table 9.1). This hypothesis is based on the identification of COS with specific-sequence roles in animals [54] and plants [55]. Moreover, different studies presented the importance of well-defined sequenced COS in different biological roles: in plant growth promotion and stress resistance it has been demonstrated that COS with DP=6,7,8 are the most effective in mitigating salt stress in wheat seedlings or in antitumor activity, chitohexaose (D6) was found as the most potent inhibitor in anti-angiogenesis effects. Therefore, a new scenario is emerging in which sequence-specific chitosan hydrolases process partially acetylated chitosan polymers, resulting in smaller modified chitosan polymers and yielding structurally more or less defined mixtures of paCOS. In this storyline, the pattern of acetylation, i.e. the distribution of the GlcNAc (A) and GlcN (D) units in the polymer chain, turns in a third and essential determinant of bioactivities. This significance could be attributed due to the different preferences of the sequence-specific chitosan hydrolases for GlcNAc or GlcN units that determine at which sites a given chitosan polymer will be cleaved, and they will simultaneously determine the occurrence and frequency of GlcNAc and GlcN units at and near the new reducing and non-reducing ends of the paCOS. On the other hand, as already stated in elicitor activity of paCOS, the charges and hydrophobic patches present at the ligand binding site of the receptors will determine which paCOS with defined PA will bind most efficiently and, thus, trigger the bioactivity. In conclusion, PA together with DP and DA will determine the bioactivities of paCOS.

In order to fully understand molecular structure-function relationships and cellular modes of action for each of the many biological interactions reported and determine the sequence defined role of paCOS in all the bioactivities previously defined, we will need to produce fully defined paCOS. The COSs used in most of the studies on biological activities are a mixture obtained by depolymerization using chitin or chitosan as raw materials and de-*N*-acetylation using chemical methods that results in heterogeneous products in DP, DA, or PA leading a lack in reproducibility and divergence in the results of the study of biological activities of COS. Therefore, obtaining COS with certain defined sequence is essential for understanding their biomechanism and the function-structure relationships in the different biological activities.

Table 9.1. Different biological activities depending on the chitooligosaccharides sequence.

Sequences	Bioactivity	References
D_n (n>3)	Growth promoting activity	[56]
ADDD>DADD>DDAD>DDDA	Priming activity	[11]
D_n (n>5)	Antimicrobial activity against *Staphylococcus aureus* (MIC value 500 and 1000 mg/mL)	[57]
D_6, D_7 > D_4, D_5	Reducing the adverse effect of salt stress and further enhanced plant growth	[58]
D_6 > D_2, D_3, D_4, D_5 > COS mix	Anti-angiogenesis activity	[59]
D_6 and D_7 > D_2, D_3, D_4, D_5	Enhance chilling tolerance activity of wheat seedlings	[60]
DA	Pro-angiogenic activity	[61]
D_3 > $D_{4\text{-}5}$, D_5, D_6, $D_{7\text{-}10}$, COS mixture	Hydroxyl radical scavenging activity and reducing power	[62]

9.1.3. Production of chitosan and COS.

Chitin, is a natural resource that arises as a side-product of the fishing industry and the processing of shrimps and crabs. It is also extracted from other sources such as squid exoskeletons, but these have a lower chitin content (15% vs 25%) [63]. On commercial scale, chitin and chitosans are two compounds of great relevance due to their high nitrogen content (between 6 and 8%) compared to synthetic celluloses. Every year, approximately 100 billion tons of chitin are produced by crustaceans, mollusks, insects, fungi and related organisms.

Consequently, its rate of replacement in the biosphere is almost double as that of cellulose [64] [65].

Currently, the most usual method to extract chitin is the chemical approach. The general extraction process involves the use of high demanding chemical and thermal conditions. Its extraction from the raw material and its transformation to chitosan involves five basic operations: pretreatment, depigmentation, demineralization, deproteinization, and deacetylation [66] [67].

The chitosan production process usually begins with the accumulation of crustacean shells or other chitin source. In crustacean shells, protein and chitin are combined to form a protein–chitin matrix, which is then extensively calcified by mineral salts to yield recalcitrant shells. Various chemical protocols have been developed and established for extraction of chitin over the last few decades. The chitin source is first subjected to physical pretreatment by washing, drying, and grinding to remove impurities and reduce its size to a powder of 0.5 mm. Then, the deproteinization step is based in the treatment with strong bases such as 1M NaOH at temperatures up to 160ºC to remove all proteins while extracting chitin [68]. Thirdly, the shell powder is sent to the demineralization unit for removal of calcium carbonate and other minerals using a diluted hydrochloric acid solution at room temperature to prevent chitin hydrolysis. Demineralization step can be performed by different acid treatments using HCl, HNO_3, H_2SO_4, CH_3COOH and HCOOH [69]. Among these acids, HCl is the preferred reagent for removing mineral constituents from crustacean shell walls (CSWs). Demineralization treatments vary with the mineralization degree of each shell, extraction time, temperature, particle size, acid concentration and solute/solvent ratio. It was reported that high temperatures can accelerate the demineralization reaction and some reactions are carried out at higher temperatures. Pigments, such as melanin and carotenoids are removed with 0.02% potassium permanganate at 60ºC, hydrogen peroxide, sodium hypochlorite or with organic solvent mixtures. Lastly, the extracted chitin is deacetylated employing a solution of sodium hydroxide at 50% w/v with ratio chitin to solution of 1/10% w/v at 110ºC and dried in an oven at 100ºC and ready to be commercialized [70] [71] [72] (Figure 9.3.). In some cases, the deacetylation reaction is carried out in the presence of thiophenol as a scavenger of O_2 or under a N_2 atmosphere to prevent chain degradation that invariably occurs due to a reaction under strong alkaline conditions. Deacetylation of chitin releases amine groups and it imparts a cationic characteristic to chitosan. This is particularly remarkable in an acid environment where the majority of polysaccharides are usually neutral or negatively charged [70].

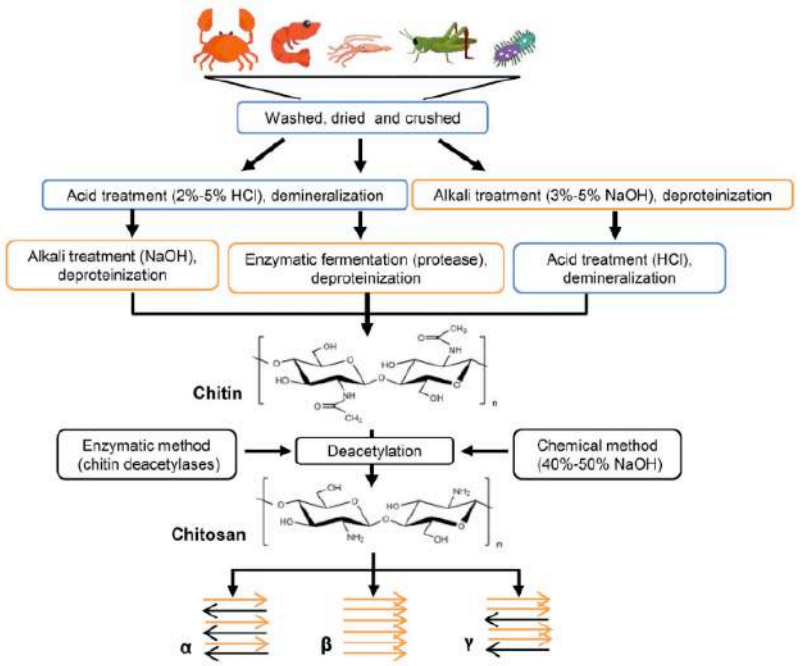


Figure 9.3. Extraction and preparation of chitin and chitosan from natural sources. Adapted from [13].

For the production of chitosan oligomers (COS), the most common chemical protocol is the hydrolysis of polymers [13]. This breaking of the glycosidic bond can ideally be carried out by three processes: chemical agents, enzymatic agents or by high energy impacts [73]. The duration of the treatment, temperature and acid concentration can affect the size of the resulting product. The most frequent conditions include 35% HCl at 80ºC for 1-2 hours [74]. An increase in hydrochloric acid concentration and in temperature also increases the rate of degradation of chitosan by concentrated hydrochloric acid [75]. Another chemical hydrolysis frequently used in industrial level is oxidative degradation by the use of free radicals (H_2O_2) in aqueous solution [76]. Afterwards, a chromatographic method is used to separate different COS depending on the DP and DA. From DP2 to DP8 with DA=0 or DA=100% could be successfully separated using ion exchange chromatography (IEC) and size exclusion chromatography (SEC) [59]. SEC separates the COS according to their size and COS with high DP are generally eluted first. However, COS with different DAs are not separated well.

Another chemical method for COS production is chemical synthesis. GlcNAc and GlcN are required as raw materials for the synthesis of COS [77] [78]. Monosaccharide molecules have multiple free hydroxyls groups, therefore, multiple steps of protection an deprotection are required. The two most important parameters of COS synthesis are the glycosyl groups and

the amino protecting groups. Different protocols for chemical synthesis of COS were reported. For instance, a fully-automated oligosaccharide synthesizer was developed which can synthetize *N*-acetyl chitohexaose [79]. However, there are still some drawbacks to be solved by chemical synthesis of COS with heterogeneous sequence such as the cumbersome steps in addition to the contamination produced. Additionally, and although each step of the reaction can ensure a high yield, the final yield is still very low after dozens of steps of reaction, which is not enough for subsequent biological activity research and application. At the same time, chemoenzymatic strategies such as the use of glycosynthases to perform the transglycosylation reaction without hydrolyzing the product (one-pot reaction) are gaining importance in chemical synthesis strategies to minimize the effect of many disadvantages presented in the other methods [80] [81]. Chitinases are glycoside hydrolases that catalyze the hydrolysis of β(1,4)-glycosidic bonds of chitin and chitosan polymers. Some chitinases have also transglycosydase activity (TG), allowing them to introduce new glycosidic bonds between donor and acceptor sugar molecules with the consequent generation of oligomeric COS. Such transglycosylating chitinases can be used for the in vitro polymerization of COS in a chemoenzymatic approach.

Taken altogether, the traditional chitin extraction and chitosan or chitosan oligomers obtention processes employ harsh chemicals at elevated temperatures for a prolonged time which can harm and alter chitin and chitosan physico-chemical properties (the final product is not a single product with a defined acetylation pattern, nor a specific degree of polymerization and nor a homogeneous degree of acetylation) and are also held responsible for the deterioration of environmental health.

Biotechnological production of chitin offers new perspectives for the production of high value chitosan with defined polymer structures given a promising inputs in biomedicine and pharmacy [68]. In a biotechnological approach, alternative proteolytic enzymes to perform the deproteinization step can be used such as papain, pepsin, chymotrypsin, alkalase, devolvase an pancreatin for extraction and separation of chitin and proteins from shrimp waste to obtain chitin polymer [82] [83]. Bioextraction of chitin is thus a greener, cleaner, eco-friendly and economical process [68]. The use of microorganisms-mediated fermentation processes are highly desirable due to easy handling, simplicity, rapidity, controllability through optimization of process parameters, ambient temperature, and negligible solvent consumption, thus reducing environmental impact and costs.

On the other hand, biotechnological obtention of COS is based in enzymatic hydrolysis of chitin or chitosan which involves the use of either specific enzymes, such as chitinase and chitosanase, or nonspecific enzymes, such as carbohydrases and proteases, to very efficiently recognize and catalyse the hydrolysis of chitin or chitosan at certain sites. The main advantages of enzymatic hydrolysis are the mild reaction conditions, less damage to the environment and yielding chitosan products with MW in the desirable range. In addition to hydrolytic enzymes, chitin and chitooligosaccharide deacetylases play a key role in the obtention of COS with a defined sequence. Most chitin deaceylases belong to the family CE4 [10].

9.2. Nano3bio project and application for paCOS obtention.

As already stated, chitosan and paCOS are some of the most promising biomolecules with a huge range of biological activities in distant fields as agriculture, cosmetics, water treatments, biomedicine and food industry due to their physico-chemical and biological properties [84]. These properties strongly depend on the DP and DA and recently demonstrated on the PA. Most of the current methods presented for chitosan or chitosan oligomers obtention render invariably mixtures of molecules differing in these parameters [10]. In this framework, alternative methods need to be developed to produce partially acetylated COS (paCOS) with a defined and known structure to guide us to understand structure-function relationships [4].

The European Project NanoBioEngineering of BioInspired Biopolymers (Nano3Bio) was created with the main goal of developing an alternative to the current industrial production of chitosans and COS using new biotechnological methods such as enzymatic, chemoenzymatic, genetic and metabolic engineering.

The work of the partners participating in this project ranges from the discovery of enzymes and genes related to the synthesis and transformation of chitin or chitosan substrates, the redesign of biocatalysts using enzymatic engineering to achieve new activities or improvements over previous activities reported, to engineering metabolic pathways in which the genes encoding these new activities are incorporated into the genetic material of microorganisms to create "bacterial cell factories". The objectives of the project also include finding new applications for biotechnologically produced chitosans as well as studying the ecological impact of their production. The main enzymes (chitin and chitosan modifying enzymes) studied within these processes include: chitin synthases, chitin deacetylases, chitinases and chitosanases.

The general designed strategy of Nano3Bio consortium consist in the use of chitin deacetylases (or other enzymes active on COS) to generate chitosans or COS with regular and defined patterns of acetylation and/or, the use of chitinases and chitosanases to hydrolyze the chitosan polymer in specific sequences to obtain defined chitosan oligomers. On the other hand, the chitinases with transglycosilation activity can be used for further polymerization of defined chitosan oligomers previously prepared. *In vitro* strategy permit us to obtain defined chitosans and address the lack of reproducibility on the biologic assays performed with chemically produced chitosans. In contrast, one of the main drawbacks of the

in vivo production is the antimicrobial activity of paCOS due to the toxicity of the product which can disturb the bacterial cell factory when this is generating the paCOS with deacetylated units. In consequence, this makes the use of CDAs as a tool for *in vitro* biocatalysts increasingly attractive. The role taken by our group in Nano3Bio project was to discover, characterize, engineer and develop new strategies for chemoenzymatic production of paCOS with a defined structure in terms of DA, DP and also the PA using chitin deacetylases and chitinases with trasnslgycosilation activity (Figure 9.4).

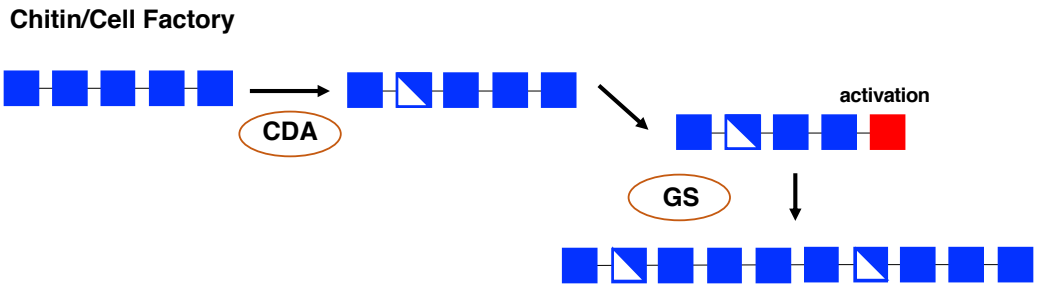


Figure 9.4. Scheme of the biotechnological production of paCOS in Nano3Bio project.

Over the past few years, the consortium have been made different advances. Regarding the *in vitro* approach and COS deacetylation, the discovery and characterization of several CDAs and related CE4 enzymes contributed to enlarge the knowledge on their specificity. With the use of enzymes with different specificities, a large portfolio of paCOS with defined structure can be obtained but more than one biocatalysis step is needed (acetylation and deacetylation steps) causing a reduction in final yield reaction. More recently, all fourteen possible partially acetylated chitosan tetramers have been obtained combining different recombinant bacterial and fungal CDAs, performing both enzymatic deacetylations and *N*-acetylations (Figure 9.5).

The use of these enzymes as a novel methodology for the production of defined partially acetylated COS is promising but still some structural information on protein-ligand complexes is needed to allow the rational design or the discovery of new deacetylates with controlled deacetylation patterns [85].

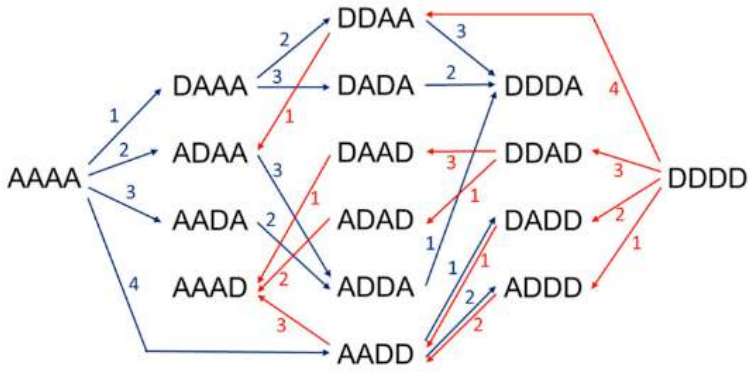


Figure 9.5. Scheme of the production routes of all possible chitin and chitosan tetramers using 4 different CDAs to specifically deacetylate or *N*-acetylate COS. A: GlcNAc; D: GlcNH2. Blue arrows: deacetylation reactions; red arrows: *N*-acetylation reactions in the presence of excess acetate. Adapted from [85].

9.3. CAZy classification, Carbohydrate Esterase family 4 (CE4).

The set of enzymes involved in the assembly, modification and breakdown of carbohydrates and glycoconjugates has been designated as Carbohydrate-Active enZymes (CAZymes) [86–88]. The variety of monosaccharide structures and intersugar linkages is abundant and, considering all types of molecules that can be glycosylated (nucleic acids, proteins, lipids, sugar themselves, metabolites, etc.), the diversity of enzymes acting on these glycoconjugates is equally rich.

Carbohydrate deacetylases belong to the carbohydrate esterase (CE) class of enzymes in the Carbohydrate Active Enzymes classification (www.cazy.org). The CAZy database (http://www.cazy.org) is a curated database that classifies and describes the families of structurally related catalytic and carbohydrate-binding modules of CAZymes. The families are defined based on significant amino acid sequence similarity and common structural folds and the classification is created module by module including only protein sequences released in GenBank (http://www.ncbi.nlm.nih.gov/genbank).

Nowadays, there are five major classes in the CAZy classification:

1. **Glycoside Hydrolases (GHs)**, that hydrolyze and/or rearrange glycosidic bonds.
2. **Glycosyltransferases (GTs)**, that form glycosidic bonds.
3. **Polysaccharide Lyases (PLs)**, that catalyze the non-hydrolytic cleavage of glycosidic bonds.
4. **Carbohydrate Esterases (CEs)**, that hydrolyze carbohydrate esters.
5. **Auxiliary Activities (AAs)**, that include redox enzymes that act in conjunction with other CAZymes.

In the carbohydrate esterase family (CEs), acetylation and deacetylation of biomolecules occurs on many different targets for a wide range of biological functions. The co-translational N-terminal protein acylation that impacts protein stability and localization and acetylation/deacetylation dynamics of histones that control their function are prime examples of structural and regulatory functions of adding or removing acetyl groups on polypeptides [89] [90]. Likewise, deacetylation of *N*-acetylated and O-acetylated carbohydrates are involved in a plethora of biological functions: pathogenic microorganisms that deacetylate their surface glycans to evade immune responses or to support biofilm formation, deacetylation of glycoconjugates in signaling events, deacetylation of structural glycans in

microbial cell wall morphogenesis, and sporulation or remodeling of plant cell wall polysaccharides are just some examples.

Sections 9.3.1 and 9.3.2 are part of the recently published review paper: Pascual S, Planas A. *Carbohydrate de-N-acetylases acting on structural polysaccharides and glycoconjugates.* Curr Opin Chem Biol. 2021 61:9-18. [91]

9.3.1. Carbohydrate esterase family, de-*N*-acetylation mechanisms and structures.

De-*N*-acetylases are grouped into families CE4, 9, 11, 14, and 18, plus few enzymes yet unclassified[1]. De-*N*-acetylation of structural polysaccharides, chitin in fungi and insects, peptidoglycan in bacteria, and poly-β-1,6-*N*-acetylglucosamine in bacterial biofilms, is catalyzed by CE4 enzymes that share a (β/α)7-barrel structure, whereas deacetylation of small oligosaccharides and glycoconjugates is catalyzed by enzymes belonging to different CE families: Diacetylchitobiose and GlcNAc- conjugates by CE14 enzymes with α/β-fold structure, *N*-acetylglucosamine-6-phosphate by CE9 enzymes with (β/α) 8-barrel structure, and UDP-3-O-acyl-N- acetylglucosamine by CE11 enzymes with two-layer-sandwich fold (Figure 9.6). Just recently, a new family CE18 has been created after characterization of a N-acetylgalactosamine deacetylase acting on bacterial exopolysaccharides [92].

Despite different folds, carbohydrate de-*N*-acetylases share a common metal-assisted acid/base mechanism. It was first proposed for the peptidoglycan GlcNAc deacetylase SpPgdA when its X-ray structure was solved in 2005 [93]. The consensus mechanism is shown in Figure 9.7.A. The metal cation (commonly Zn^{2+}, Co^{2+}, Fe^{2+},Mn^{2+},or Cd^{2+}) is coordinated by the conserved motif Asp-His-His. First, the general base residue activates a water-bound molecule to form a tetrahedral oxyanion intermediate, which is stabilized by metal coordination and by other residues from the catalytic cleft. Then, the nitrogen group of the intermediate is protonated by the general acid residue to assist C-N bond breaking, generating a free amine in the de-*N*-acetylated product and releasing acetate as a sub-product [24] (Figure 9.7). The first 3D structure of a chitin deacetylase (CE4, *Vc*CDA) in complex with substrates solved in our group showed that a sugar hydroxyl group also

[1] Family CE-10 has been withdrawn because the majority of its members were esterases active on non-carbohydrate substrates. Thus, there are 15 effective CE families.

participates in metal coordination [94]. Most CE families with de-*N*-acetylase activity contain a single metal cation, except family CE9 that contains both mononuclear and binuclear metal binding sites (Figure 9.7.B) [95].

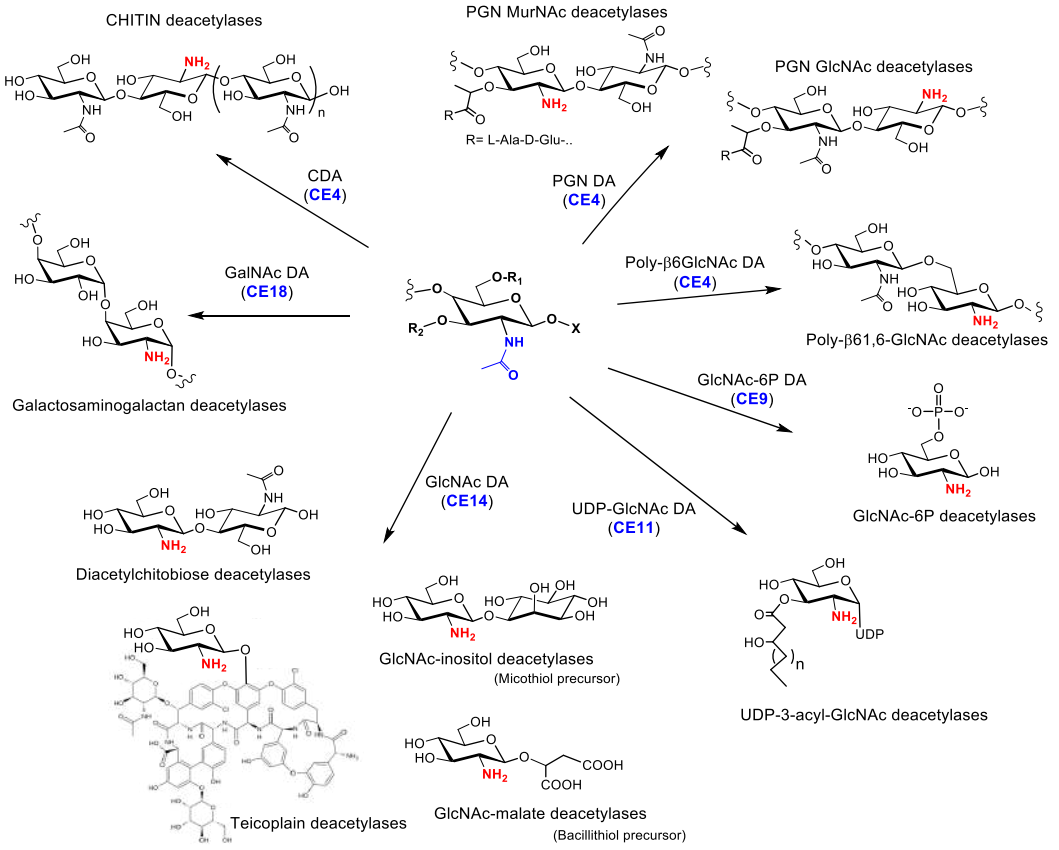


Figure 9.6. Carbohydrate de-*N*-acetylases showing the substrate specificities of each CE family. NH_2- in red is the deacetylated–NAc moiety in the corresponding product. Adapted from [96]

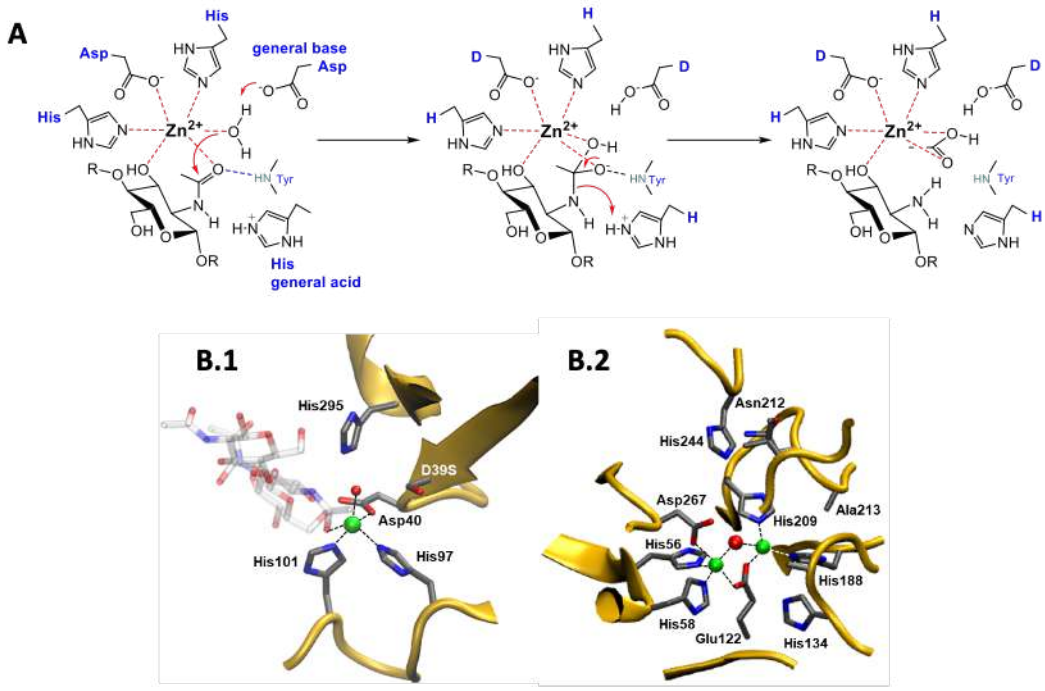


Figure 9.7. Catalytic mechanism of carbohydrate de-*N*-acetylases. (A) Metal-assisted acid/base mechanism of carbohydrate de-N- acetylases. Conserved Asp-His-His metal-binding triad, general acid (conserved His) and general base (conserved Asp). (B) Metal binding sites in de-N- acetylases, B.1) mononuclear metal coordination (VcCDA, PDB 4OUI, mutant D39S at the general base) and B.2) binuclear metal coordination (MsNagA, PDB 6FV3) found in some CE9 enzymes.

9.3.2. Deacetylases acting on structural polysaccharides: chitin de-*N*-acetylases.

The CE4 family is composed of not only chitin (CDA, EC 3.5.1.41) and chitooligosaccharide deacetylases (COD, EC 3.5.1.-), peptidoglycan *N*-acetylglucosamine deacetylases (PGN GlcNAc DA, EC3.5.1.104), peptidoglycan *N*-acetylmuramic acid deacetylases (PGN MurNAc DA, EC 3.5.1.-), and poly-β-1,6-*N*-acetylglucosamine deacetylases (PNAG DA, EC 3.5.1.-) but also some acetylxylan esterases that are de-O-acetylases (AXE, EC3.1.1.72).

CE4 enzymes share a conserved region known as NodB homology domain (for NodB COD involved in Nod factors biosynthesis [97]) of≃150 aa long and defined by five conserved motifs, which were proposed after the 3D structure resolution of SpPgdA in 2005 [98] and ClCDA in 2006 [99]. These motifs form the active site (Figure 9.8) and, as new 3D structures are released, their description is refined based on more extensive sequence and structural alignments (Figure 9.9) [100].

- **Motif 1 (MT1):** TFDDG, includes the general base aspartate (first D) and the aspartate that is part of the complex with the catalytic metal (second D).
- **Motif 2 (MT2):** H(S/T)XXHP, in which the two histidine residues are part of the metal-binding system and the serine or threonine residue is a hydrogen bond acceptor that stabilizes the structure of the loop-shaped motif in this region. These two histidine residues and the second aspartate of motif 1 form the characteristic His-His-Asp metal-binding triad of CE4 enzymes.
- **Motif 3 (MT3):** RXPY, that is forming one of the sides of the active site groove and provides stabilizing interactions in this area of the structure.
- **Motif 4 (MT4):** DXXD(W/Y), that is shaping the other side of the active site groove and includes a hydrophobic residue exposed to the solvent and an aspartate buried in the structure.
- **Motif 5 (MT5):** I(V/I)LXHD, includes the catalytic general acid histidine and a leucine that is in the hydrophobic patch that accommodates the acetate methyl group.

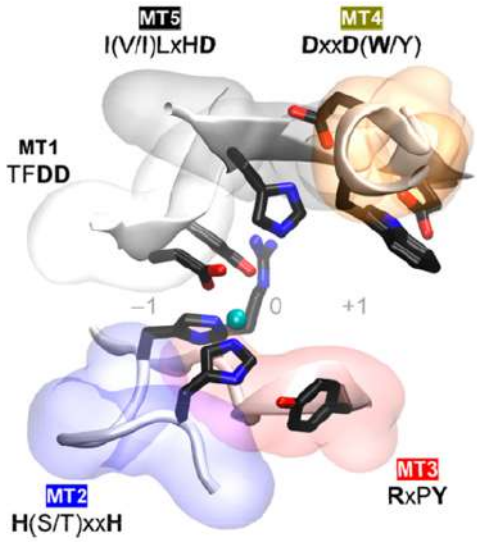


Figure 9.8. Conserved motifs MT1–MT5 defining the active site of CE4 enzymes. Motif 1 (TFDD) includes the general base aspartate (first D) and the metal-binding aspartate (second D) that, together with two His of motif 2 (H(S/T)xxH), form the characteristic Asp-His-His 3D structures of representative and novel chitin deacetylases (CDAs).

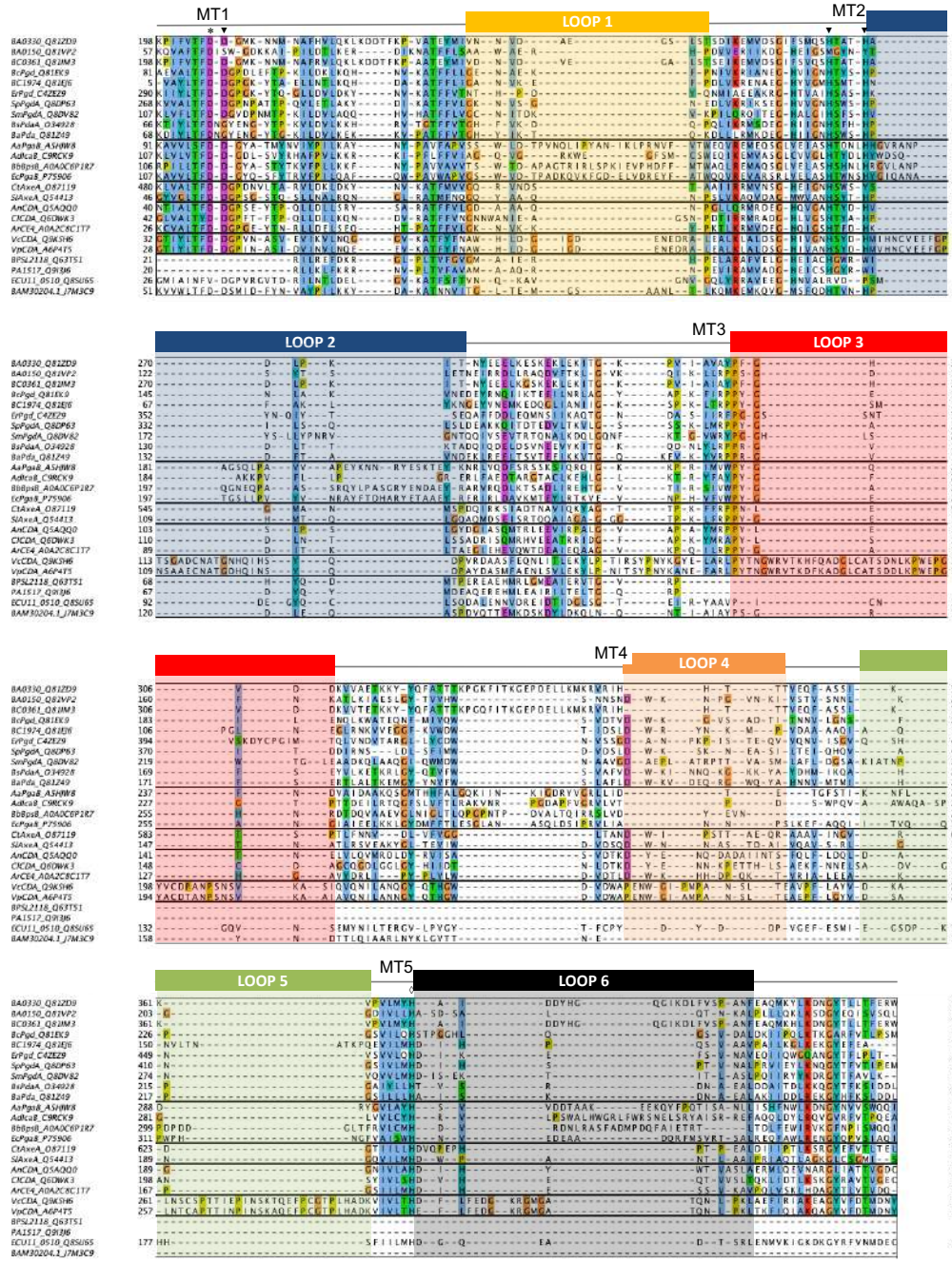


Figure 9.9. Multiple sequence alignment of CE4 enzymes. Loops are highlighted with coloured boxes according to [94]**. Conserved catalytic motifs are labelled MT1-5. "His-His-Asp" metal binding triad (▼), catalytic base (✱), and catalytic acid (⧫) are labelled. The alignment includes sequences of the enzymes, except for BmCDA1 and BmCDA8 from the insect Bombyx more, which were added later to the list and show some differential structural features to other CE4 members.**

Many CDAs and CODs have been characterized in archaea, marine bacteria, fungi, and insects [10,101]) and localized in different cellular compartments. Periplasmic fungal CDAs are generally tightly coupled to a chitin synthase to rapidly deacetylate newly synthesized chitin to chitosan. Extracellular fungal CDAs are secreted to alter the physicochemical properties of the cell wall, either for protection against exogenous chitinases and evasion of host's immune defenses or to initiate sporulation or autolysis. In bacteria, CDAs are either intracellular, as those involved in Nod factors biosynthesis in Rhizobium species, or extracellular, as those involved in the catabolism of chitin in marine bacteria [10]. In addition to numerous studies addressed to decipher the role CDAs play in the biology of their natural organisms, there is a current interest in the discovery and biochemical characterization of novel CDAs to use them as biocatalysts for the production of partially deacetylated chitooligosaccharides (paCOS) as bioactive molecules in different application fields, or to inhibit them since they are potential targets against pathogenic microorganisms [102].

Multiple CDAs have been characterized over the last years showing that they are capable of covering a very wide range of conditions for the production of paCOS. Its molecular weight varies from just 12 kDa to 150 kDa. Many of the CDAs in fungi are usually glycosylated. Their isoelectric points vary from 2.6 to 4.8. The optimal pH range varies from 4.5 to a value of 12, another remarkable characteristic is that there are surprisingly thermostable enzymes, their optimum activity being from 30ºC to 60ºC, although some of them can be subjected to 90ºC during certain periods of time conserving a considerable percentage of its activity.

CDAs exhibit diverse deacetylation patterns resulting from different action mechanisms (multiple-attack, multiple- or single-chain modes of action) and different pattern recognition of their substrates (Figure 9.10) [100].

A. *Multiple-attack or processive mechanism*: after binding the polysaccharide chain, the processive enzyme performs sequential deacetylations before dissociating and binding to another region of the chain [103].
B. *Multiple-chain or distributive* mechanism: the enzyme can form more than one active complex with the substrate and the deacetylation of only one residue of the polymeric chain takes place in each binding event.
C. *Single-site* mechanism: that covers the few enzymes with specificity for deacetylating a single position in their substrates [94].

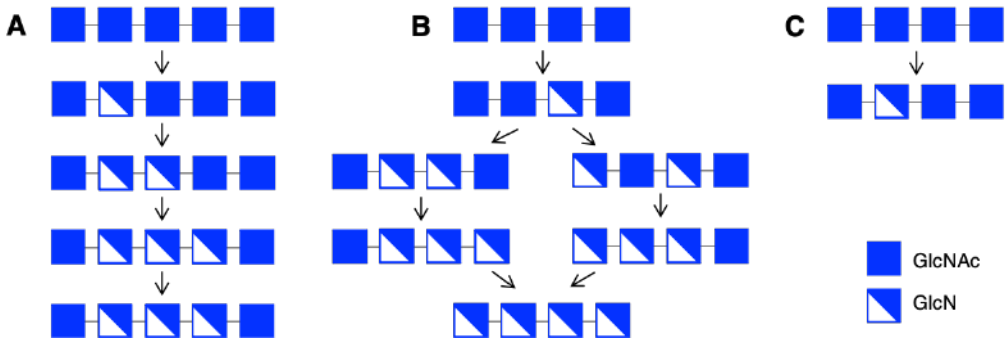
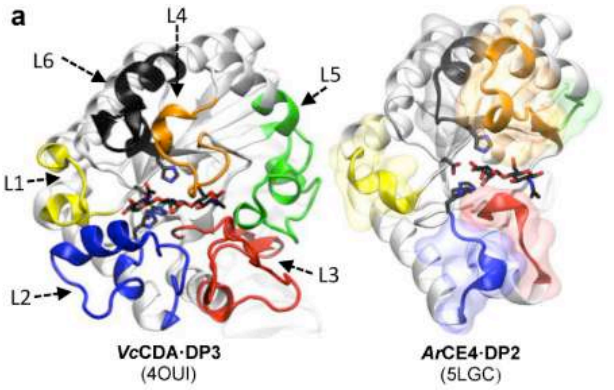


Figure 9.10. Modes of enzymatic action patterns for CE4 enzymes. A) Multiple-attack or processive mechanism; B) Multiple-chain or distributive mechanism and C) single-site mechanism.

A fundamental question is how these enzymes define the pattern of deacetylation. The X-ray structure of *Vc*CDA-Substrate complexes (PDB 4NZ1, 4OUI) highlighted the role of a series of loops that shape the binding cleft in substrate binding and led to the proposal of the "subsite capping model" (loop 1 to 6, Figure 9.11.A) [94]. The deacetylation patterns of characterized CDAs until 2018 were reviewed in Ref. [10] and in the previous PhD thesis on deacetylases in our group (Hugo Aragunde) [104]. Since then, new CDAs are being identified to unravel their biological functions and specificity on COS substrates, thus expanding the toolbox of biocatalysts for the preparation of paCOS with defined patterns of de-*N*-acetylation (Figure 9.11.B). Most CDAs also act in reverse mode following the same acetylation regioselectivity of $(GlcN)_n$ substrates than in the de-*N*-acetylation reaction. Significantly, the combination of different CDAs has allowed the enzymatic production of all partially acetylated chitosan tetramers [85].



b

CDAs from	PA
M. rouxi, C. lindemuthianum, A. nidulans, P. anserine, ***C. cinerae, N. aquimarinus***	D_n
P. graminis	AAD_{n-2}
Pestalodiopsis sp., ***C. neoformans* (CDA4)**	$AAD_{n-3}A$
P. chlamydosporia	$ADDA_{n-3}$
S. meliloti (NodB), ***Bombix mori* (CDA1,8)**	DA_{n-1}
V. cholera	ADA_{n-2}
V. alginolyticus	AD
S. woodyi, S. baltica	AD; $[ADA_{n-2}]$
Arthrobacter. sp	$D_{n-1}A$

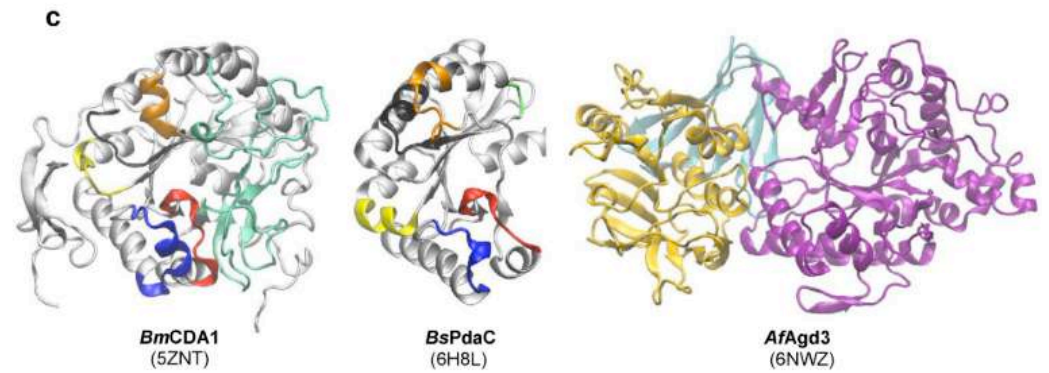


Figure 9.11. Different X-ray structures of carbohydrates de-*N*-acetylases and deacetylation patterns. (a) X-ray structures of Vibrio cholera CDA (*Vc*CDA, long loops, deep and narrow cleft) and Arthrobacter CDA (ArCE4, short loops, open and shallow cleft) in complex with chitooligosaccharide substrates. Loops 1 to 6 (color-labelled) illustrate their function according to the "subsite capping model". (b) Reported CDAs with known deacetylation pattern on chitooligosaccharides. PA: pattern of acetylation of end products from COS. A: GlcNAc; D: GlcN. In bold, characterized enzymes since 2018. (c) Relevant new 3D structures of CE enzymes since 2018 (left) insect CDA, loops 1 to 6 characteristic of CE4 enzymes labelled as in A), new loop insertion labelled in magenta, (middle) BsPdaC has dual MurNAc DA activity on peptidoglycan and GlcNAc DA activity on chitooligosaccharides (right), AfAgd3, GalNAc exopolysaccharide deacetylase, N-terminus domain (yellow) CBM87, C-terminus (pale green), founding member of family CE18.

Recently identified fungal CDAs include *Coprinopsis cinerea* Cda1-3 [105,106] and *Aspergillus fumigatus* Cod4 [107]. *Cryptococcus neoformans* contains four CDAs; three of them are GPI-anchored to the membrane acting on cell wall chitin during chitosan biosynthesis, but the fourth (CnCDA4) is a secreted enzyme with exceptional specificity for a GlcN residue at subsite -1. It has been proposed that it may further deacetylate exposed cell wall chitosans or released oligomers by human chitinases, thus making the fungus less susceptible to host immunosurveillance. The characterization of the secreted CDA4 from this pathogenic fungus showed that it is more active on partially deacetylated chitosan oligomers is involved in virulence. It is proposed that further deacetylation of cell wall chitosan or released oligomers by human chitinases is a strategy to make the fungus less susceptible to

host immunity [108]. CDA1-3 from the hypervirulent *Cryptococcus gattii* have shown that *C. gattii* has evolved an alternate regulation of chitosan biosynthesis during mammalian infection compared with *C. neoformans* [109]. CDAs have been identified for the first time in diatoms, *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*, which might be acquired by horizontal gene transfer from bacteria and fungi, respectively [110]. Several bacterial CDAs have been recently identified from *Actinobacter schindler* [111], *Bacillus licheniformis* [112] and *Rhodococcus equi* [113] (not yet characterized for their deacetylation pattern), whereas the *Nitratireductor aquimarinus* CDA was shown to act on COS with DP<4 and fully deacetylated DP4 [114]. Insect CDAs is an emerging field of study to unravel the biology of cuticle formation and peritrophic membrane modifications [115] and represent promising targets for insecticide development. Knock-out experiments have identified specific functions of CDA-encoding genes [116–118]. Different CDAs have been expressed and characterized from Bombyx mori [119,120], reporting the first 3D structures of insect CDAs (BmCDA1, 5ZNT; BmCDA8, 5ZNS) [119]. This article describes the first X-ray structures of insect CDAs, family CE4 deacetylases but with two unique loop insertions that define a much longer and wider substrate binding cleft than microbial CDAs. These new structures significantly contribute to further structure–function studies and to evaluate these CDAs as targets for insecticide development. Remarkably, they have two unique loop regions that contribute to the distinct architecture with a much longer and wider substrate-binding cleft than the microbial CDAs (Figure 9.11.).

9.3.3. Chitin de-*N*-acetylase from *Vibrio cholerae*

Vibrio cholerae is a gram-negative curved rod with facultative anaerobic metabolism. It is a bacteria with very high motile capacity due to the presence of polar flagellation. It can be isolated from aquatic and estuarine environments, although it is also a facultative human pathogen. *Vibrio cholerae* is the causative agent of the disease known as cholera. In the aquatic environment it is able to bind to the surfaces of plants, filamentous green algae, copepods, crustaceans and insects [121]. The chitin-oligosaccharide deacetylases from *Vibrio cholera* (also known as *Vc*COD or *Vc*CDA) has attracted the attention of many researchers due to its involvement in the chitin degradation pathway in the marine aquatic environment. A considerable number of species of the *Vibrio* genus have been identified: *V. algynolyticus* [122], *V. parahaemolyticus* [123], *V. cholerae* and others [124].

In the presence of chitin, a series of signals are transduced to regulate the gene expression of proteins related to cell binding to this substrate and its degradation [124]. *Vibrio cholerae* secretes chitinases that are responsible for reducing the size of this polymeric substrate to dimers of *N*-acetyl-D-glucosamine (chitobiose), which are rapidly deacetylated by *Vc*CDA in the second position from the non-reducing end. The resulting product (GlcNAc-GlcN) has two main destinations [125]: it can be incorporated into metabolic pathways as a source of N or C or it can act as a signal molecule to induce the expression of other enzymes related to the transformation and metabolization of chitin.

9.3.3.1. *Vc*CDA: Catalytic mechanism and structure

*Vc*CDA is a extracellular multidomain protein which can be found in various oligomeric forms. At the N-terminal end, it presents the typical deacetylase domain of CE4 enzymes followed by two chitin-binding domains at the C-terminus (Figure 9.12). Its monomeric sequence is composed of 431 aminoacids and the total molecular weight is 47.8 kDa with an isoelectric point of 5.41 according to Expasy-Protparam [126]. In the original sequence, *Vc*CDA has a 26 amino acid signal peptide.

*Vc*CDA was the first enzyme of the CE4 family to crystallize in a productive complex with its substrates. This milestone in the research in the chitin deacetylases family was carried out by the collaboration of the Laboratory of Biochemistry of the Institut Químic de Sarrià and the group of Dr. Marcelo Guerin (Biophysics Unit of the University of the Basque Country-CSIC, CICBiogune, Derio). Thanks to the resolution of enzyme-substrate complexes, it was possible to define their catalytic mechanism and propose a general hypotheses about the elements that dictate the substrate specificity and the deacetylation pattern of the different enzymes of the CE4 family [94].

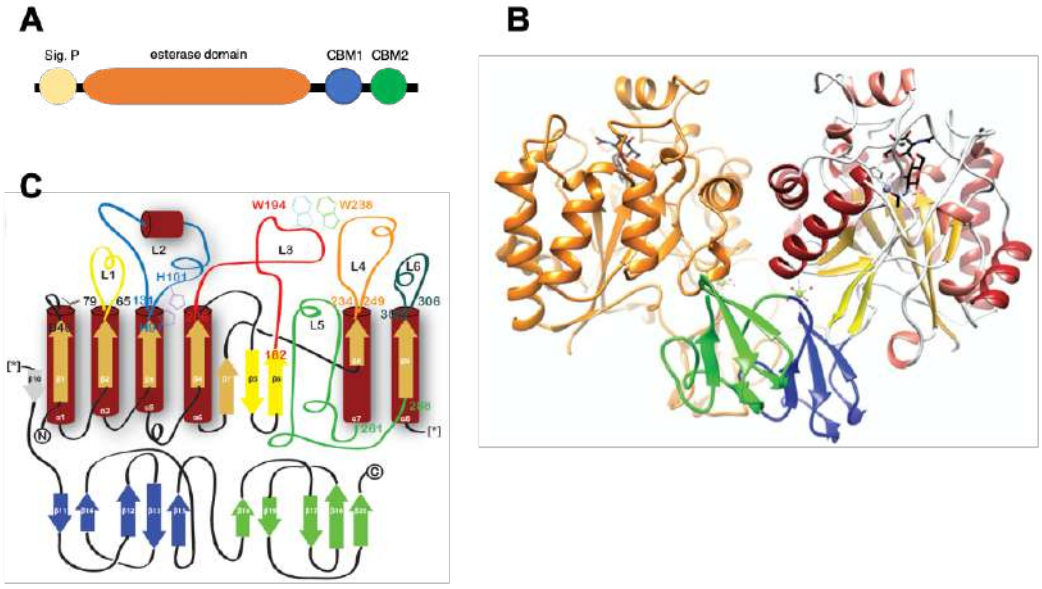


Figure 9.12. Structure of chitin de-*N*-acetylases from *vibrio cholerae*. (a) Modules present within the *Vc*CDA enzyme sequence. From the N end to the C terminal is the CE4 domain (26-338) and 2 CBMs of family 5 (336-382 and 385-431). (b) Schematic representation showing the general crystallized structure of *Vc*CDA as a homodimer. One monomer is represented in orange while the other domains that compose it are coloured in the other monomer: the deacetylase domain in red and yellow and the CBMs in blue and green. (c) Topological diagram of the *Vc*CDA monomer. The upper part shows the deacetylase domain while the lower part corresponds to the two CBMs.

The catalytic domain that ranges from residues 26 to 338 has the typical $(\beta/\alpha)_7$ barrel fold of the deacetylases of the CE4 family (Figure 9.12.B). The core of this structure is formed by a central cylinder defined by β sheets surrounded by α helices, forming a distorted β-barrel structure [94]. These secondary structures are connected to each other by a series of loops that will end up defining the active site where the substrate will bind.

In the active center of the enzyme there is a divalent cation identified as Zn^{+2}. This metal is retained due to its complexing interactions with a series of active site residues. This is known as the metal-binding HHD motif, in *Vc*CDA composed of the His97, His101 and Asp40. The enzyme operate by a two-step acid/base mechanism of action (mechanism presented in Figure 9.7.A).

The *Vc*CDA binding cleft is defined by a series of 6 loops. As will be explained in detail in the section defining the Subsite Capping Model, these structures are relevant at different levels: they establish interactions with the substrate and can define substrate specificity. Some of these loops are stabilized by disulfide bridges within the loop sequence itself. To identify them, they have been numbered from 1 to 6 (Figure 9.13):

- Loop 1: Residues from Gly71 to Arg79
- Loop 2: Residues from His97 to Asp131
- Loop 3: Residues from Asp182 to Ser206
- Loop 4: Residues from Ala234 to Glu249
- Loop 5: Residues from Leu261 to Asp288
- Loop 6: Residues of Glu300 to Met306

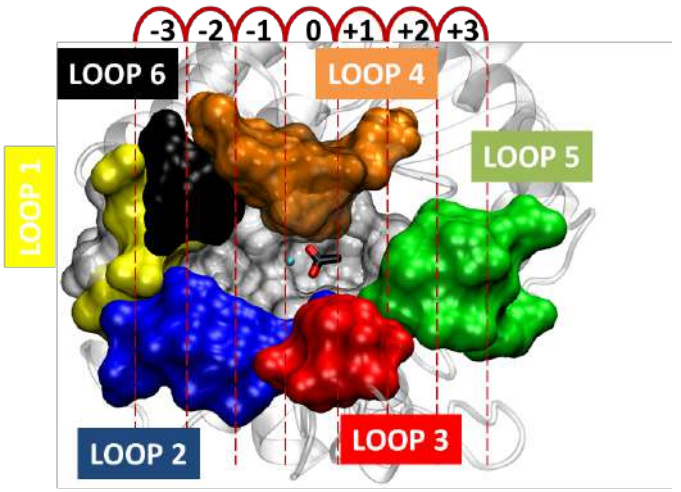


Figure 9.13. Loops of *Vc*CDA. Representation of the loops that decorate the periphery of the active site of *Vc*CDA.

9.3.3.2. Biochemical characterization of *Vc*CDA.

For its biochemical characterization, the *Vc*CDA full length (FL) was heterogously expressed in *E. coli* by cloning in a expression vector pET22b(+), which is an expression vector regulated by the pET system.

A Strep-Tag II has been included at the C-terminal end for purification by affinity chromatography. Five oligomeric fractions were obtained by size exclusion chromatography (SEC) purification, being the monomer the one that developed the highest deacetylase activity using the analog methylumbelliferyl acetate (AcOMU) as substrate. Furthermore, the molecular weight of each oligomeric fraction was verified by HPSEC and the monomeric fraction by MALDI-TOF MS (Figure 9.14).

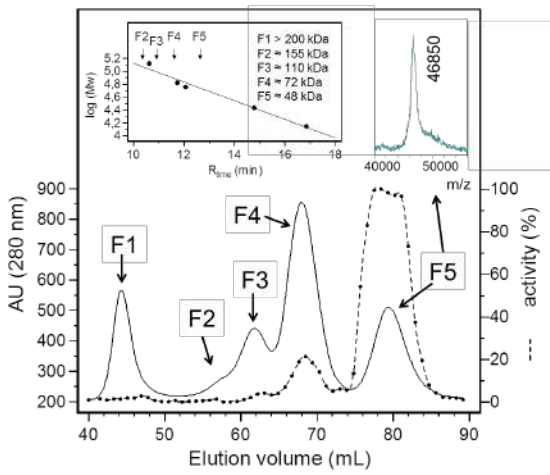


Figure 9.14. Chromatogram resulting from the preparative GFC of *Vc*CDA expressed in E. coli after being purified by affinity chromatography (Strep-Trap). - Absorbance at 280nm (···) enzymatic activity using AcOMU as substrate. Upper left graph: molecular weight of each of the fractions estimated by HPSEC; Upper right graph: MALDI-TOF MS analysis of fraction 5 (monomeric). [94]

The range of optimal conditions was determined. A study was carried out with different buffers to verify which was the optimum pH using two substrates: AcOMU and chitobiose (DP2). In both cases it was observed that the optimum pH is between 8 and 9 regardless of the buffer and substrate used (Figure 9.15).

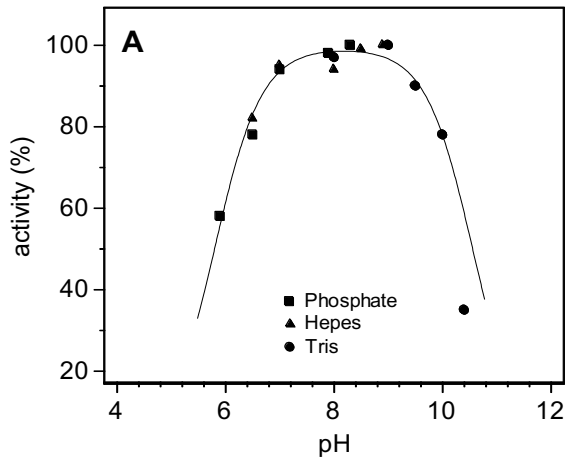


Figure 9.15. Determination of the optimal pH of *Vc*CDA. Dependence of the enzymatic activity of VcCDA as a function of pH using chitobiose (DP2 4 mM).

The kinetic parameters were characterized for various substrates from DP2 to DP4, since this enzyme is inactive with GlcNAc. DP2 had previously been described as the natural substrate for this enzyme [127].

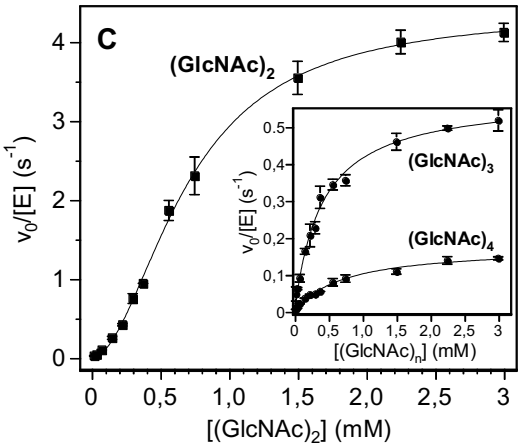


Figure 9.16. Characterization of the kinetic parameters using the monomeric fraction of *Vc*CDA FL and the chitooligosaccharides DP2 to DP4 in 50 mM phosphate buffer pH 8.5 NaCl 300 mM and at a temperature of 37ºC.

Once the kinetics parameters were obtained, it was observed that not all the substrates followed the same type of enzyme kinetics. While DP3 and DP4 followed a mechanism described by the Michaelis-Menten equation (Equation 9.1.A), DP2 showed an allosteric behavior which was adjusted to the Hill equation (Equation 9.1.B).

The value of Hill's coefficient h for DP2 shows a value close to 2, which roughly indicates that two substrate molecules are necessary for catalysis. This manifests itself in a kind of lag phase at low substrate concentrations when v0 / [E] is plotted against [S] (Figure 9.16).

$$a)\ {}^{V_0}/_{[E]} = \frac{k_{cat} \cdot [S]}{K_M + [S]} \qquad b)\ {}^{V_0}/_{[E]} = \frac{k_{cat} \cdot [S]^h}{(K_M)^h + [S]^h}$$

Equation 9.1. (a) Michaelis-Menten equation (b) Hill equation. Allosteric sigmoidal equation.

Based on the results obtained in the characterization of the enzyme, it was observed that its activity decreases as the DP of the substrate increases.

9.3.3.3. The structures of *Vc*CDA in complex with substrates. Induce-fit mechanism.

For crystallogenesis of *Vc*CDA structure soaking tests were performed using DP2, DP3 and DP4 as substrates [94]. Of all the conditions tested, crystal of the apo protein and a productive complex with DP2 and DP3 were obtained. Otherwise, it was not possible to get a crystal with the chitotetraose substrate in the active site.

These data provided first insights into structure-function relationships for this family of enzymes and highlighted the role of the 6 loops (already identified at a sequence level in the multiple alignment of Figure 9.9) that shape the binding site cleft in substrate binding.

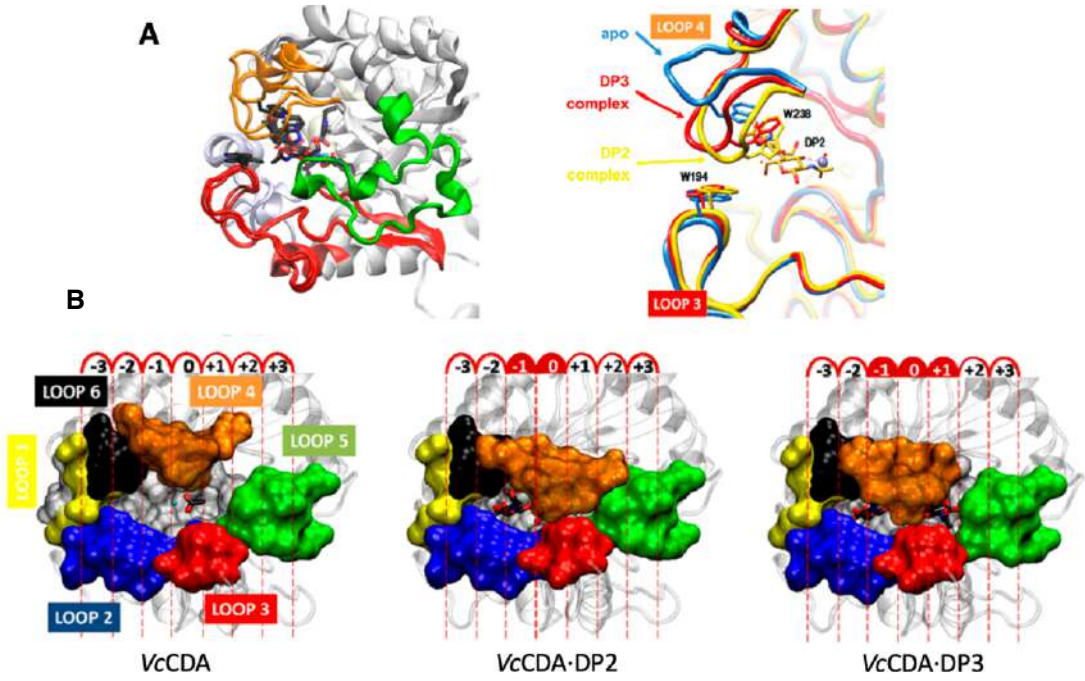


Figure 9.17. Crystallographic structures of VcCDA in the unligated form, in complex with $(GlcNAc)_2$ and in complex with (GlcNAc)3 [94]**. A) Superimposition of the three structures. Left: loop 4 (orange) has different conformations; right: magnification of the active site with loop 4 in the unligated form (blue), and in enzyme-substrate complexes with $(GlcNAc)_2$ (yellow) and $(GlcNAc)_3$ (red) ligands. Only DP2 ligand is shown. B) Overall structures. *Vc*CDA:unligated, *Vc*CDA-DP2:in complex with DP2, *Vc*CDA-DP3:in complex with DP3. Loops 1 to 6 are coloured as in Figure 9.13.**

The 3D structures obtained help us to defined the nomenclature for chitin de-*N*-acetylase of *Vibrio cholerae*. Subsite nomenclature is a generally accepted procedure for establishing monomer status in the case of active enzymes on polymers or oligomers. This nomenclature was initially proposed by Planas et al in 1995 in the CBM meeting and extended by G.J. Davies et al in 1997 for hydrolases [128]. Since this nomenclature was not applicable to enzymes that did not produce substrate hydrolysis, it was adapted by Tokuyasu et al in 2000 for its application to deacetylases [129]. The subsite 0 is the one in which the residue to be deacetylated is located, negative subsites the ones that are situated towards the non-reducing end and positive subsites the ones located to the reducing end of the bound substrate.

In the case of *Vc*CDA, the substrate binds into the active site in a manner that the second GlcNAc unit from the non-reducing end is located at subsite 0 where deacetylation occurs.

Moreover, negative subsites -2 and -3 are physically blocked by loops 1, 2 and 6 and, on the other site of the catalytic cavity, subsite +2 is also blocked by loop 5 (Figure 9.17).

In the absence of substrate the enzyme shows an open substrate binding cleft (Figure 9.17.B). When the substrate is bound, there is a movement of loop 4 upon substrate binding that changes the structure of the enzyme's active site from an open to a semi-closed or closed conformation depending on the length of the bound substrate. When chitobiose is bound, loop 4 adopts a closed conformation. This conformational change is interpreted as an "induced fit mechanism", the general meaning translates into the dynamic modulation of part of the structure of an enzyme due to substrate binding. This conformational change allow the creation of the perfect environment for catalysis to take place. Furthermore, this complex structure with DP2 allows us to define the subsites of the active site of the *Vc*CDA (Figure 9.17)

On the other hand, when the structure of the enzyme in complex with DP3 was analyzed, it stands out that loop 4 does not adopt a completely closed conformation. To accommodate a longer substrate, its conformation is semi-closed in such a way that the trisaccharide is able to bind in the subsites -1, 0 and +1.

Figure 9.17.A shows in detail the different conformations of loop 4 in the three structures described. In addition, the presence of tryptophan (W238 and W194) are capable of establishing stacking interactions with the pyranose of the substrate. W238 is of special relevance for the closing dynamics in the induced fit mechanism after substrate binding.

9.3.4. Determinants of substrate specificity: Subsite Capping Model.

9.3.4.1. Subsite Capping Model statement.

Once the importance of shape and movement of loops on substrate specificity within the *Vc*CDA structures was identified, the possibility that this phenomenon was not isolated and that it could be extrapolated to other deacetylases was considered. It is therefore we proposed and named this phenomenon as Subsite capping model [94].

The Subsite Capping Model was first proposed in our group as a hypothesis to explain why CDAs, as a subfamily of CE4 enzymes, present differences in their substrate preference and generate diverse patterns of deacetylation and specificities.

Based on Figure 9.18 and Figure 9.11., it is observed that the different CE4s studied present loops with highly differentiated three-dimensional structures. This fact was to be expected due to the great diversity in sequences that were found. Unifying these results with the fact that the rest of the structure of these proteins is highly conserved and, however, the PA of their products are different, it is proposed that these variable loops are the possible determinants of the substrate preference and the deacetylation mode.

If we go deeply in the comparation between the different available sequences and structures at the moment, it was observed that all enzymes shared a very conserved core, but the loops showed high variability in sequence, length and shape.

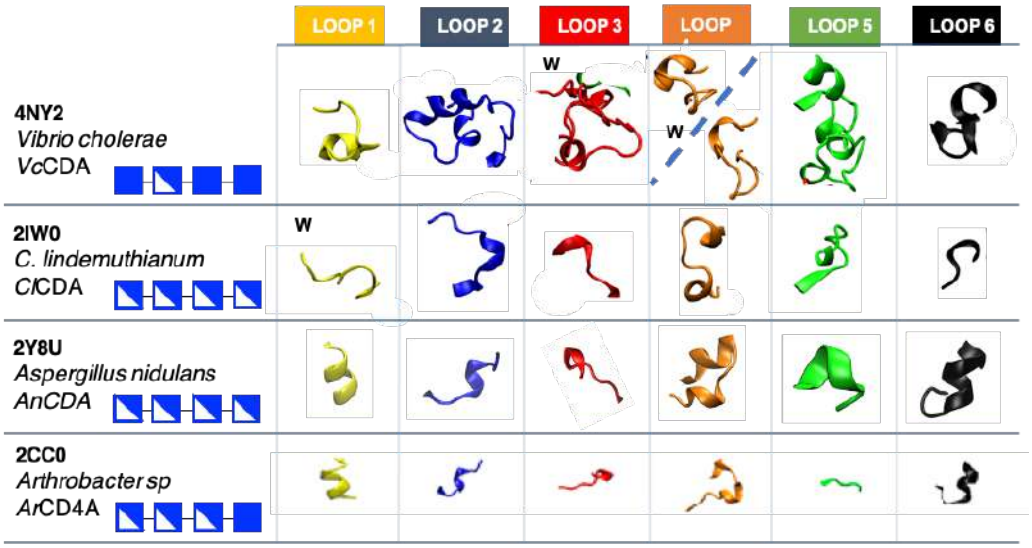


Figure 9.18. CDAs and CODs with known pattern of deacetylation on chitooligoshacarides.

The definition of the Subsite Capping Model was based on the following key aspects:

- The presence of potential subsites that are blocked by the structure of the loops that surround the active site.
- The loops are dynamic structures.
- The displacement of the loops allows the exposure of subsites necessary for the accommodation of certain substrates.
- Different deacetylases show important differences in loops sequence, size and structure, decorating differently the active site of each enzyme.

These four key features translate into the Subsite Capping Model statement: the pattern of deacetylation and substrate preference of CE4 enzymes are governed by specific loops and their dynamics.

Using this hypothesis, a solid base for deacetylase engineering could be generated in which modulating the final result of catalysis would be as simple as exchanging predefined loops. However, to reach this utopian situation, a great effort is necessary in demonstrating the universality and applicability of this theory.

9.3.4.2. Expanding the Substrate Capping Model.

Since 2014 when the subsite capping model was initially proposed, new 3D structures of CE4 enzymes have been solved and characterized, allowing a deeper study of the structure-function relationship of substrate specificity in this family of proteins.

The structural superposition of all available structures of CE4 enzymes active on COS confirmed the significance of loop topology as a determinant of substrate binding specificity. The characteristic α/β barrel is highly conserved, variability being provided by a series of loops surrounding the active site that connect the α/β elements of the barrel. The difference in length, structure and sequence composition of these six loops is what modulates the shape of the binding cleft exposed to the substrate in each particular enzyme. In general, loops 1, 2 and 6 define the negative subsites of the binding cleft and loops 3, 4 and 5 define the positive ones, being the catalytic site located at the center of these loops (subsite 0) that differentially block the accessibility to additional substrate binding subsites.

Currently, only two structures of CE4 enzymes in a productive complex with substrates are available. The *Vc*CDA complexes [94], and recently, the 3D structure of ArCEA, a CDA from an Arthrobacter sp [130]. These two structures are really useful for explaining the binding topology and the substrate specificity.

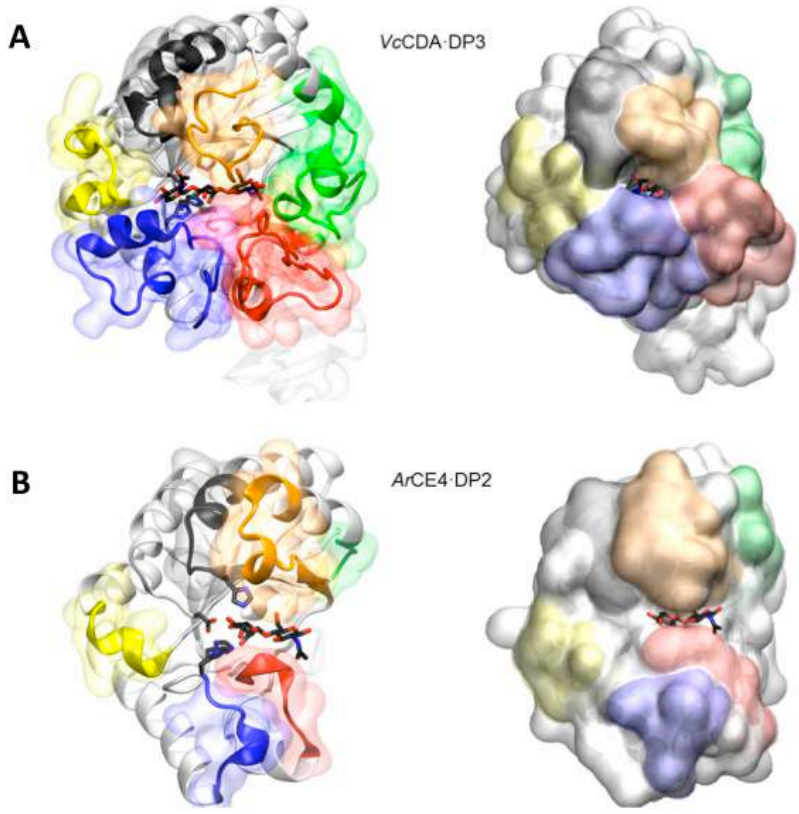


Figure 9.19. 3D structures of enzyme·substrate complex. (a) VcCDA with $(GlcNAc)_3$ substrate. (b) ArCE4 with $(GlcNAc)_2$ substrate. Loops 1 to 6 are colored as in Figure 9.13.

For long loops and high specificity, *Vc*CDA is a great example. As already stated, the negative subsites at the non-reducing end of the substrate are defined by loops 1, 2 and 6, while the positive subsites at the reducing end of the substrate are defined by the loops 3, 4 and 5. Loops 4 and 5 are quite mobile allowing the accommodation of long substrates such DP4 or DP5. On the other hand, the loops that define the negative subsites present an extensive network of interactions which very efficiently block this region, preventing any other PA from being obtained in the deacetylation products. These loops are capping both the reducing and non-reducing ends of the substrate, and this constricting topology has two main consequences. In the first place, the enzyme is highly active on chitobiose and its catalytic efficiency drops substantially on longer oligomers such as chitotetraose or chitopentaose. Secondly, the enzyme only deacetylate the second unit from the non-reducing end owing to there is no space for the substrate to slide along the active site towards new binding

conformation, thus only one GlcNAc can be accommodated at the catalytic site (subsite 0) where deacetylation takes place (Figure 9.19.A).

On the other hand, the CE4 enzyme from *Arthrobacter sp*. is a deacetylase with short loops and broad specificity. ArCE4 structure in complex with chitobiose shows the substrate bound to a much more open binding cleft. Although the enzyme was co-crystalized with $(GlcNAc)_4$, only two units are observed in the structure indicating a weak interaction of part of the oligomeric chain with the flat protein surface topology. The main difference between the two model enzymes is the size and shape of the loops surrounding the active site, having both the catalytic center in the same site with respect to the protein core. For ArCE4's enzyme, the short loops result in a significantly open binding cleft that can accommodate longer substrates (Figure 9.19.B). Furthermore, the lack of protein caps at either ends of the substrates allows different binding modes, exposing different GlcNAc units of the oligomeric chain at the catalytic site. These observations are in agreement with the enzyme's properties, showing that the enzymatic activity of ArCE4 increases with the length of the substrates and that it follows a multiple-chain mechanism fully deacetylating the COS substrates except for the reducing-end units.

Gaining further structural information of enzyme-substrate complexes of CE4 enzymes will be crucial to the design or discovery of novel CDAs with controlled specificities in a rational way. It could be possible to modulate the final result of the catalysis towards new specificities, for example, exchanging pre-defined loops, but it is necessary to demonstrate the universality and applicability of this hypothesis.

9.4. Protein engineering: Rational design and Directed evolution

The application of enzymes in industrial processes and other fields often require properties that are not commonly found in systems isolated from their natural environment. Scientists have been always looking at nature in order to find a source of inspiration to mimic the great power and efficiency of its processes. One of these astonishing processes is the catalytic properties of the enzymes; very complex molecular machines capable of enhancing the rate constant of chemical reactions [131,132]. Without the need for human intervention, enzymes have already undergone an exhaustive optimization process: Natural selection. Natural selection is the process through which populations of living organisms change and adapt as a key factor in organism evolution, but the particular mechanism, by which it is carried out is not well understood in the macroscopic world of phenotype. It is widely accepted that many enzymes evolved from pre-existing enzymes via gene duplication. Notwithstanding, protein engineers look for proteins with high stability under the conditions in which the transformation process is going to take place and a high catalytic activity for natural and non-natural substrates towards new specificities [133].

The defining characteristic of protein engineering is the end product, a protein with a new (or modified) sequence and not a living organism. Since proteins cannot reproduce, many of the concerns and biases generated by genetically modified microorganisms are not applicable in protein engineering. As regards to the latter, protein engineering is looking for new chemical compounds from non-biological sources but which, by their very nature, are easily biodegradable (although other safety and toxicity aspects must be taken into account) [134].

Protein engineering is a powerful tool to obtain proteins and enzymes with a desired trait, like an increased activity, higher tolerance against stressors or higher protein yield. Three major strategies can be applied for protein engineering: knowledge-based mutagenesis or rational design (RM), computational protein design (CPD) and directed evolution (DE).

9.4.1. Rational design of enzymes

The rational design is one of the two basic enzyme optimization protocols, the first to be applied was used to modify the specificity of certain enzymes [135]. It is based on the performance of specific mutants under the hypothesis that these mutations are the ones that will determine an improvement in their properties or a new activity that is being look. However,

to carry out this type of strategy, an in-depth knowledge of the structure, the catalytic mechanism, its regulation, and the similarities with other related enzymes is required. For instance, early site-directed mutagenesis studies of trypsin, to attain chymotrypsin like activity, did not achieve the desired result. The active-site mutation D189K was initially thought to be important in switching the selectivity but further work showed that mutations at surface residues were also required. This exemplifies the difficulties of modifying natural enzymes even where there is detailed sequence and structural knowledge of a protein. Nowadays, with the use of bioinformatics tools and the increase in their calculation capacity, has allowed the generation of increasingly precise hypotheses [136,137].

In general, these types of strategies require a high planning effort and the results are not always as expected. In cases where the relationship between structure and function clearly indicates the modifications to be made, the results are exceptional. Even within the rational design, procedures based on the coupling of enzymatic modules can be considered without altering their internal sequences in order to combine their properties and thus generate new structural organizations [138]. However, when the search for a new function is not extracted in a trivial way from the observation of the structure, the results are not always satisfactory. This is because, despite the accumulated knowledge, there are a large number of parameters that are not yet controllable, such as protein folding, establishment of stabilizing substrate interactions, and protein expression problems.

9.4.2. Directed evolution of enzymes

Directed evolution mimics natural evolution in the lab, in an accelerated fashion, and is used to modify proteins at will, to endow them with more desirable properties. In this procedure, a detailed knowledge of the protein to be worked is not necessary; although it is not a requirement, it can facilitate the design of the experiments [139]. Directed evolution tries to reproduce what happens spontaneously with natural evolution with living organisms in the environment, but in this case in the laboratory and with a much shorter time scale. The classical approach to obtaining new enzymes is based on the search in new isolated organisms, thus, the diversity of these enzymes is naturally generated through natural selection as a force for change. In directed evolution, diversity is created artificially by genetic engineering methods and selection is obtained by introducing an artificial selection pressure. Directed evolution campaigns are based on iterative steps that are repeated until

the desired variant is obtained. This process will make impossible to obtain mutants whose properties are adequate to carry out a specific activity as a consequence of the evolutionary forces introduced by the researcher [140,141].

The process of engineering a protein is complicated. Natural proteins are biologically complex, they are the product of millions of years of evolution and with that comes significant complexity that can be difficult to understand through chemical knowledge alone. Through time and the pressures of evolution, natural proteins have traversed an evolutionary landscape that does not focus on one aspect of the design of an enzyme. Secondary function, operability under a certain physiological condition and control over the flux of substrate to product are all selected for. Proteins found in nature are therefore evolutionarily complicated and difficult to modify.

In any process of directed evolution, it is necessary to take into account three basic elements (Figure 9.20):

a) Selection of the starting protein or protein family.
b) Selection of a variability generation method.
c) Selection or screening method.

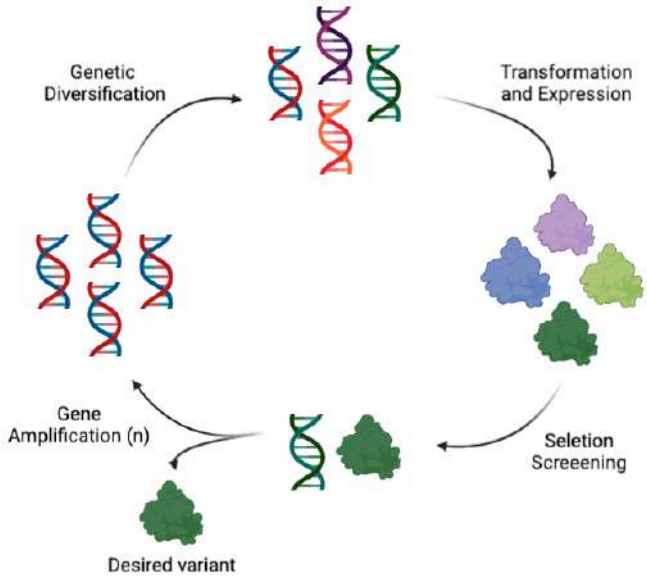


Figure 9.20. General steps of a directed evolution protocol. The starting gene encoding the protein that will be evolved is subjected to a gene diversification strategy; the quality and size of the resulting library will depend on the criteria chosen for its construction. The individual variants of the genetic library are expressed in an appropriate host or by cell-free *in vitro* transcription and translation (IVTT). Using an appropriate assay, a screening will then be implemented to identify one or more variants of interest, typically showing a desired activity, thermostability, or selectivity. These variants can be subjected to a

new round of directed evolution, until the improvements are satisfactory enough to enable the biocatalytic process.

The selection of the starting protein or family of proteins is the first relevant element. If only one enzyme is selected and we want to improve its properties, the variability generation techniques will be primarily those that introduce mutations along the gene. On the other hand, if a set of related enzymes needs to be combined, the variability generation methods are essentially based on the generation of chimeras by recombination or by DNA shuffling [142,143].

The second key element is how this variability is introduced. In nature, variability is generated due to errors introduced by polymerases in the DNA replication stage. Because polymerases present extremely low error rates (e.g. 1 base per billion replicated bases [144], the appearance of mutants is sporadic, also prevented by DNA repair and correction mechanisms, hence natural evolution requires such a long time scale. At laboratory scale, how to generate the variability in libraries of mutants is an essential step. Protein libraries obtained from *in vitro* mutagenesis are widely used for directed evolution. In general three different approaches can be applied: random mutagenesis, homologous recombination or focused mutagenesis (Figure 9.22). Multiple protocols have emerged from each of these three strategies to obtain mutants or chimeras. The selection of one protocol or another within the different varieties is not a simple decision, this step will dictate the size of the mutant library to be tested. In most cases, the libraries are located between 10^3-10^6 mutants [145], but even so the sequence space of a protein is not covered. Sequence space is the total of possible amino acid combinations for a given protein size. However, it is also not interesting to cover the entire sequence space, since only those proteins with a high degree of similarity to the original may present an activity similar to the starting one (Figure 9.21) [146,147].

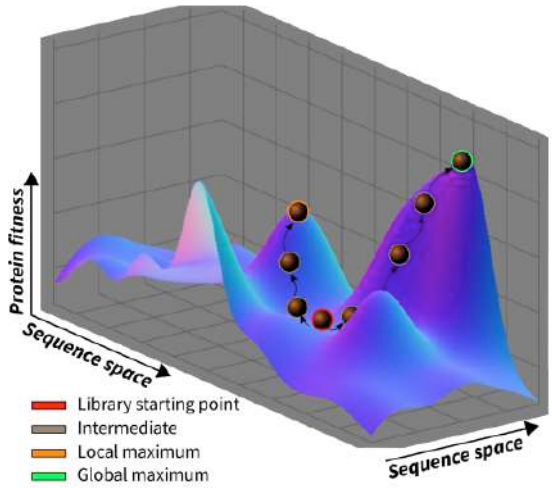


Figure 9.21. Relationship between catalytic promiscuity and the evolution of a new function. The aim of a DE experiment is to traverse a fitness landscape towards a maximum; ideally the global maximum. Here, the library starting point is represented by the red ball, intermediates by grey balls and local and global maximums represented by amber and green balls respectively. Adapted from [148].

Random mutagenesis is normally conducted by error prone PCR (epPCR). It exploits the inherently high error rate of Taq DNA polymerase, which can be further increased by suboptimal PCR conditions. The resulting PCR products contain a predefined number of mutations, which can be controlled by the amount of Mg^{2+} and Mn^{2+} ions or an unequal dNTP concentration [149]. Library generation using epPCR is very time efficient and can be conducted without prior information about the target, except for its DNA sequence. However, the library size increases rapidly with the target length, which requires an efficient screening or selection method.

A different approach is homologous recombination, which was first described by Stemmer as DNA shuffling [150,151]. A DNase cleaves a set of homologous target genes and the fragments are reassembled during a PCR reaction without any primers, where opposing strands bind at homologous regions. The resulting fragments resemble a full length gene composed of different homologs. As stated by Stemmer, this method can be more efficient in terms of protein optimization compared to epPCR, since existing beneficial mutations are combined. The generation process is rather simple and can be conducted without prior structural knowledge. However, phylogenetic information is necessary to identify suitable gene homologs, whereby distant homologs cannot be used [152].

Contrary to random mutagenesis and homologous recombination, focused mutagenesis does not target the whole protein but one or a few previously selected residues. Mutations are

incorporated by synthetic DNA oligonucleotides, containing one or more degenerated codons. The most commonly used codons are NNS or NNK, covering all possible amino acids, while only one stop codon is introduced (Table 9.2). Since some amino acids such as serine are encoded by six codons, whereas others such as methionine are only encoded by one codon, a strong bias is introduced by degenerated oligonucleotides. This can be avoided by adding trinucleotide phosphoramidites, one coding for each amino acid. The resulting library is easy to generate and allows for a very detailed analysis of the substituted residues. However, only a small number of residues can be targeted like this, until library generation gets very labor-intensive [153].

Table 9.2. IUPAC nucleotide code.

IUPAC code	Base
A	Adenine
C	Cytosine
G	Guanine
T (or U)	Thymine (or Uracil)
R	A or G
Y	C or T
S	G or C
W	A or T
K	G or T
M	A or C
B	C or G or T
D	A or G or T
H	A or C or T
V	A or C or G
N	any base

Additional libraries which are available resemble a combination of the previously described ones. For example, random mutagenesis can be applied on small sections of the protein to reduce the library size, allowing to screen a larger proportion of the library. Furthermore, if several rounds of mutagenesis and screening or selection are conducted, different methods such as homologous recombination and random mutagenesis can be combined.

Other methods like alanine scanning are difficult to categorize. Here, a small library is created by site-directed mutagenesis, where single alanine mutations are introduced for every amino acid. Alanine is known to be present in α-helices and β-sheets, not disturbing the backbone

conformations and it does not impose significant electrostatic or steric effects. Thus, solely the effect of the missing side-chain is analyzed in subsequent activity or binding assays [154].

Library of mutants or recombination libraries need to be assembled in a plasmid for further protein expression. Most common techniques for assembly are Golden Gate, Gibson-base, CPEC cloning or QuickLib [155,156]. After assembly, the library must be integrated into an expression host such as bacterial, yeast or cell-free host for protein library expression [157,158]. While high expression is usually desirable for maximizing screening sensitivity, the associated burden on the host may generate important fitness costs and genetic instability. Enzyme variants are classically expressed in the cytoplasm, and subsequent activity screening is then performed with cell-penetrable substrates or after cell lysis.

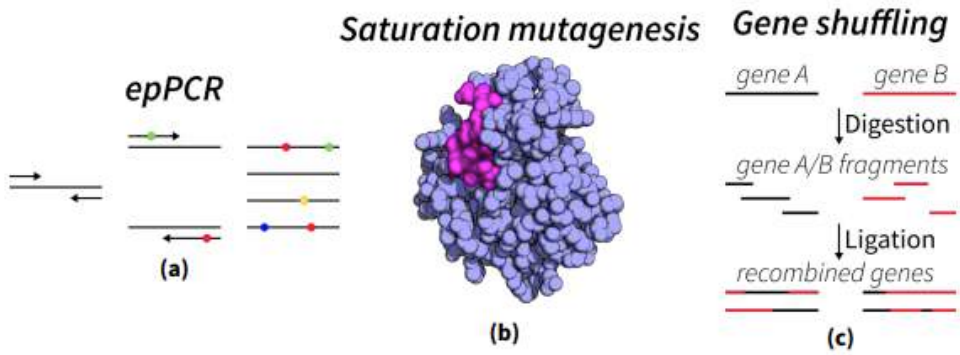


Figure 9.22. Three commonly used library generation techniques are illustrated. (a). An epPCR schematic is shown where a gene of interest is amplified by a low-fidelity polymerase producing errors in amplification (coloured circles) in the final product. (b). A protein region of interest is designated, in cyan, for optimisation via saturation mutagenesis. (c). Two phenotypically equivalent genes are identified for a gene shuffling experiment. This involves restriction and ligation to create a library random homologous genes.

The last determining element in directed evolution is the method that differentiates between enhanced mutants and those that are not. There are two basic types of assays: selection and screening [141,159]. In selection, only those system hosts that contain a protein with enhanced function are able to survive. Selection is the easiest methodology to interpret, but their application is very limited, e.g. resistance to some type of toxic compound. In screening assay, the activity of the enzyme can be measured and a value can be assigned for each mutant of the library. Screening assays are of pivotal importance, not only because their robustness is one of the major predictors of success but also because the throughput rate determines the maximum library size that can be handled. In this case, it is not about survival or non-survival, it is about having an objective tool for selecting mutants that exceed/fall short

a level of activity greater/lower than the threshold determined by the WT protein. This stage also conditions the size of the library since, for example, if the substrates are difficult to obtain, the activity measurement protocol is very elaborate or there is no possibility of automating the experiment, the hypothetical size of the library is reduced against the opposite conditions (accessible and cheap substrate, simple test and high automation capacity).

Two major strategies can be distinguished for library screening. In screens of spatially separated variants, all mutants are singulated and analyzed individually, whereas methods such as fluorescence-activated cell sorting (FACS) can be applied directly to a bulk population. This latter requires fluorescent phenotype or marker, which is automatically sorted. Mutants, that enable a higher fluorescence or binding of a fluorescent marker are singled out and analyzed further.

Screening for a desirable phenotype, in most cases, is a blind process, it is highly desirable to develop a high-throughput screening (HTS) methodologies to facilitate screening of extremely large libraries. These methodologies are particularly valuable for enrichment and isolation of rare mutants with beneficial activity from large number of mutants. In recent years there has been tremendous interest in the development of enzyme assays in connection with the high-throughput screening of enzymes for use in biocatalysis and drug discovery [141,160,161]. In this respect, both new reaction types and new assay concepts have been reported. Macroscopic screening formats such as colonies on Petri dishes or bacterial culture in 96-well plates are relatively easy to handle, miniaturization in wells/cell-free droplets, oil-in-water microdroplets or with single cells may greatly increase the assay throughput and reduce the costs. Most of these high-throughput assays are based on chromogenic and fluorogenic substrates or sensors. One of the emerging technologies in HTS development for directed evolution of enzymes is the use of fluorescence-activated cell sorting (FACS) to allow the isolation of enzyme variants with significantly improved activities, altered substrate specificities, or even novel functions [162].

In general, enzymatic activities can be determined using direct assays with the real substrates and aim at directly monitoring the formation of desired product, although direct screening based on common analytical techniques have a comparably low throughput, remarkable achievements using HPLC- and GC-based assays have been reported [163]. On the other hand, indirect methods based in a signal developed as a consequence of enzyme turnover (e.g. pH change, release a fluorophore) may require a coupled reaction or the use of a

surrogate substrate. Indirect assays cover a deep range of throughput values and not always require a physical separation between variants. But, it is still not uncommon to use a mixed approach to integrate direct and indirect assays, especially when enantioselectivity is the target for evolution. For enzymes catalyzing reversible reactions, screening assays may also be designed to monitor the reverse reaction relative to the final objective.

OBJETIVES

10. OBJECTIVES

TOPIC 1: Engineering substrate specificity of chitin deacetylases

Regarding CE4 family, the main objective of the project is to redesign the chitin deacetylase of *Vibrio cholerae* to increase the activity towards longer substrates using directed evolution and rational design approaches.

The second main objective is to understand the molecular basis of loops dynamics in *Vc*CDA enzyme which dictate the enzyme specificity and permit us to experimentally demonstrate the subsite capping model approach.

Finally, the last goal is the attempt to engineer and redesign the *Vc*CDA enzyme using a semirational approach towards a new deacetylation pattern.

This knowledge will contribute to exploit chitin deacetylases both as biocatalysts and as therapeutic targets against fungal infections. In this context, the specific goals are:

CHAPTER 1. **Biochemical characterization of chitin deacetylase from *Vibrio Cholerae***

1. Biochemical characterization of *Vc*CDA in terms of protein expression and purification, kinetic characterization by HPLC-MS and protein thermostability.

CHAPTER 2. **Screening assay for directed evolution of chitin deacetylases.**

1. Implemented a convenient and sensitive HTS assay to screen directed evolution libraries of CDAs and other CE4 enzymes active on COS.
2. Address a directed evolution approach using HTS assay to generate *Vc*CDA mutants with higher activity towards longer substrates.
3. Biochemical characterization of selected mutants.

CHAPTER 3. **Engineering loop flexibility to tune enzyme specificity in *Vc*CDA.**

1. Demonstrate the functional mechanism of the Subsite Capping Model using two strategies: Loop 5 blocking and loop 5 flexibilization.
2. Elucidate the dynamics of loop 5 using NMR techniques.

CHAPTER 4. **Engineering *Vc*CDA enzyme towards new deacetylation patterns**

1. Development and validation of a HTS rapid screening assay for deacetylase activity.
2. Loop 1 engineering by screening a focused full library.

CHAPTER 1. BIOCHEMICAL CHARACTERIZATION OF *Vc*CDA

11. CHAPTER 1. Biochemical Characterization of *Vc*CDA

11.1. Expression and purification of VcCDA FL wt

For the expression of the *Vc*CDA enzyme in *E.coli*, it is necessary to clone its encoding gene into a suitable protein expression vector. As already stated in section 9.3.3.2, the gene coding for the complete protein *Vc*CDA FL (full length) was previously cloned in the laboratory in a vector pET22b. The plasmid must present the genetic elements necessary to regulate or induce its production. The high copy number plasmid pET22b(+) was chosen, which allows the fast production of a high quantity of protein once induced. Since CE4 enzymes contain metal ions in their catalytic center, the possibility of the presence of His-Tag negatively influencing the enzyme's activity was considered. For this reason, the peptide Strep-Tag II was included in the C-terminus end of the enzyme, followed by a stop codon. This additional peptide allows protein purification by affinity chromatography without the presence of metal ions. Furthermore, the amino acid sequence of Strep-Tag II (WSHPQFEK) is rare in nature and allows obtaining higher purity levels in comparison to the original His-tag. Once was cloned in a expression vector suitable for *E.coli*, the desired enzyme could be expressed.

The expression of this enzyme in the described system had already been successfully carried out in previous works in the laboratory [94]. For the expression, a protein autoinduction protocol was used due to its convenience and the high yields achieved with this enzyme. In the autoinduction media, a base of LB medium can be used in which a series of additives such as salts and sugars are added. Mainly three sources of carbon are included: glucose, glycerol and lactose. In the early stages of cell growth, the use of glucose as the main carbon source allows protein expression to be inhibited and all cell metabolism to be involved in multiplication. When this sugar is exhausted, the other carbon sources are used. Since lactose is an inducer of the pET system, the expression of the protein is initiated [164]. With this method (contrary to the IPTG induction strategy) is not necessary to add an inducer to the culture at a certain point since the growth of the expression strain is performed in a medium that promotes the induction of the T7 promoter, leading to the expression of the protein of interest.

The protocol for the expression of this protein starts with a culture of a given volume in an Erlenmeyer flask. This culture is incubated at 25ºC for 48 hours. After this step, the culture is

centrifuged and lysed using a sonicator. Insoluble cell debris resulting from lysis are removed by centrifugation and filtration (see MATERIALS AND METHODS, protocol 21.4.1).

The first purification step is affinity chromatography using Strep-Trap column affinity chromatography (GE-healthcare). The basis of this process is based on the biotin-streptavidin affinity, this complex has one of the strongest non-covalent interactions known. The strep-tag is an 8 amino acid sequence which was selected to specifically bind to streptactin which is an optimized version of streptavidin. The steps of this technique are exemplified in Figure 11.1. In this type of chromatography, the clarified cell lysate is loaded into the column, in this phase the recombinant protein with the Strep-Tag II sequence is retained through its affinity to streptotactin. The rest of the unbound proteins are washed by using the buffer in which the lysate is found. Once the column is washed, the elution of the protein is carried out using a desthiobiotin solution. This compound has a higher affinity for strep-tactin than the recombinant protein, making it to elute from the column. Finally the column is regenerated with 2-hydroxy-4-aminobenzoic acid (HABA). This compound has the objective of displacing the desthiobiotin; however, it has lower affinity for the stationary phase, but its large excess manages to displace the desthiobiotin. With a second wash with an excess of buffer, the HABA is eliminated and the column is ready for the next loading and elution steps (Figure 11.1).

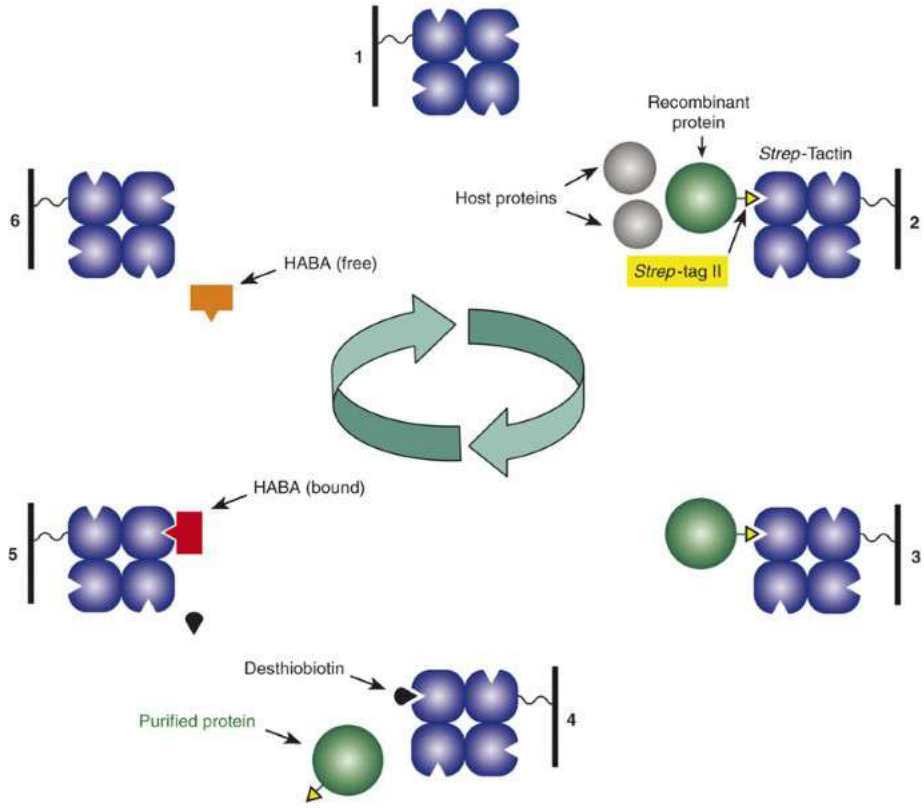


Figure 11.1. Strep-Tag II affinity purification process. Outline of the Strep-Trap affinity chromatography protocol: 2. The protein with Strep-Tag II is retained by its affinity with Strep-tactin, 3. *E. coli* proteins are washed with buffer, 4. Desthiobiotin displaces recombinant enzyme and eluted protein of interest, 4. HABA displaces desthiobiotin to regenerate column, 5. HABA is washed away and column regenerated.

In the purification of *Vc*CDA FL (Full length), several loading and elution steps of the protein were carried out (see protocol 21.4.4). The previously described process is applied several times in series on the same column dividing the culture lysate into several volumes (Figure 11.2). A common Strep-tag column purification of *Vc*CDA FL wt is shown in Figure 11.2.

All *Vc*CDA proteins expressed in this work shows several oligomeric fractions and the monomeric fraction is the one with the highest activity (see section 9.3.3.2.). To isolate the different oligomeric states, a preparative size exclusion chromatography was carried out in Superdex 200 (16/600) column. Size exclusion chromatography (SEC) or gel filtration is a chromatographic technique used to separate molecules by their molecular weight or size while they pass through the column packed gel filtration medium. In contrast to affinity chromatography, molecules do not bind to this medium and the buffer composition does not directly affect the resolution (separation between peaks). Gel filtration medium is a porous matrix in the form of spherical particles that are chosen for their physical and chemical stability

and because they are inert (without adsorption properties or reactivity). The medium is packed inside a column and it is equilibrated with a buffer that fills the matrix pores as well as the space between the particles. The liquid inside the pores is referred as stationary phase and it is in equilibrium with the liquid outside the particles (mobile phase). In this environment, the larger molecules will not be retained in the matrix particle's pores and their retention time in the column will be shorter.

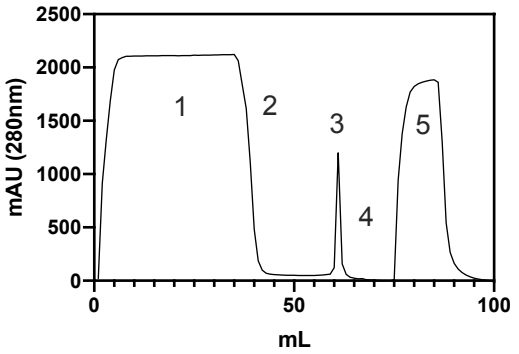


Figure 11.2. Strep-Trap purification chromatogram of *Vc*CDA FL Chromatogram resulting from affinity chromatography purification of *Vc*CDA in which three column loads were performed. The numbers indicate the stages of a charge / elution cycle: 1. Charge, 2. Wash, 3. Elution, 4. Wash, 5. Column Regeneration.

All pure protein fractions eluted from the affinity column are pooled into a single volume and concentrated to a maximum of 2 mL volume to be loaded onto the gel filtration column. Elution was carried out at a flow of 1 mL/min and using the 50 mM phosphate buffer pH 8.5 and 300 mM NaCl, the same buffer that was used in the previous steps.

Figure 11.3 shows the chromatogram obtained for the *Vc*CDA FL where the monomeric fraction elutes in a volume of approximately 80 mL (fraction F5). It has been observed throughout the work that the relative amount of monomeric fraction compared to the other fractions corresponding to oligomers of the protein varies between batches. The parameters that affect this phenomenon are unclear. Using the column calibration equation, the value obtained for this enzyme is 41.7 kDa, while the theoretical Mw is 46.9 kDa. This variation may be due to the three-dimensional organization of the enzyme that makes it migrate in a different way depending on the hydrodynamics radio. The hydrodynamic properties of macromolecules in solution has been studied for the determination of the conformation or shape of proteins.

Using samples from the different purification steps, an SDS-PAGE electrophoresis gel was made to check the purity of the enzyme (Figure 11.4).

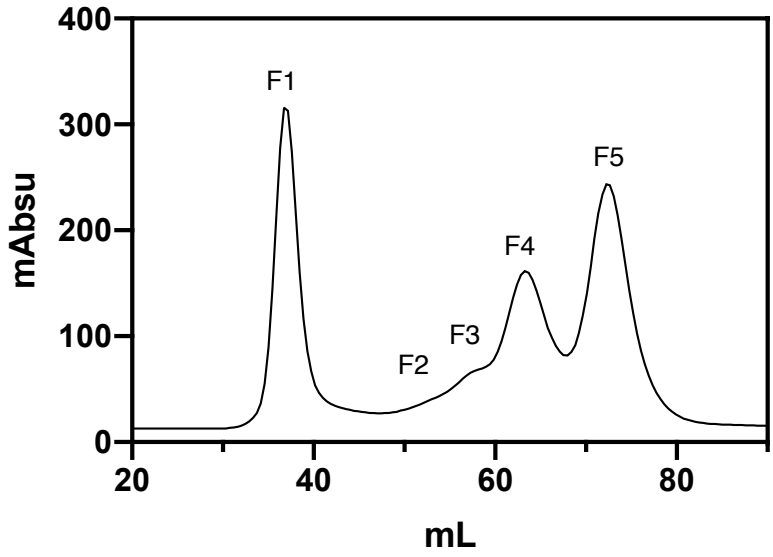


Figure 11.3. SEC chromatogram of *Vc*CDA FL Chromatogram obtained in filtration gel chromatography using a Superdex 200 column (16/600). Five multimeric fractions are shown. F5 corresponds to the monomer.

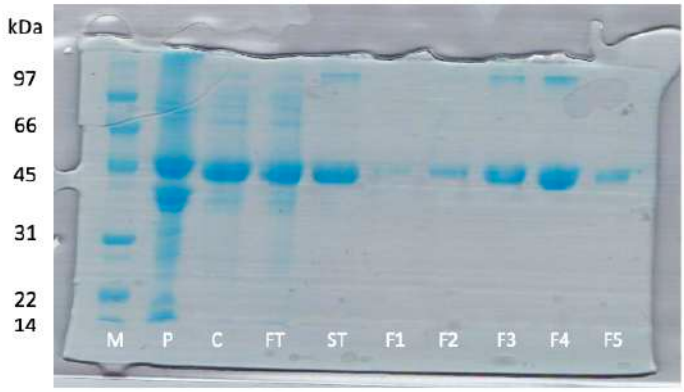


Figure 11.4. 14% SDS-PAGE gel of the entire purification process of *Vc*CDA FL. M: Marker Low Range, P: Pellet from the centrifugation after the lysis, C: crude, soluble fraction of the lysate, FT: Flow through the column, ST: Protein purified by the Strep-Trap, F1-F5: SEC fractions 1 to 5.

The presence of a band with a molecular weight similar to *Vc*CDA FL is observed in the cell debris lane. This could be due to the fraction of the protein expressed using this protocol is not properly fold and forms insoluble inclusion bodies, or on the other hand, that cell lysis is

not complete and this band corresponds to a correctly expressed enzyme but only released after boiling samples for loading onto denaturing SDS-PAGE gel. In the fourth lane corresponding to flow through (FT), interpreted as the volumetric fraction that has passed through the column in its loading process, the presence of a band that corresponds to the expected molecular weight of the protein of interest is observed. This is due to the load is exceeding the capacity of the column and part of the expressed protein cannot bind to the matrix because it is saturated. Although the nominal capacity described by the manufacturer (GE Healthcare) should be close to 5 mg of protein for each mL of medium, under the conditions used it has been observed that this is not applicable to our protocol, indicating that the cause is due to having exceeded the binding capacity of the column. In the last 5 lanes, all different fractions from gel filtration purification are shown. All 5 fractions show a high purity, monomeric fraction (F5) was selected for further studies.

11.2. Kinetic characterization of the *Vc*CDA FL

11.2.1. Reported methods for analysis of deacetylase activity

Once the enzyme was purified, characterization using all the available substrates was done: DP2, DP4 y DP5. In order to address the problem of evaluating the degree of acetylation of chitosan oligomers and polymers for enzymatic activity analysis there are published several different methods based on several distinct properties of these molecules such as infrared spectroscopy [165], UV-Spectrophotometry [166], NMR [167], potentiometric titration [168], etc. Two general groups of methods of monitoring CDA activities in chitin oligomers can be identified: Direct methods that are based on many characteristics of the chitosans released in the enzymatic reaction or the use isotopic labeled substrates. One of the most direct methods is the measure of the absorbance at 216 nm, the base of this method is the difference in extinction coefficient between the carbonyl functional group in the chito-oligosaccharide substrates and the acetate product released [169]. Despite being simple, this method is generally regarded as having low sensibility and low specificity. On the other hand, radiometric methods have been applied for the evaluation of deacetylase activity using [^{3}H]acetylglycol-chitin [170]. Notwithstanding being very sensible methods, the main inconvenience comes from the need of labelling the substrate and how uncomfortable radiometric assays are. Direct methods have evolved towards incorporating chromatographic (or other type of separation) stages in order to separate oligomers in terms of their DP. There

are several modes of detection used to monitor product formation: HPLC-UV methods [171], Refractive index [172], MS [173]. But those have not shown the throughput enough to analyze a high number of samples. Indirect methods are based in coupled chemical or enzymatic reactions in order to measure product formation. Procedures based on the chemical modification of the produced chitosans are commonly centered in derivatization of the free amine generated in the deacetylation in order to create a compound measurable by spectrophotometry (either colorimetric or fluorometric). There are two basic chemical reactions for generating colored products: those that use indole [123], 3-methyl-2-benzothiazolone hydrazine (MBTH) [174] or those that use ninhydrin [175]. Methods that generate fluorescent products derived from chitosan often use fluorescamine [176]. Enzymatic coupled assays are generally focused on measuring the accumulated acetate. The advantage of these methods is that they are presented in commercial kit format (Megazyme, R-biopharm). These kits vary in the enzymes that will translate the acetate concentration into the conversion of NAD^+ into NADH (kit acet, Megazyme; R-Biofarm) [177] or NADH into NAD^+ (kit acetk, Megazyme) by measuring changes in absorbance. Disadvantages of indirect methods include: their complexity and the interference of other products present in the reaction (free amines, acetate not resulting from deacetylation, etc). There are also indirect colorimetric methods that have incorporated a separation step in their protocols either HPLC or electrophoresis. Such is the case of labelling the products to increase the sensibility of the method itself with the 1-phenyl-3-methyl-5-pyrazolone (PMP) [178] or crystal violet [179]. Other methods for the estimating CDA activity by using artificial substrates like 4-methylumbelliferyl acetate [180] or p-nitroacetanilide [181]. The basis of these assays is that after the cleavage of the acetate the resulting molecule is able to be excited at a defined wavelength and develop fluorescence.

11.2.2. Development and validation of an HPLC-MS method with high capacity for kinetic studies

Unfortunately, there is no technique described so far that will be able to measure the direct accumulation of the deacetylation product with the yield and sensitivity necessary for the rapid characterization of a number of natural or mutant deacetylases. This highlights the need for an analytical method able to measure the deacetylase activity in a HTS approach .

In previous work in our group [104], it was reported the optimization and validation of a fast and direct analytical method using HPLC-MS system integrating hydrophilic interaction liquid chromatography (HILIC) step and detection by mass spectrometry for the quantitative analysis of soluble chitosan oligosaccharides in reactions with free purified protein or from cell extract. HILIC is a separation mode used in HPLC techniques designed in a way that it allows the separation of polar molecules, such as chitin and chitosan. This method is applicable for a large variety of compounds independent of their origin, degree of polymerization, degree of acetylation, and pattern of acetylation.

The protocol described includes a general methodology for enzymatic characterization of CDAs that allows automatic sampling of kinetic assays, allowing the separation between the sampling stages and the kinetic measurement step. This property makes it possible to significantly increase the analysis capacity per operator. The use of an HPLC-MS system has the advantage of matching the performance of robotic platforms in the sampling, automatic injectors in the chromatography section and the selectivity of the MS detectors. The result is a fast, versatile and reliable method for the direct quantification of chitosan oligomers.

Taken altogether, an HPLC-MS method was set up for the characterization of chitosan oligosaccharides resulting from the analytical deacetylation of chitoligosaccharides from the action of de-*N*-acetylases. *Vc*CDA FL wt enzyme was selected for setting up the HPLC-MS assay. As already stated, *Vc*CDA wt enzyme deacetylate the second unit from the non-reducing end in chitoligosaccharides. Preparative reactions with chitoligosaccharides ranging from DP2 to DP4 were performed in order to obtain pure standards of the products for the optimization of the separation and detection protocol.

All analytic procedures were carried out using an Agilent Series 1260 (Agilent, USA) liquid chromatograph equipped with a vacuum degasser, a binary pump, a UV-vis detector , autosampler and a mass spectrometer using an electrospray ionization font and a single quadrupole MS detector. Chromatographic separation was performed on a XBridge BEH Amide 2,5 μm 2.1x100 mm Column XP (Waters), an injection volume of 5μL and isocratic elution at 60ºC with acetonitrile/water 65:35 v/v 1% formic acid at a flow rate of 0,4 ml/min. The electrospray source was set for only positive mode ion detection. Capillary voltage was defined at 4kV and the desolvation gas (nitrogen) was set at 12L/min and a temperature of 300ºC.

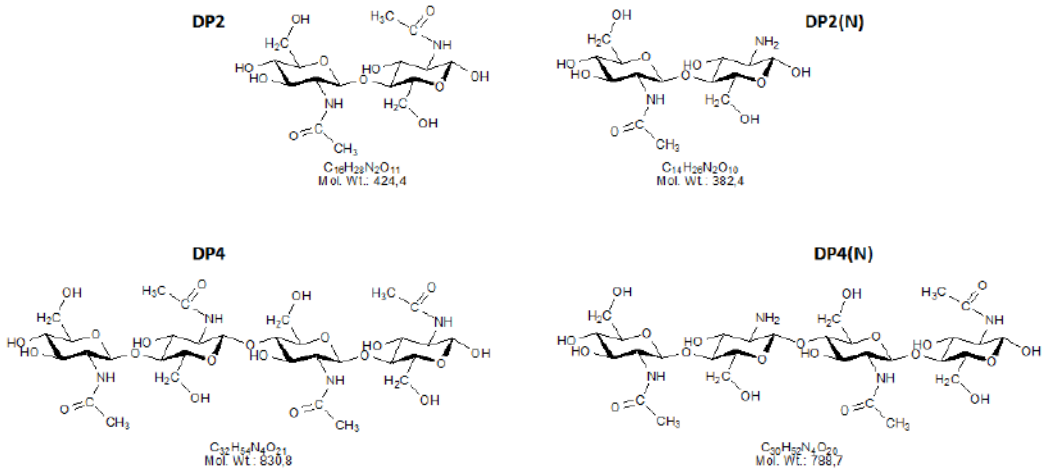


Figure 11.5. Structures of chitin oligosaccharide DP2 and DP4 and their respective chitosan homologs after being monodeacetylated by the *Vc*CDA. This enzyme deacetylates the second position from the non-reducing end.

The accumulation of the deacetylated products was measured by Single Ion Monitoring (SIM) $[M+H]^+$ species, in this case m/z 383, 586, and 789 for monodeacetylated di-, tri- and tetrasaccharide products, respectively. In order to transform abundancies into concentrations, the area of each peak was interpolated into a standard curve.

Three different reaction cocktails were performed, one for each substrate (DP2, DP3, DP4), and mixed with a defined volume of *Vibrio cholerae* chitin oligosaccharide deacetylase. The course of the reaction was monitored by means of a preliminary MS method. When total substrate consumption or the absence of reaction evolution were confirmed, chitosan oligomers were purified by ion-exchange chromatography using a previously washed and conditioned Dowex 50wx8 resin. Elution of pure products was accomplished by means of a NH_3 gradient. After freeze drying, the overall efficiency of the process reached around 50% of the theoretical expected de-*N*-acetylated products. Small samples of the pure compounds were dissolved in n-propanol:water 50:50 and ion spectra were acquired in the MS spectrometer in positive detection mode (see protocol 21.5.3). MS analysis of the resulting compounds indicated high degree of purity letting us to confirm their suitability for being used as references in the optimization of the method and later for the definition of standard curves.

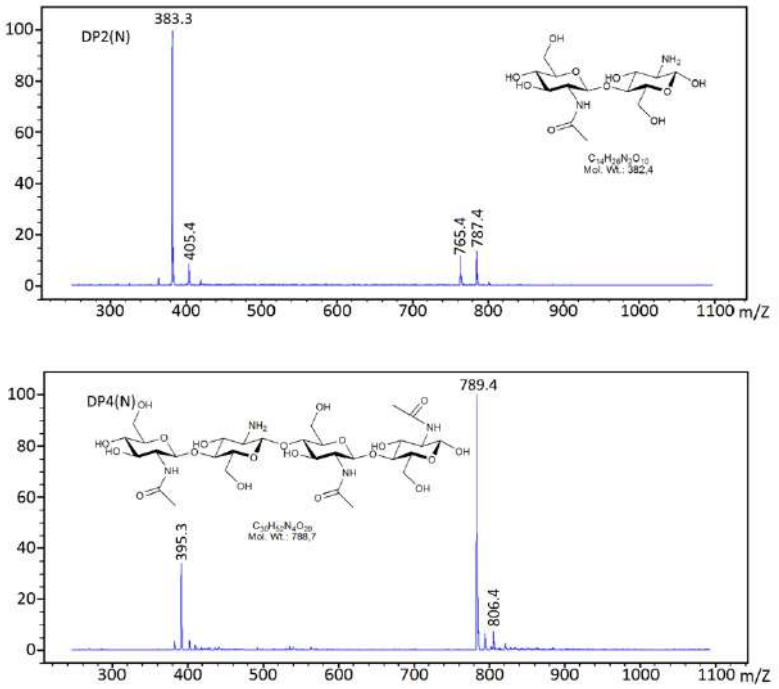


Figure 11.6. A) Ion scan spectra for the DP2(N) products after deacetylated by *Vc*CDA and purified. m/z 383.3=[DP2(N)+H]$^+$; m/z 405.4= [DP2(N)+Na]$^+$; m/z 765.4= [2·DP2(N)+H]$^+$; m/z 787.4= [2·DP2(N)+Na]$^+$. B) Ion scan spectra for the DP4(N) products after deacetylated by VcCDA and purified. m/z 789.4=[DP4(N)+H]$^+$; m/z 806.4= [DP4(N)+Na]$^+$.

As shown in the corresponding mass spectra (for DP2(N) and DP4(N) respectively), the most relevant peaks illustrate that the most present form is the mono protonated charged ion, being other cations anecdotic as it is the case of the Na^+ adduct. This piece of information lead us to choose the monoprotonated charged ion as the element that we should measure in the final SIM method. It is needed to mention that for the application of this method standard compounds must show the same pattern of acetylation (PA) as the products to be studied. Different PA will show different response factors even if they have the same degree of acetylation. Afterwards, in previous work different parameters such as stop buffer, method repeatability, and analytical validation among others were analyzed in order to set up the HPLC-MS method [100]. All volumetric manipulations, incubation and sampling was automatized using a Bravo Liquid Handler robotic platform (Agilent).

In order to check the validity of the whole procedure to monitor deacetylation, two enzymatic reaction were prepared, one for each substrate DP2 and DP4. Substrate concentration used was 2 mM dissolved in Tris 50 mM pH 8.5 and NaCl 300 mM to a final volume of 150 μl, the enzyme concentration used was 0.75 μM. Samples were withdrawn from the cocktail and

stopped. Samples (10 μL) were withdrawn from the reaction and mixed with 90 μL of 1-propanol:water 1:1 in order to stop the reaction at different time intervals. The follow up of the reactions was sustained for as long as 40 hours. Consequently, all the generated samples and standards were analyzed with the previous developed HPLC-MS methodology. Concentrations were obtained by using response factor values calculated from a second set of standards equivalent to the ones showed in Figure 11.6 (both for substrate and deacetylated product).

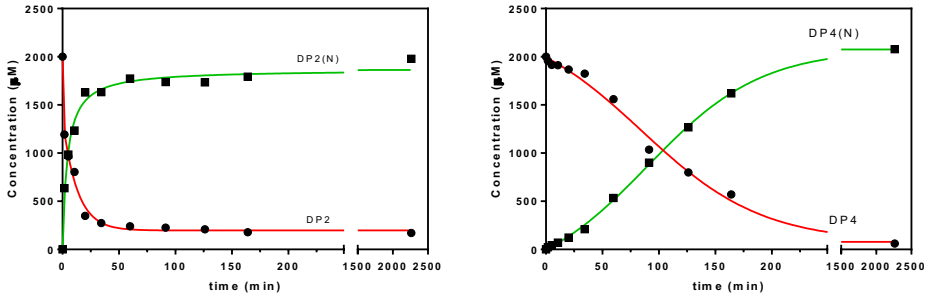


Figure 11.7. *Vc*CDA reaction progress curves at pH 8.5 and 37°C of the chitobiose and chitotetraose substrate. Concentrations are calculated using a response factor for each compound: DP2 or DP4 as substrate and DP2(N) or DP4(N) as product.

This experiment allowed us to corroborate its applicability in the case of measuring chitosan oligosaccharides resulting from an enzymatic catalyzed reaction.

Steady state kinetic parameters were determined for the two chitooligosaccharides (DP2 and DP4). Two enzymatic assays were carried out as replicates for each DP. Initial velocities were calculated from the linear region and V_0/[Enzyme] were plotted against substrate concentration. Experimental results were submitted to a non-linear regression, for DP2 and DP4 a common Michaelis-Menten ($V_0=k_{cat}\cdot[S]/(K_M+[S])$) behavior

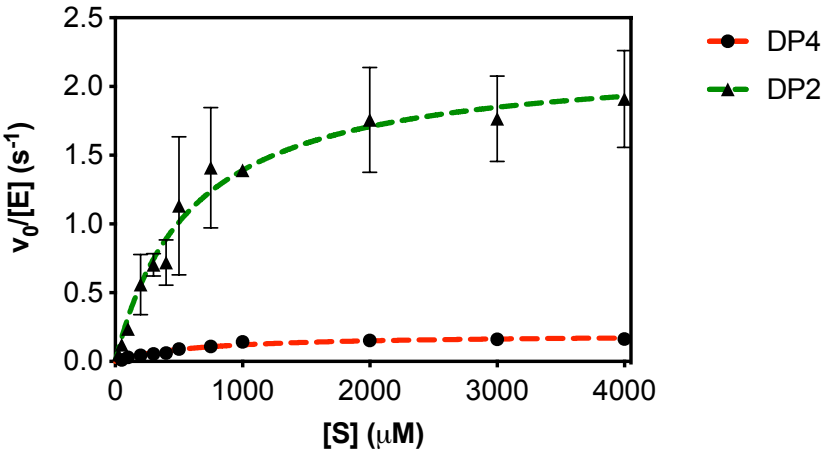


Figure 11.8. Michaelis-Menten curves for the purified monomeric *Vc*CDA deacetylation reaction at substrate concentrations ranging from 4 mM to 0.025 mM in Tris 50 mM pH 8.5 and NaCl 300 mM. Temperature was kept constant at 37°C and a final volume of 100 µl. Substrates used were DP2 and DP4.

Table 11.1. Kinetic parameters of wt enzyme. Conditions: 25 μM to 4 mM substrate, 0.05-0.8 μM enzyme, 50 mM phosphate buffer pH 8.5, 300 mM NaCl, 37°C.

		M&M parameters		
	Substrate	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m ($M^{-1}\cdot s^{-1}$)
WT	DP2	2.02 ± 0.19	0.46 ± 0.10	$3.7\cdot10^3$
	DP4	0.20 ± 0.01	0.64 ± 0.07	$3.1\cdot10^2$

In previous work [104], values of k_{cat} for DP2 and DP4 were 2.19 s^{-1} and 0.11 s^{-1} and for K_M of DP2 and DP4 were 0.36 mM and 0.52 mM. In terms of k_{cat} for DP2, values are similar however for DP4 previous work obtain half of the value obtained in Table 11.1. K_M values are slightly higher in this work. Different parameters can affect these values such the buffer preparation, scientist hands, among others. Values obtained are within the expected results.

11.3. Protein thermostability characterization

Thermostability is the resistance to irreversible chemical or physical changes of a substance due to elevation in temperature [182]. Protein thermostability is, therefore, the preservation of the unique structure and chemical properties of polypeptide chains under high temperatures and is a fundamental property of enzymes because it enhances productivity in industry, as enzymes operate at higher temperatures where more reagents and compounds are available. Enhanced thermostability is one of the most common properties desired as output from a protein engineering study and is often an important economic factor. The stability of a protein is affected by many factors, such as temperature, pH, solvent, and the presence of surfactants. Among all possible deactivating factors, temperature is the best studied. At elevated temperatures, many enzymes tend to become (partly) unfolded and/or inactivated, meaning that they are no longer able to perform the desired tasks. There are two types of protein stability, thermodynamic stability and thermotolerance, that are crucial from an applied perspective. Thermostability, however, is a consequence of the combination of several factors acting in synergistic manner and not due to a single specific major factor. We next sought to determine different thermostability-related parameters to further perform a characterization of enzyme properties.

11.3.1. Temperature profile of enzyme

As with many chemical reactions, the rate of an enzyme-catalysed reaction increases as the temperature increases. However, at high temperatures the rate decreases again because the enzyme becomes denatured. Optimal temperature of enzyme was obtained by temperature activity profile (Figure 11.9). Specific activity using DP4 as a substrate was analyzed at different temperatures (30ºC, 37ºC, 50ºC and 60ºC). Deacetylase activity was determined by HPLC-MS method according section 11.2.

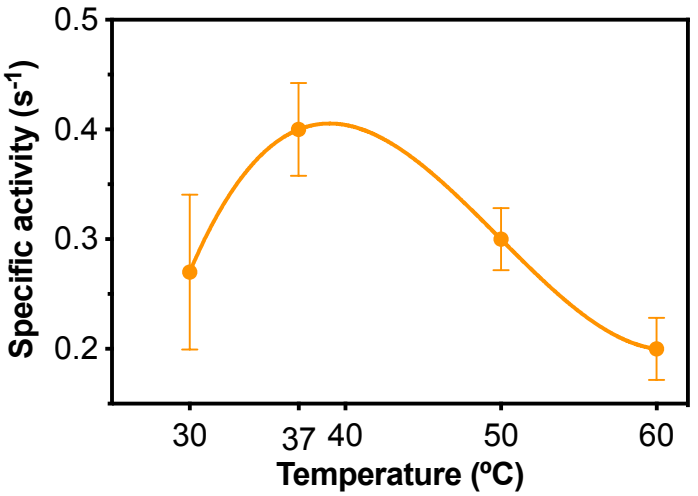


Figure 11.9. The effect of temperature on the specific activity of *Vc*CDA FL wt. Reaction conditions: Buffer: 50 mM phosphate buffer pH 8.5, 300 mM NaCl, 37 °C.

Table 11.2. Specific activity (s^{-1}) for wt at different temperatures. Reaction conditions: Buffer: 50 mM phosphate buffer pH 8.5, 300 mM NaCl, 37 °C.

	Specific activity (s^{-1})
Temperature (ºC)	**WT**
30	0.27 ± 0.07
37	0.40 ± 0.04
50	0.30 ± 0.03
60	0.24 ± 0.03

Maximum activity is reached at 37ºC (Figure 11.9 and Table 11.2). Enzyme activity drops as the temperature increases.

11.3.2. Melting temperature

One of the easiest parameters to estimate the stability of a protein is the determination of its Melting temperature (Tm). This parameter is defined as the temperature at which half of the protein population is found in denatured form and the other half in native form. The more thermally stable a protein is, the greater the energy input to destabilize its structure and, consequently, the higher its Tm. The study of Tm is also used in many other aspects such as

the binding of ligands, protein-protein interactions, the effect of denaturing agents, etc. An assay developed for the study of Tm is called Thermal Shift Assays or TSA. A TSA was developed based on the use of a fluorescent probe, Thermofluor which was later adapted to miniaturized assays in 96-well microplates. A later revision of this methodology established the use of a new probe, the Sypro Orange, whose advantages are a greater sensitivity and that the wavelengths used both in its excitation and the detection of the emitted radiation (483nm and 560nm respectively) are found in most real-time PCR equipments. This fact facilitates the monitoring of the fluorescence of many samples in a system in which the denaturation temperature ramp could be controlled very efficiently. Once this method was established as a reference protocol for the study of protein denaturation, it was named Differential Scanning Fluorimetry or DSF [182,183].

The mechanism that makes the Sypro Orange a denaturation probe is exemplified in the image set in Figure 11.10. This compound nonspecifically binds to hydrophobic surfaces. In this way, when the temperature increases and the enzyme reaches states of denaturation, areas of the hydrophobic core that were previously shielded inside are exposed and become accessible to the probe (Figure 11.10.). The exclusion of the water layer that surrounds the Sypro Orange allows it to increase its fluorescence very significantly. At a certain point where denaturation is well advanced, aggregates of denatured protein can form, reducing overall fluorescence. The mathematical treatment of the results allows the adjustment of a sigmoidal equation which provides the Tm of that protein under these conditions (Equation 11.1).

Equation 11.1. Boltzman Sigmoidal equation.

$$F = F_{min} + (F_{max} - F_{min}) / (1 + e^{(Tm-T)/slope})$$

Melting temperature for VcCDA FL wt enzyme was determined in presence and absence of Zn^{2+} (0,1 mM $ZnCl_2$) (see protocol 21.1.13). As already stated, this enzyme is metal dependent using Zn^{2+} as a metal. 5 µL of Sypro Orange solution were added to 20 µl of protein solution (4 µM final concentration) in 50 mM phosphate buffer pH 8.5, 300 mM NaCl. A thermocycler Rotogene 3000 (Corbett Research) was used to quantify the fluorescence (see protocol 21.1.13.). Tm (°C) values are given in Figure 11.11.

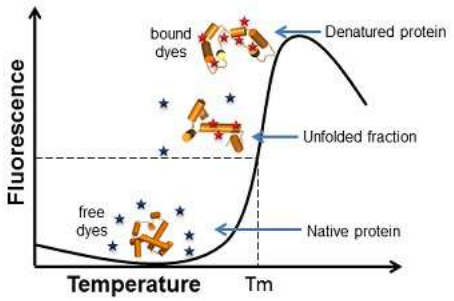


Figure 11.10. Mechanism of action and calculation of Tm using Sypro Orange in a TSA test. Scheme of operation of the Sypro Orange reagent used in a TSA protocol in which the increase in its fluorescence is monitored by binding to the hydrophobic areas of the proteins exposed to the solvent in the denaturation. Example of the crude result obtained in a TSA experiment in which, once a maximum point is reached, the fluorescence is reduced by the formation of protein aggregates. Adapted from [184].

In figure Figure 11.11, presence of Zn^{2+} increases the melting temperature in wt enzyme. This phenomena could be explained through the role of the catalytic metal structure in the stabilization of the enzyme. The presence of Zn^{2+} could stabilize the protein due to the binding to secondary binding sites.

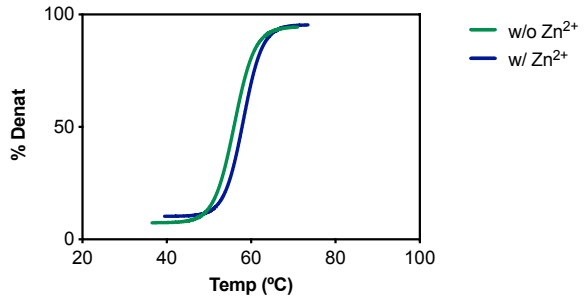


Figure 11.11. Thermal stability of *Vc*CDA wt in presence or absence of Zn. Green: Tm of VcCDA FL wt without Zn^{2+}: 56.0 ± 0.17, Blue: Tm of VcCDA FL wt with Zn^{2+} (0,1 mM $ZnCl_2$): 58.2 ± 0.2.

11.3.3. Thermotolerance

The structural stability of proteins has been traditionally studied under conditions in which the folding/unfolding reaction is reversible, since thermodynamic parameters can only be determined under these conditions. In our work, thermotolerance has been studied. Thermotolerance concept study the reversibility of the denaturation process. Proteins were incubated at defined temperature (50ºC) for different intervals of time, then protein was cooled down at 37ºC and the residual activity was determined. Thermotolerance is described by the half-life time ($t_{1/2}$) that is the time period necessary for the residual enzymatic activity to decrease to half of its initial value after incubation at a certain temperature, and the rate of

thermal inactivation (K_i) that determines the rate of the decreasing of the residual enzymatic activity.

In figure Figure 12.28, purified enzyme (2 μM) as incubated in PBS pH 8.5 buffer at 50°C. Then, aliquots were taken at different time intervals, diluted 1:10 in PBS pH 8.5, and kept at 37°C for 10 min (final enzyme concentration 0.2 μM). Residual activity was determined using DP4 substrate at 2 mM concentration (see protocol 21.1.14.).

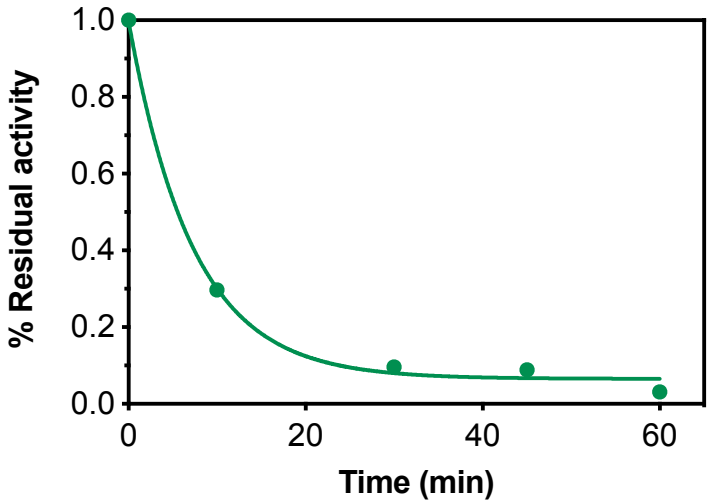


Figure 11.12. The thermotolerance of *Vc*CDA FL WT was determined with free enzymes at 50°C. Exponential decay equation has been adjusted (Equation 21.1).

Table 11.3. Thermotolerance of *Vc*CDA FL wt free enzyme at 50°C. Half-live (t_{50} min) monitoring residual activity with DP4 substrate (HPLC-MS).

t_{50} at 50°C (free *Vc*CDA FL wt)	**K_i (min^{-1})**	**0.14 ± 0.01**
	$t_{1/2}$ (min)	**5.4**

*Vc*CDA wt enzyme in free state has a half-life value of 5.4 min at 50ºC using DP4 as a substrate.

CHAPTER 2. SCREENING ASSAY FOR DIRECTED EVOLUTION OF CHITIN DEACETYLASES

Part of this Chapter has been published in:

Screening Assay for Directed Evolution of Chitin Deacetylases: Application to Vibrio cholerae Deacetylase Mutant Libraries for Engineered Specificity.
Sergi Pascual and Antoni Planas
Anal. Chem. 2018, 90, 18, 10654–10658,
https://doi.org/10.1021/acs.analchem.8b02729

12. CHAPTER 2. Screening assay for directed evolution of chitin deacetylases

12.1. Introduction

CDAs and CODs show diverse deacetylation patterns, some fully deacetylating their COS substrates, but few recently characterized CDAs showing specific deacetylation patterns (Figure 9.11). Ongoing research programs search for new CDAs from fungal and bacterial origin with the aim of discovering novel specificities and expanding the toolbox of CDAs and CODs for the enzymatic production of tailored paCOS and chitosans for biotech and biomedical application. The structural determinants of CDA specificities are poorly understood, with some new insights coming from the recently solved X-ray structures of enzyme·substrate complexes (*Vc*CDA [185] and ArCE4 [186]), and the proposed “subsite capping model” [100,185] to rationalize specificity and deacetylation patterns. Protein engineering will also offer the opportunity to engineer substrate specificity, either by rational or directed evolution approaches. In this context, an efficient assay for deacetylase activity implemented in HTS format is needed to screen either natural CDAs or directed evolution libraries in the search of novel specificities and improved efficiencies.

Here we report the design and development of a high-throughput assay for the screening of directed evolution libraries of CDA and other CE4 enzymes active on COS and its proof-of-concept application to the engineering of substrate specificity of a *Vibrio cholerae* chitin deacetylase (*Vc*CDA). The main goal of the directed evolution program presented is redesign the *Vc*CDA wt enzyme to increase the deacetylase activity in terms of k_{cat} for longer substrates such as DP4 and DP5, but maintaining the high specificity for the deacetylated position and maintaining the high activity towards short substrates like DP2. The HTS assay developed is based an the detection of deacetylase activity in a library of mutants.

Enzyme assays for deacetylation activity of CDAs and CODs are diverse. Whereas unspecific substrates for general amidase activity (4-methylumbelliferyl acetate, [187] 4-nitrophenyl acetate or p-nitroacetanilide [181]) are often used to monitor protein purification, assays using COS as substrates are specific for kinetic characterizations. These are based on monitoring acetate release (UV absorbance changes [188], radiolabeled substrates [189,190] and coupled enzymatic assays [191,192]), or formation of free amino groups with chromogenic or fluorogenic reagents such as fluorescamine, [193] o-phthalaldehyde (OPA), [194] or ninhydrin

[195]. All of these methods use purified enzymes and are not directly applicable to an HTS format with cell extracts as required for screening in directed evolution approaches. On the other hand, more complex HPLC-MS methods have been developed to monitor deacetylation and sequence the acetylation pattern of the products [196] but all of these methods have not been able to perform the assay in High Throughput Screening (HTS) format that is necessary to evaluate large numbers of samples, as might be necessary for screening large mutant libraries.

The HTS assay, implemented in 96-well microtiter plates, is based on fusion of the target CE4 catalytic domain to a chitin binding module (CBM), capture of the expressed proteins from cell-free extracts with chitin-coated magnetic beads, and evaluation of the deacetylase activity of the immobilized enzyme variants on natural COS substrates by monitoring product formation with a coupled assay leading to a fluorescence readout (Figure 12.1).

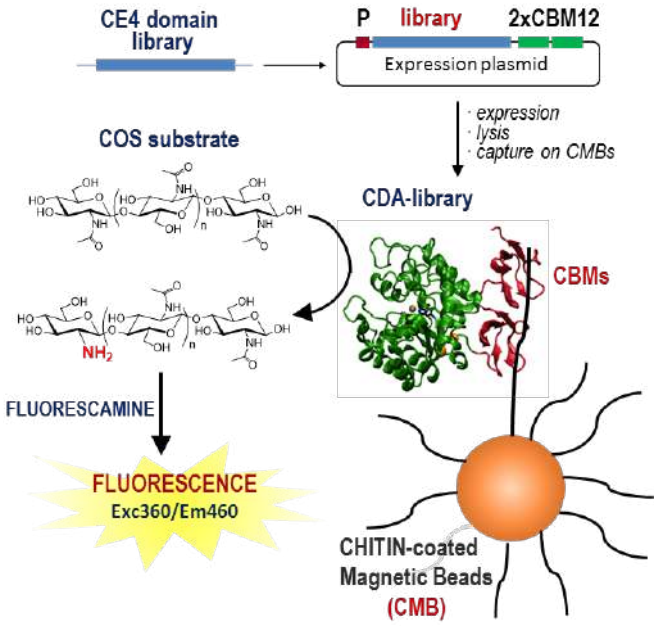


Figure 12.1. Principle of the HTS assay for deacetylase activity on COS.

12.2. Development of High Throughput Screening (HTS) assay.

For the HTS assay development, different parameters were first optimized using the *Vibrio cholera* chitin deacetylase (*Vc*CDA wt), [124,185] either as purified enzyme or as cell-free extracts of recombinant *E.coli* expressing the enzyme. *Vc*CDA is a multidomain protein composed of a CE4 catalytic domain with chitooligosaccharide deacetylase activity and two

family 12 chitin binding modules (CBM12) at the C-terminus, which have shown to bind insoluble chitin (Figure 9.12) [197]. This tandem of two CBM12 domains will be used in the general HTS protocol; the target CE4 catalytic domain (library of mutants to be screened) will be fused to the CBMs acting as a C-terminus tag for binding to the chitin magnetic beads (CMB) in order to immobilize the enzyme.

Enzyme immobilization is confinement of enzyme to a phase (matrix/support) different from the one for substrates and products. Inert polymers and inorganic materials are usually used as carrier matrices. Apart from being affordable, an ideal matrix must encompass characteristics like inertness, physical strength, stability, regenerability, ability to increase enzyme specificity/activity and reduce product inhibition, nonspecific adsorption and microbial contamination. Different techniques are used for protein immobilization: physical immobilization, absorption, covalent binding, affinity immobilization and entrapment [198].

As a consequence of enzyme immobilization, some properties such as binding, catalytic activity or thermal stability may become altered. These effects have been demonstrated and exploited. The concept of stabilization has been an important driving force for immobilizing enzymes. Moreover, true stabilization at the molecular level has been demonstrated.

Chitin magnetic beads (CMB) are an affinity matrix for the small-scale isolation of target proteins fused to a chitin binding domain (CBD). In our work, CBM12 domains act as a CBD. Chitin magnetic beads have been prepared having a magnetite core, thereby permitting the magnetic isolation of CBD-fusion proteins from cell culture supernatants. In several studies, CMB were used as single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein-splicing element

Taken altogether, different parameters and conditions of the screening assay for deacetylase activity were analyzed and developed.

12.2.1. Evaluation of protein binding to Chitin Magnetic Beads (CMB).

Different conditions were tested to study and determine the best conditions for binding and immobilization of *Vc*CDA tagged with CBMs to the CMB. Traditionally, the estimation of immobilized enzyme on the support is calculated from the enzyme content in the solution before and after immobilization using protein quantification methods. Conditioning of the CMB (before the binding step) and binding of the protein to CMB were tested using different buffers

and conditions (see protocol 21.1.1.). Protein bound and unbound were analysed by BCA assay (Figure 12.2) and SDS-PAGE (Figure 12.3).

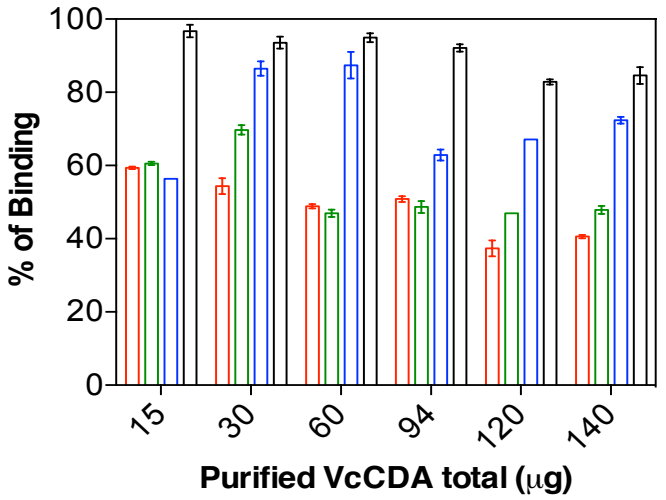


Figure 12.2. Percentage of bound protein to 25 μL CMB suspension at increasing purified protein loads. The binding assay was performed at 4°C, 650 rpm for 1 hour using different buffers. Buffers: A (red column). PBS (50 mM phosphate, 300 mM NaCl, pH 8.5) as conditioning and binding buffer. B (green column). PBS+Triton X-100 (0.05%) as conditioning buffer and PBS as binding buffer. C (blue column). CBB (20 mM Tris·HCl, 500 mM NaCl, 1 mM EDTA, 0.05% Triton X-100, pH 8.0) as conditioning buffer, and PBS as binding buffer. D (black column). CBB as both conditioning and binding buffer.

SDS-PAGE gels of bound and unbound (supernatant) protein (0-140 μg *Vc*CDA incubated with 25 μL CMB suspension with PBS pH 8.5 as conditioning and binding buffer at 4°C for 1 h) are shown in Figure 12.3. After washing the CMBs with PBS, both the beads (CMB gel) and the supernatant (Supernatant gels) were analyzed by SDS-PAGE (see protocol 21.9.1.).

Maximum binding was obtained using CBB buffer in conditioning and binding steps. For the screening assay, the buffer for conditioning and binding to the CMB was selected not to contain detergents and EDTA (as commonly used with commercial CMB). Although some CDAs are not inhibited by EDTA, indicating strong metal cation coordination, [99,185] it cannot be assumed general when applying the screening to directed evolution mutant libraries. Therefore we chose a simple PBS buffer (PB 50 mM, NaCl 0,3M at pH 8.5) where the efficiency of binding was about 50-60% in the range of concentrations tested (Figure 12.2). To determine if the bounded protein is still active using different buffers, deacetylation activity of the protein bound to CMB was determined using DP4 (2 mM) as a substrate by HPLC-MS (Figure 12.4) (see protocol 21.1.2.) (see section 11.2).

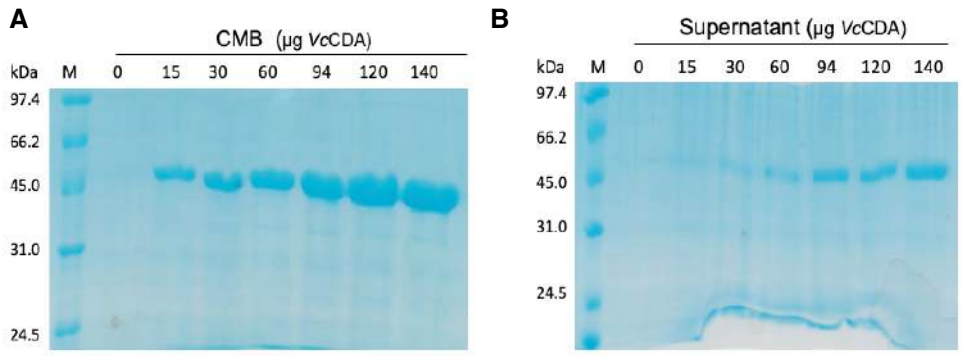


Figure 12.3. Binding efficiency of VcCDA to the CMBs using buffer A: PBS (50 mM phosphate, 300 mM NaCl, pH 8.5) as conditioning and binding buffer. M: molecular weight marker.

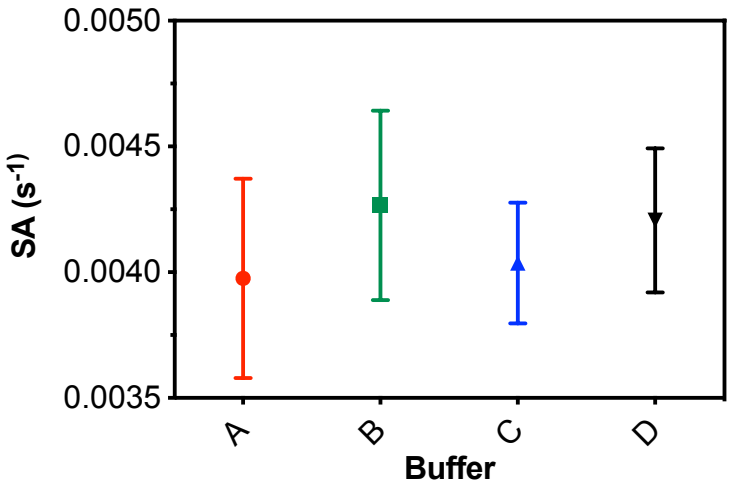


Figure 12.4. Binding efficiency of *Vc*CDA to the CMBs using different conditioning and binding buffers. Deacetylase activity (60 µg *Vc*CDA incubated with 25 µL CMB suspension with different buffers as conditioning and binding step at 4°C for 1 h) was determined using tetraacetylchitotetraose (DP4) as a substrate (2 mM). Buffers: A (red column). PBS (50 mM phosphate, 300 mM NaCl, pH 8.5) as conditioning and binding buffer. B (green column). PBS+Triton X-100 (0.05%) as conditioning buffer and PBS as binding buffer. C (blue column). CBB (20 mM Tris·HCl, 500 mM NaCl, 1 mM EDTA, 0.05% Triton X-100, pH 8.0) as conditioning buffer, and PBS as binding buffer. D (black column). CBB as both conditioning and binding buffer. SA= Specific activity.

The presence of detergents and chelating agents does not affect the activity once the protein is bound. Therefore Strategy A with standard PBS buffer (PB 50 mM, NaCl 0.3M at pH 8.5) was selected as a conditioning and binding buffer for further steps.

12.2.2. Specificity of *Vc*CDA binding to the CMBs.

In screening campaigns, a library of mutants will be expressed and obtained from cell extract. A cell extract contains a huge variety of cytoplasmatic proteins that may interact with CMB and affect the binding of the target *Vc*CDA. To study this effect, 50 μL of purified *Vc*CDA solution containing 0-140 μg of protein were mixed with 100 μL of a cell-free extract of *E. coli* BL21(DE3) cells harboring an empty pET22b plasmid and the suspension incubated with 25 μL of CMB in PBS buffer at 4°C for 1 h with shaking (650 rpm). Bound and unbound proteins were analyzed by SDS-PAGE (Figure 12.5.). (see protocol 21.1.3.)

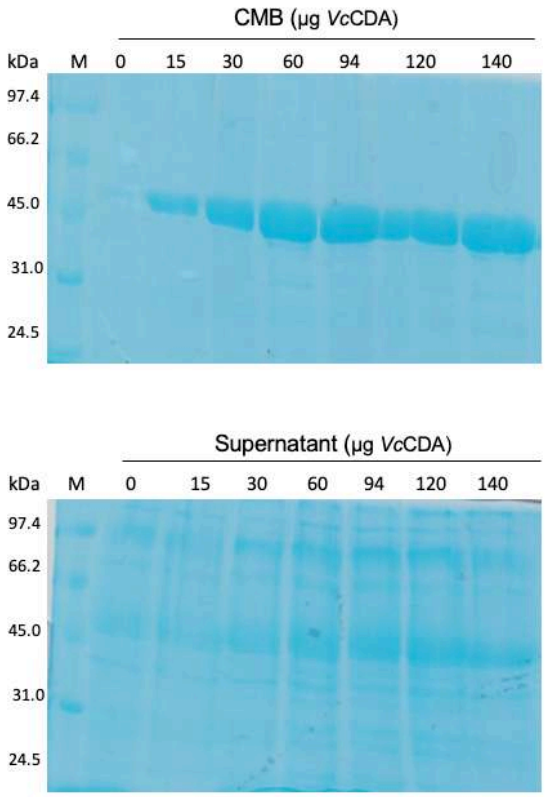


Figure 12.5. Specificity of *Vc*CDA binding to the CMBs. Upper gel: CBM-immobilized proteins, Lower gel: supernatants.

No interference from cell extract were observed. CMB are able to capture *Vc*CDA wt protein from the cell extract media without presence of other proteins (Figure 12.5).

On the other hand, BSA is highly used as a protein stabilizer. The high solubility of albumin (up to 40% w/v) at pH 7.4, its stability at pH values of 4 to 9, and temperature variations (up

to 60 ◦C when heated for 10 h) without any deleterious effects make it an attractive macromolecular stabilizer. BSA is able to stabilize an enzyme by promoting hydrophobic interactions and increasing the viscosity of the enzyme solution; however, this must be carefully balanced as increasing the protein concentration and viscosity increases the probability of aggregation which could lead to decreased enzyme availability or non-uniform nanoparticles formation [199]. As a result, BSA is used in some HTS since it provides a protective mechanism similar to increasing the enzyme concentration. However, BSA has discrete binding sites for anionic or hydrophobic molecules. To study BSA as a stabilizer (Figure 12.6), 40 μg of purified *Vc*CDA protein solution were mixed with 0-10 μg/μL of BSA in PBS buffer and 25 μL of CMB in PBS buffer were added and the suspension incubated at 4ºC for 1 h with shaking (650 rpm). After washing, the beads were analyzed by SDS-PAGE (see protocol 21.1.4.).

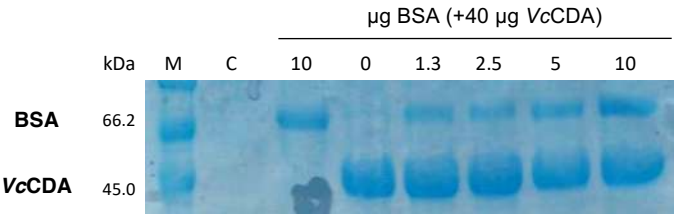


Figure 12.6. Specificity of *Vc*CDA binding to the CMBs in presence of BSA. M: molecular weight marker; C: no BSA nor VcCDA. In BSA lane, 10 μg of BSA without *Vc*CDA were added.

Binding was shown to be specific when cell extracts were used, not detecting any other bound proteins (Figure 12.5) discarding possible interactions with CMB. Moreover, if BSA is required as stabilizer, it had some affinity for the beads but it did not displace the bound CDA-CBM fusion protein (Figure 12.6)

The rate of protein binding to the CMB step is crucial for screening assays to reduce the global timing of the screening. To study the kinetics of binding, 25 μL of CMB in PBS pH 8.5 buffer were added to 90 mg of purified *Vc*CDA protein in 150 ml PBS pH 8.5, and the suspension was incubated at 4°C for different time intervals (15 to 90 min) with shaking (650 rpm). After washing, both the beads (CMB gel) and the supernatant (Supernatant gel) were analyzed by SDS-PAGE (Figure 12.7.A) (see protocol 21.1.5.).

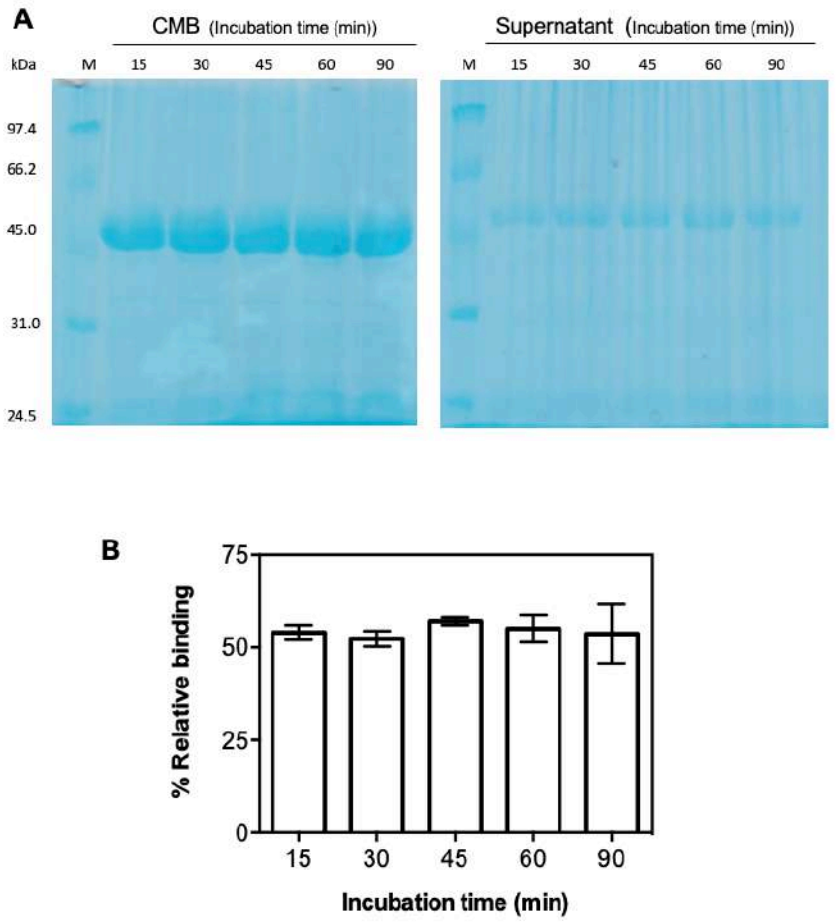


Figure 12.7. Kinetics of *Vc*CDA binding to the CMBs. A) SDS-PAGE of protein-bound and unbound to CMB after different incubation times; B) quantification of bound protein (by quantification of the unbound protein in the supernatant (S) by the BCA assay [200]).

For this, the kinetics of protein binding to CMB determined that maximum binding was achieved in less than 15 min at 4ºC with shaking (Figure 12.7.). The incubation time for binding can be varied from 15 to 90 minutes of incubation at will to be sure that protein is bound to CMB. In the final HTS protocol, an incubation time of 60 minutes was selected.

12.2.3. Deacetylase activity of protein bound to CMB.

Ideal immobilization should yield a stable biocatalyst while minimizing distortions in its structure and function. However, in many cases, immobilization results in a decreased enzyme activity relative to the free enzyme, and only there are few examples where the activity is unaltered or even increased. Importantly, in all cases, it is critical to check enzyme

activity after immobilization, and determine the effective concentration of active sites available. Comparation of the deacetylase activity between free enzyme and immobilized enzyme is studied using chitobiose and chitotetraose to show how binding affects the natural activity of *Vc*CDA. For this assay, purified *Vc*CDA or cell extracts were mixed with DP2/DP4 at a final concentration of 2 mM in a final volume of 100 μl, and the solution was incubated at 37°C. At different times intervals, aliquots (10 μl) were withdrawn and analyzed by HPLC-MS (see protocol 21.1.6.).

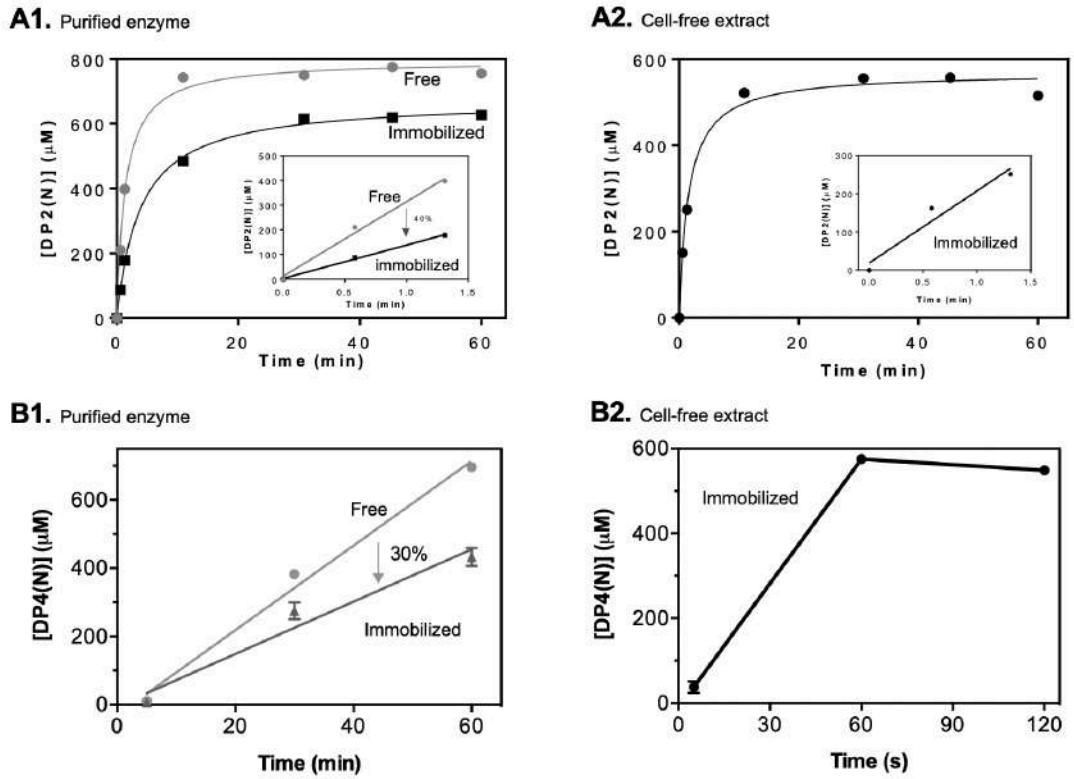


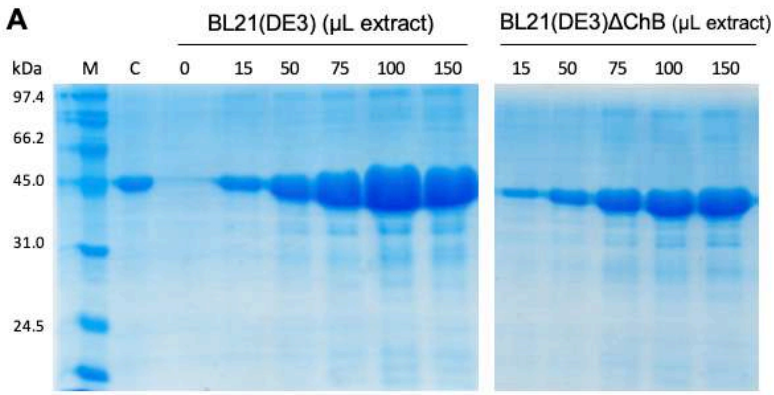
Figure 12.8. Deacetylase activity of free and CMB-immobilized *Vc*CDA using DP2 and DP4 as substrates. A) Activity with DP2 substrate monitored by HPLC-MS (protocol). A1) Purified enzyme: Free enzyme (grey circles), CMB-immobilized enzyme (black squares). Inset graph shows specific activity. A2) Cell-free extracts. 150 μl (black cicles) of a cell-free extract containing *Vc*CDA. B) Activity with tetraacetylchitotetraose (DP4) substrate monitored by HPLC-MS. B1) Purified enzyme. 30 μg of purified protein in free form (grey circles) and 60 μg of protein bound to CMB (black squares). B2) Cell-free extract. Deacetylase activity using DP4 was performed using the same protocol described above (A2) working with 100 μl of cell free extract.

Immobilization of *Vc*CDA on the CMBs resulted in 40% reduction of it activity for DP2 and 30% for DP4 relative to the free enzyme in solution using the same reaction buffer (volume of *Vc*CDA added to CMB was selected taking in to account the binding capacity of CMB to get the same amount of protein as free enzyme (Figure 12.2)), as determined by monitoring

product formation from diacetylchitobiose substrate by HPLC-MS (Figure 12.8). Immobilized cell-free extract shows deacetylase activity with both substrates DP2 and DP4. Importantly, both purified enzyme or cell extract bound to CMB are active for the substrate to be evolved.

12.2.4. Evaluation of *E.coli* expression hosts.

E. coli BL21 (DE3) was used as host for *Vc*CDA expression. Since *E.coli* expresses an endogenous chitin deacetylase, a ΔChib strain (knock-out of the chib operon)[201] was also tested as expression host to analyze protein binding and specific activity in order to determine possible interferences in screening assays. Here, *E. coli* expression hosts, BL21(D3) and BL21(D3)ΔChb were transformed with a pET22b plasmid containing a synthetic gene coding for *Vc*CDA [185]. Cell cultures of both strains were grown until OD 1.2 and binding efficiency was analysed by SDS-PAGE and the deacetylase activity, once the protein was bound by HPLC-MS (see protocol 21.1.7.)



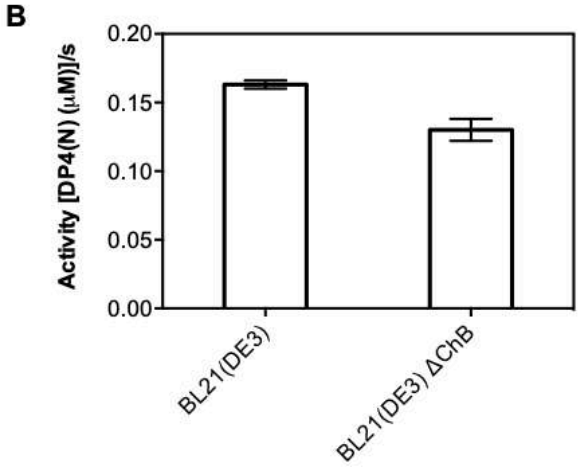


Figure 12.9. *E.coli expression hos*ts. A) Protein expression: 0-150 μl of cell-free extract from each strain were mixed with CMB in PBS pH 8.5 buffer, and the suspensions were incubated at 4°C for 1 h with shaking (650 rpm). After the binding step, the beads were retained with a magnet and the supernatants transferred to new tubes (S) analyzed by SDS-PAGE. B) Activity of the immobilized enzymes: The beads (CBM-immobilized proteins) coming from 75 μl of cell-free extract from both strains as above were kept at 37°C for 5 min. Reactions were initiated by the addition of 100 μl DP4 (2 mM). After 60 min, the beads were retained with a magnet, and 10 μl of supernatant were withdrawn and added to 90 μl of water/1-propanol (1:1) to stop the reaction and quantify product formation by HPLC-MS.

The same binding behaviour to the CMBs, as well as similar specific activity, was observed with a cell-free extract from both strains. Therefore, standard BL21 (DE3) cells were further used to express the CDA libraries as they showed a slightly higher protein expression level (Figure 12.9).

12.2.5. Evaluation of cell lysis method.

Cell lysis, performed to disrupt the cell wall structure to release the intracellular contents, is a routinely used technique in the field of biology, for example, in extraction of cellular products (e.g., nucleic acids, proteins, and metabolites) from microorganisms [1, 2], protein or enzyme screening, clinic diagnosis and bioanalytical detection [5, 6].

Traditional methods, employing mechanical (e.g., sonication, french press, bead milling), physical (e.g., freeze-thawing, osmotic shock), chemical (e.g., alkali, detergents), or enzymatic (e.g., lysozyme) treatments to disrupt the cell wall, have been extensively reviewed elsewhere. New methods have emerged in recent years to address the technical needs of micro-scale sample processing, for example in high-throughput screening or miniaturized assays. Here, cell lysis in 96-well deep well plates was studied using 3 different protocols:

Bugbuster Reagent®, freeze-thaw cycles, and sonication. BugBuster® Protein Extraction Reagent is formulated for gentle disruption of *E. coli* cell wall to liberate active proteins. The proprietary formulation utilizes a Tris-buffer based mixture of non-ionic and zwitterionic detergents that is capable of cell wall perforation without denaturing proteins.

The lysis efficiency was determined by analyzing the free protein released to the medium. Cell cultures were grown in 96-well deep-well plates following the autoinduction protocol (see protocol 21.4.1). Cells were harvested by centrifugation at 3500 rpm and the supernatants were discarded. Cell pellets were lysed by Bugbuster® commercial reagent, freeze-thaw cycles and sonication. Protein released to the media was quantified using the BCA assay (see protocol 21.1.8.).

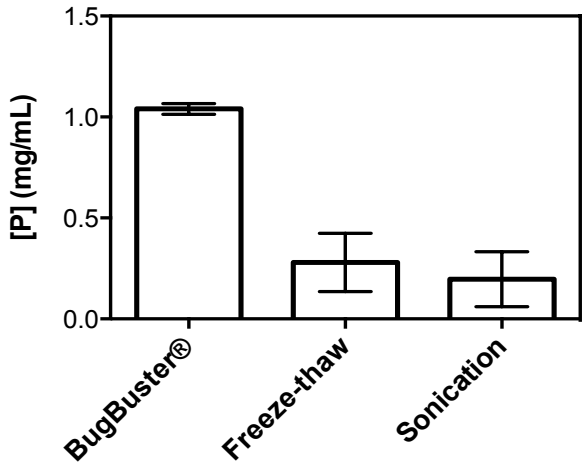


Figure 12.10. Study of cell lysis method. a) Bugbuster Reagent®: the cell pellet was resuspended with 250 µl of Bugbuster reagent with shaking at 650 rpm for 20 min at rt. The lysate was centrifuged at 3500 rpm for 1 h and the pellet discarded. b) Freeze-thaw cycles: the cell pellet was frozen by submerging the microplate in a dry-ice/ethanol bath for 2 min. Then it was thawed by transfer to an ice/water bath for 8 minutes. The cycle was repeated 10 times. The lysate were resupended in 250 µL PBS pH 8.5, centrifuged at 3500 rpm for 1 h, and the pellet discarded. c) Sonication: the cell pellet was resuspended in 250 µL PBS pH 8.5 and lysed by sonication during 7 minutes (50% amplitude, 5 s ON, 15 s OFF, 2 mm diameter probe). The lysate was centrifuged at 3500 rpm for 1 h and the pellet discarded. For the three method, the protein in the clarified cell-free extracts was quantified using the BCA assay [200]**. All the procedure was done in triplicate.**

Cell lysis using Bugbuster releases 3 times more protein to the media than sonication and frezze-thaw cycles. BugBuster is more expensive than other classical methods, but protein lysis is a key step to maximizes the protein released to the media which can be a limiting factor in a library of mutants due to the different expression levels. Cell lysis of the cultures grown on 2 mL deep-well plates was optimized using a Bugbuster commercial lysis cocktail

(Figure 12.10) for final screening protocol. Moreover, the presence of BugBuster buffer does not affect the overall activity of the enzyme after lysis.

12.2.6. Development and Validation of final HTS assay.

High-throughput screening is an essential process in drug discovery or directed evolution programs. The ability to identify true active compounds depends on the high quality of assays and proper analysis of data. An assay is typically designed to be as fast, as reliable, as sensitive, and as suited to the objective as possible, while keeping a good compromise between costs and quality of the assay. In this regard, the Z-factor (Equation 12.1) is a useful dimensionless statistical parameter that combines data variability and signal dynamic range to provide a quality score for a certain assay.

Once basic parameters were studied and determined in previous sections, a final protocol for HTS assay needs to be defined. Consequently, a sensitive activity assay for the CMB-immobilized CDA enzymes using COS substrates with DP from 2 to 6 in HTS format in 96-well microtiter plates was studied. Methods to quantify acetate release or amine formation were evaluated for sensitivity, reproducibility, and automation on a liquid handling platform (Bravo Agilent). Acetate detection using a coupled enzyme assay has been discarded due to the low capacity to be implemented in a HTS. In consequence, amine quantification by reaction with fluorescamine and fluorescence readout proved to be the best method for HTS implementation (Figure 12.11). *Vc*CDA has maximum activity on diacetylchitobiose (DP2) as substrate, and a 10-fold lower activity on tetraacetylchitotetraose (DP4). Since the assay will be used to screen libraries for gain-of-function or increase of activity, assay optimization was performed with the slow-reacting DP4 substrate to define a wider dynamic range. Using a fixed concentration of the fluorescamine reagent, the volumes of cell-free extract and CMB were varied to set the working conditions, Cell-free extracts of cultures expressing wt VcCDA, an inactive D39S mutant [185], wt CD (catalytic domain with no CBMs), and a void plasmid (NoP, no CDA expression) were incubated with conditioned CMBs in PBS buffer, DP4 substrate (2 mM) was added, and deacetylated product formation quantified by reaction with fluorescamine testing different situations (see protocol 21.1.9. modified).

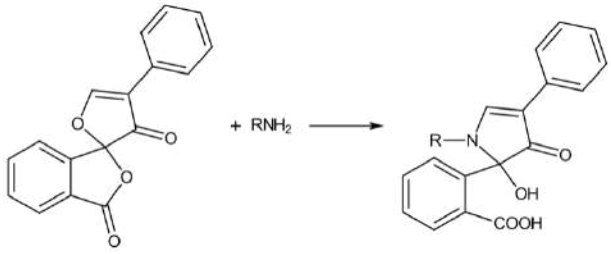


Figure 12.11. Reaction between fluorescamine and primary amine.

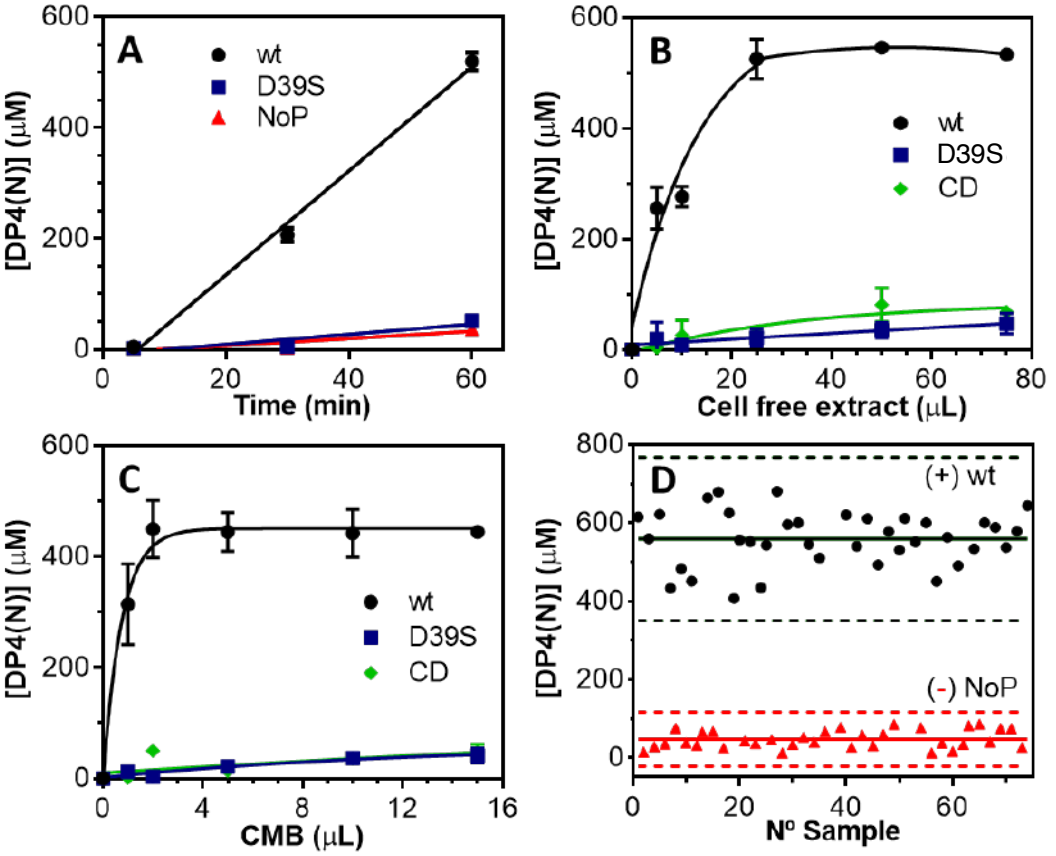


Figure 12.12. Deacetylase activity by the fluorescamine assay using the HTS protocol. A) Kinetics of deacetylated product (DP4(N)) formation). B) DP4(N) formation at 60 min reaction, varying the cell-free extract (5-75 μL) at fixed CMB (10 μL). C) DP4(N) formation at 60 min reaction, varying CMBs (1-15 μL) at fixed cell-free extract (25 μL). D) Z'-factor calculation. DP4(N) formation at 60 min reaction using final conditions (10 μL CMB, 25 μL cell-free extract) for 75 repetitions of wt and inactive D39S mutant. Solid lines, average value; dotted lines, ± 3 SD (standard deviation). All deacetylation reactions were done at 37°C, PBS pH 8.5.

The maximum difference between the deacetylase activity of active wt *Vc*CDA Full-length (FL) protein and the inactive mutant D39S and a void plasmid is at 60 min of reaction (Figure 12.12.A). The use of 20 to 80 μl of cell-extract of *Vc*CDA FL and 2 to 16 μl of CMB (previously

conditioned with PBS and washed after binding step) shows the highest activity compared to the inactive D39S mutant and the catalytic domain of *Vc*CDA (with no CBMs) at 60 min of reaction. The hypothesis that the binding of the protein *Vc*CDA FL to the CMB is given by the CBM is demonstrated in Figure 12.12.B. The catalytic domain protein does not show any activity after the binding and washing steps concluding that the CBMs are essential for the interaction with chitin polymer present in CMB to immobilize the enzyme.

The final optimized and validated protocol is given in the Figure 12.13. For the final HTS assay adapted to 96-well plates, different features were studied and optimized. Protein is expressed using autoinduction method in 1.5 mL in 96 deep-well microplates. Cell lysis is done using a commercial buffer BugBuster. After lysis, 25 μl of cell extract were transferred to a new plate (final volume of 150 μl) and mixed with 10 μl of the CMB suspension in PBS previously conditioned. The 96-well microtiter plates were incubated on an shaker (Eppendorf MixMate) at 650 rpm, 4°C for 1 h to allow protein binding to the CMB. Washes of the chitin magnetic beads (protein immobilized on the CMB) were performed using a magnet platform. The reaction plate with CMB-immobilized proteins and the containers with PBS buffer and substrate solutions were preheated 10 min to the corresponding reaction temperature. 70 μl of DP4 2 mM were added to the complex protein-CMB and the reaction plate was placed on an Eppendorf MixMate at the reaction temperature, and shaken at 650 rpm for 1 h. The deacetylase activity was quantified using 20 μL fluorescamine as reagent (2 mg/mL final concentration) to detect the formation of primary amines. The deacetylase reactions were stopped by transferring 40 μl of reaction supernatants from the reaction plate (applying the magnet to settle the beads) to a black 96-well microtiter plate (for fluorescence reading). Each plate contains a set of glucosamine standards, positive controls (*Vc*CDA FL wt) and negative controls (*Vc*CDA FL D39S). Detailed protocol is presented in below (see protocol 21.1.9.).

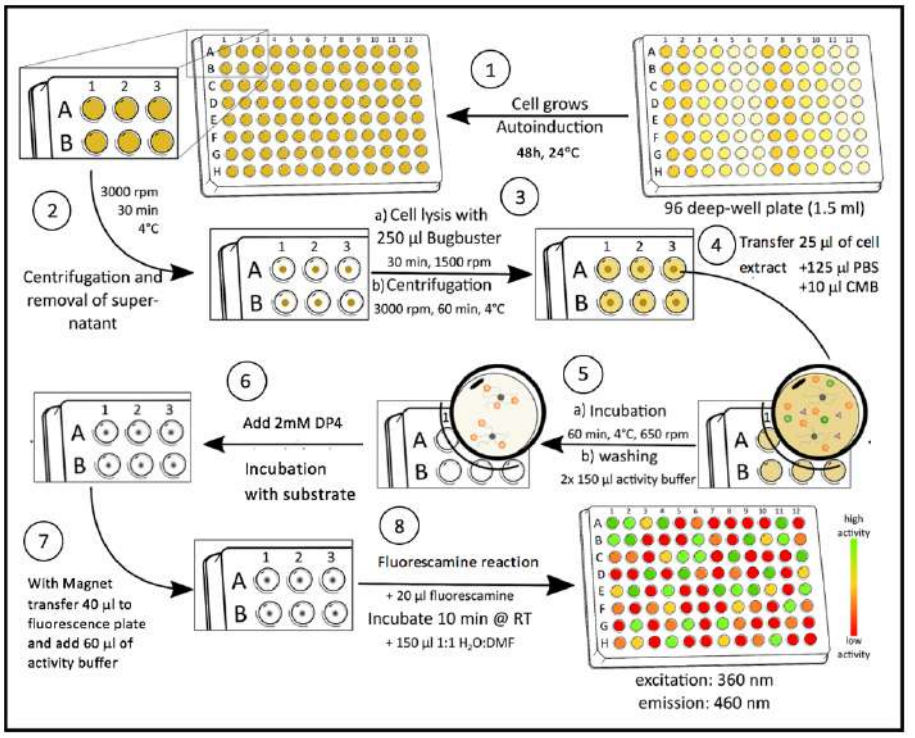


Figure 12.13. HTS protocol of mutant libraries.

By far the most common quality metrics reported for HTS assays validation are Z- and Z′-factor. Z′-factor is often used during assay optimization because it is based on controls, while Z-factor may be used during screening to assess performance of the screen on actual samples (Equation 12.1).

Equation 12.1. Z-factor equation. Equation is defined by: the means (μ), and standard deviation (σ), of both the positive (p) and negative (n) controls.

$$Z - factor = 1 - \frac{3(\sigma_p + \sigma_n)}{|\mu_p - \mu_n|}$$

The final method is able to discriminate between positive (active wt enzyme) and negative (void plasmid) levels with a calculated Z'-factor of 0.5 (Figure 12.12.D), which reflects a good quality for an HTS assay.

On the other hand, in measuring the quality of experiments, repeatability and reproducibility are key parameters. To study the reproducibility, the HTS protocol was applied to six glycerinates of *E.coli* BL21(DE) cells harboring the pET22b (wt *Vc*CDA) expression plasmid. After cell lysis, the cell-free extracts from each culture were split in six aliquots of 25 µL, and

the protocol was continued for each of the 6x6 samples. In the graph, the concentration of DP4(N) product formed is given for each of the 6 aliquots (repetitions) from each of the 6 cultures (X-axis). The mean values (of 6 repetition) and coefficient of variation (CV%) for each culture are given (see protocol 21.1.10.).

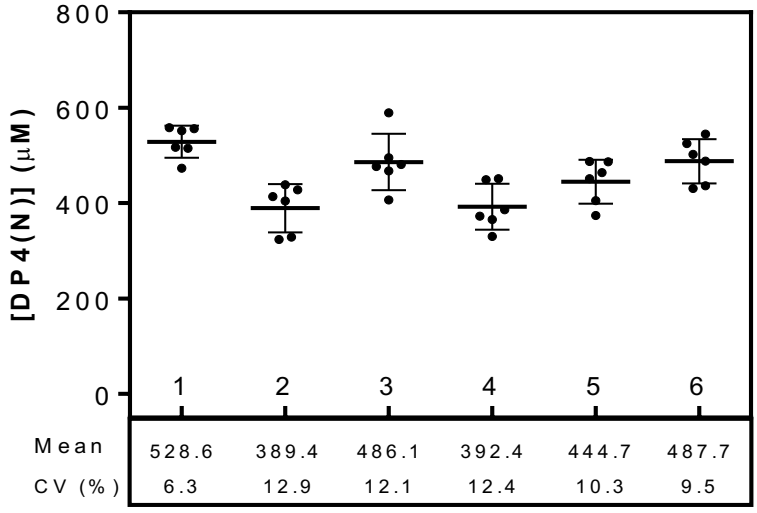


Figure 12.14. Reproducibility of HTS assay applied in 6 cell cultures.

The reproducibility of the assay had a coefficient of variation (CV) in the range of 6 to 12% for different independent samples (Figure 12.14) confirming the validation of the HTS assay developed. For CV lower than 15%, the final screening assay is approved as reliable in terms of reproducibility.

12.2.7. Application to *Vibrio cholera* deacetylase mutant libraries for engineered specificity.

The HTS assay was applied to screen mutant libraries of the model *Vibrio cholera* enzyme in the framework of a directed evolution program. *Vc*CDA is highly specific for deacetylating the penultimate GlcNAc residue from the non-reducing end of COS, with DP2 being the most active substrate, and activity rapidly decreasing with increasing DP (Figure 9.16). For applications in biocatalysis we were interested in engineering the enzyme to have higher k_{cat} values for longer substrates (DP4 and DP5) but maintaining the high specificity for the deacetylated position and maintaining the high activity towards short substrates like DP2.

12.2.7.1. Development of library screening strategy.

To this end, random mutant libraries of the catalytic domain of *Vc*CDA (CE4 deacetylase domain) were prepared by error-prone PCR (epPCR) using GeneMorph kit from Agilent at low mutation frequency (1-2 bp per kb) and subcloned into a recipient plasmid consisting of a pET vector containing the sequence of the two CBM12 modules to be fused at the C-terminus of the catalytic domain (Figure 12.15.) using CPEC technique.

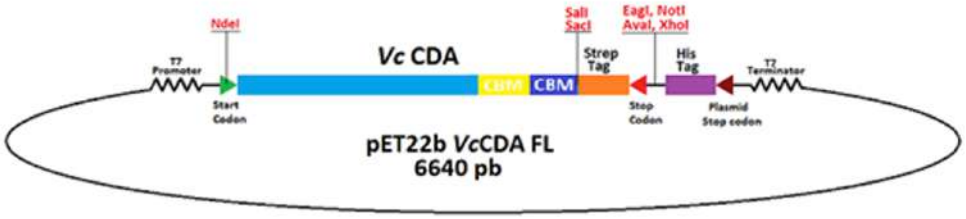


Figure 12.15. Plasmid map of VcCDA FL. Map of the original vector in which the complete VcCDA proteins with the Strep-TagII sequence at the C-terminal end were cloned. Library of mutants were clone in a vector containing the CBM.

CPEC is one of many sequence-independent cloning techniques used to assemble multi-fragment complex libraries in a one-step PCR reaction. It uses a single enzyme, making it inexpensive and fuss-free compared to other strategies that use multiple enzymes and steps (Figure 12.16).

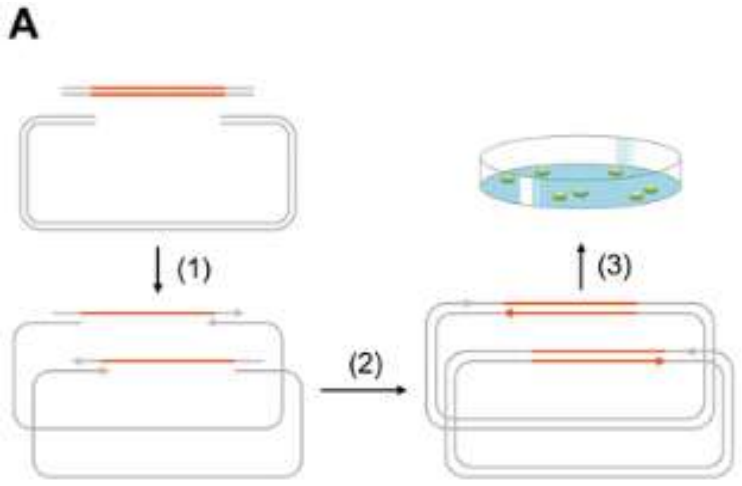


Figure 12.16. CPEC technique. Linearized vector and inserts (with overlapping sequences) are added to a single reaction tube where PCR occurs with the help of a high-fidelity polymerase. The vector and insert use each other as templates to give cloned products. Adapted from [202].

Primers for insert and vector obtention were designed taking into account that mutagenesis target sequence starts after ATG codon start and finish in the last proline of catalytic domain (Figure 12.17). This strategy was designed to avoid the randomization of CBMs and only include the catalytic domain.

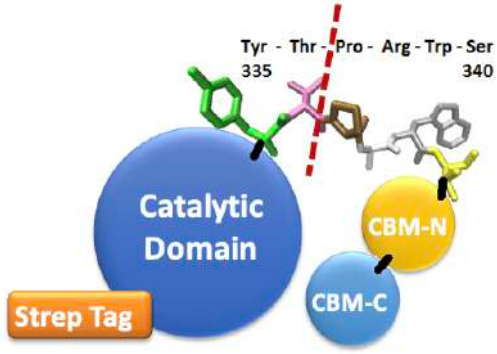


Figure 12.17. Linker between CBMs and catalytic domain of *Vc*CDA. Scheme of the WT FL structure in which the T336 cleavage site is indicated.

Different PCR conditions were optimized for CPEC cloning [203] (protocol 21.1.11.). After expression of the library, selective pressure to screen for more active mutants on the DP4 substrate was introduced by performing the activity step in the HTS protocol at a higher temperature. Previously, temperature profile of the CMB-immobilized wt *Vc*CDA FL was studied to determine the residual activity of the complex in the range of temperatures between 37ºC to 70ºC. To study the temperature profile, 25 μL of cell extract were mixed with CMB in a final volume of 150 μL. The protein-CMB complex was washed and reactions were initiated by addition of 70 μL of DP4 (2 mM, final concentration) at different temperatures (37ºC-70ºC). At 60 min reaction time, deacetylase activity was quantified using fluorescamine. Fluorescence was measured on a fluorescence reader (see protocol 21.1.12).

The temperature profile of the CMB-immobilized wt *Vc*CDA (Figure 12.18) showed that the enzyme retained less than 40% residual activity at 50°C compared to its activity at 37°C. By carrying out the activity step at 50ºC in the HTS protocol, the wt enzyme becomes the negative level and active mutants will show up either because they are more thermostable or more active with higher residual activity at 50°C obtaining a mutant with higher activity.

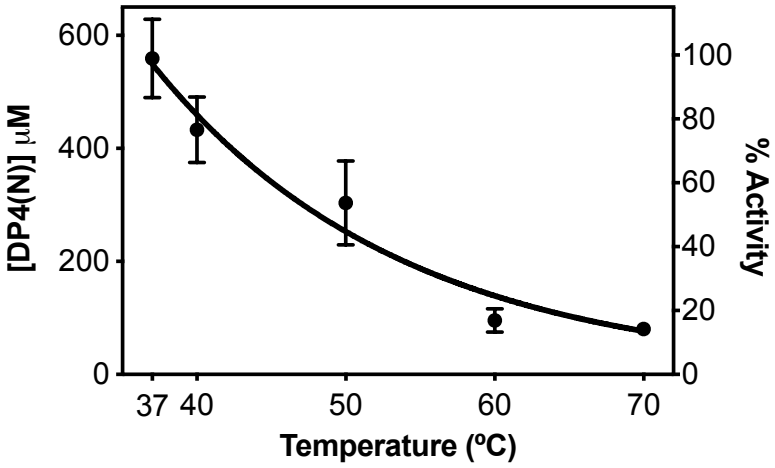


Figure 12.18. Temperature profile of CMB-immobilized wt *Vc*CDA.

12.2.7.2. Library screening of *Vc*CDA and mutant characterization.

Around 400 mutants were screened for the first generation and rendered few active mutants from which mutant G1 had a significantly higher activity (Figure 12.19.A). Screening assay was done following the scheme presented in figure 4.12. Gene sequencing showed that the G1 mutant had a single amino acid substitution, K275E. This mutant increases deacetylase activity 2.5 fold relative to wt enzyme. Afterwards, mutant K275E has been selected as a template for a second round of epPCR following the same procedure, this time selecting for active mutants at a higher temperature of 60°C (Figure 12.19.B). More than 400 mutants were analyzed in this second generation of mutants. Of the few active mutants, the most active (G2) contained two mutations (K275E/H127R). Both G1 and G2 mutants were recovered, expressed, and the proteins purified for biochemical characterization. All screenings were done following the protocol 21.1.9.

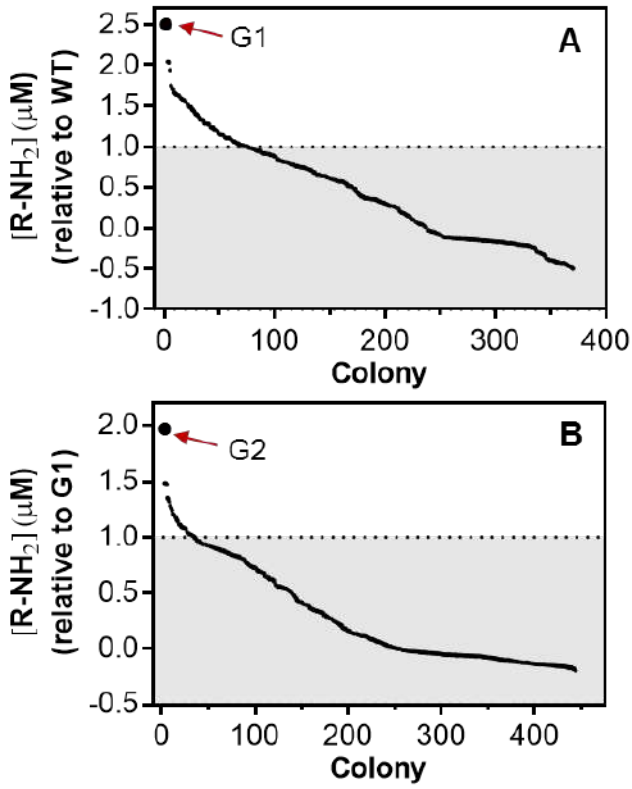


Figure 12.19. Random libraries by epPCR. A) First generation, template: wt enzyme, screening for activity at 50°C. G1= *Vc*CDA K275E mutant B) Second generation, template: G1 mutant, screening for activity at 60°C. G2= VcCDA K275E-H127R.

Protein expression of wt *Vc*CDA, *Vc*CDA K275E and *Vc*CDA K275E-H127R in preparative culture was performed by autoinduction media in BL21(DE3) cells (see protocol 21.4.1). For the isolation of the enzyme from the cellular lysate, two chromatographic steps were performed: an affinity chromatography with a StrepTrap column (see protocol 21.4.4) followed by a Size Exclusion Chromarography (SEC) with a Superdex 200 (16/600) column (protocol 21.4.6) as described in CHAPTER 1, section 11.1. Results for Streptag purification of *Vc*CDA FL K275E is highlighted as example (Figure 12.20).

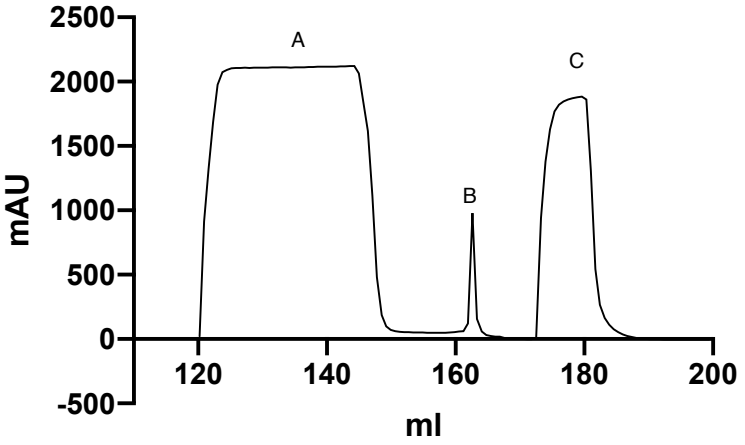


Figure 12.20. Chromatogram of Strep-trap purification of *Vc*CDA K275E mutant. A) loading and washing, B) elution; C) regeneration with HABA and washing. mAU: 280nm.

In order to identify and separate possible oligomeric forms of the expressed protein, a second purification step by SEC was performed. All elution samples from the first affinity chromatography were pooled together, concentrated up to a 2 ml volume and loaded into a Superdex 200 (16/600) column for this second purification step. The elution was performed at a rate of 1 mL/min using PBS (50 mM phosphate, 300 mM NaCl, pH 8.5) as buffer, with the same composition used in the previous steps (Figure 12.21).

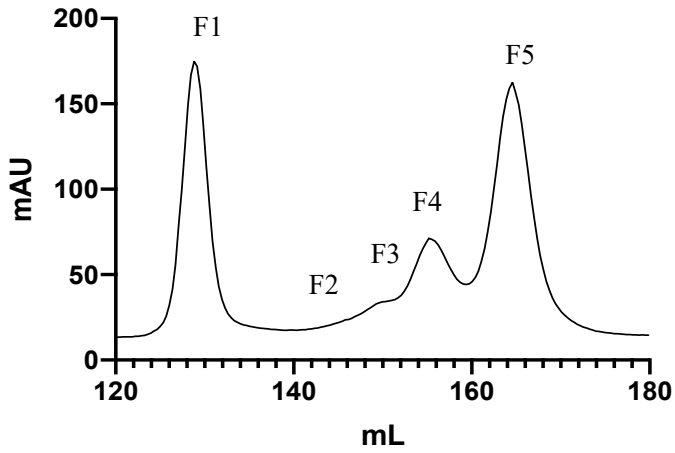


Figure 12.21. Chromatogram of SEC purification of *Vc*CDA K275E. Chromatogram obtained in the gel filtration chromatography using a Superdex 200 (16/600). Five multimeric fractions are shown, from which fraction 4 (F4) corresponds to the dimer and fraction 5 (F5) corresponds to the monomer of the protein.

Five different peaks were observed in the chromatogram (Figure 12.21). Calibration of the column with known standards (see protocol 21.4.6.2.) allowed the identification of each fraction composition. Fraction 1 (F1), fraction 2 (F2) and fraction 3 (F3) probably corresponds to soluble oligomers, fraction 4 (F4), corresponds to the dimeric form of the enzyme, and fraction 5 (F5), eluted approximately 80 minutes after injection) is the monomeric form of *Vc*CDA K275E. Several purifications of the enzyme were carried out during the performance of this work and it was observed that, although the chromatographic profile always presented the same five characteristic fractions, their relative amount slightly varied between different batches. Oligomeric states are closely related to cell and expression conditions but the parameters that affect this phenomenon are unclear. Using the column calibration equation, the value obtained for this enzyme is 39.3 kDa, while the theoretical Mw is 46.9 kDa.

SDS-PAGE analysis (see protocol 21.9.1) was performed with samples corresponding to the different chromatographic steps and it revealed that the recombinant enzyme migrated with an approximate molecular mass of 45 kDa, which agrees with the theoretical mass (46.9 kDa) (Figure 12.22.). The five fractions from SEC show the same band under denaturing SDS-PAGE, confirming the non-covalent oligomeric nature of fractions F4 and F5 (see protocol 21.9.1).

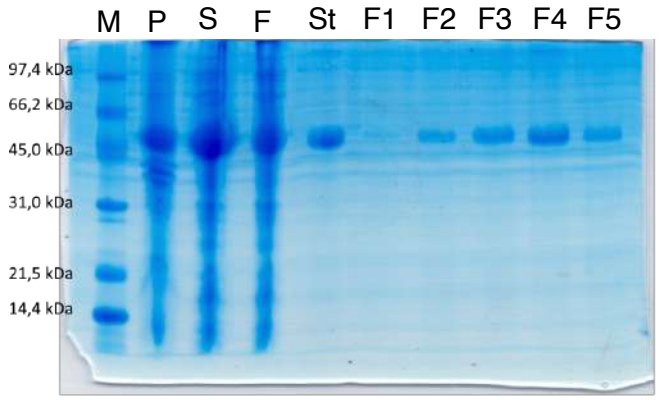


Figure 12.22. SDS-PAGE 14% acrylamide of *Vc*CDA FL K27E purification. M: Marker Low Range (Biorad); P: cell lysate, S: soluble fraction; F: flow through during column loading and washing; St: elution of protein purified by Strep-trap column; F1: fraction 1, F2: fraction 2, F3: fraction 3, F4: Fraction 4, F5: fraction 5 from GFC.

Protein yields for protein expression and purification were determined using BCA as quantification method. For *Vc*CDA wt enzyme 7 mg/L of culture, for *Vc*CDA FL K275E 1.25 mg/L of culture and for *Vc*CDA FL K275E-H127R 1.5 mg/L of culture.

For protein characterization, Michaelis-Menten kinetics were evaluated with diacetylchitobiose (DP2), tetraacetylchitotetraose (DP4), and pentaacetylchitopentaose (DP5) at 37°C and pH 8.5 (Figure 12.23 and Table 12.1) using HPLC-MS method (see protocol 21.5).

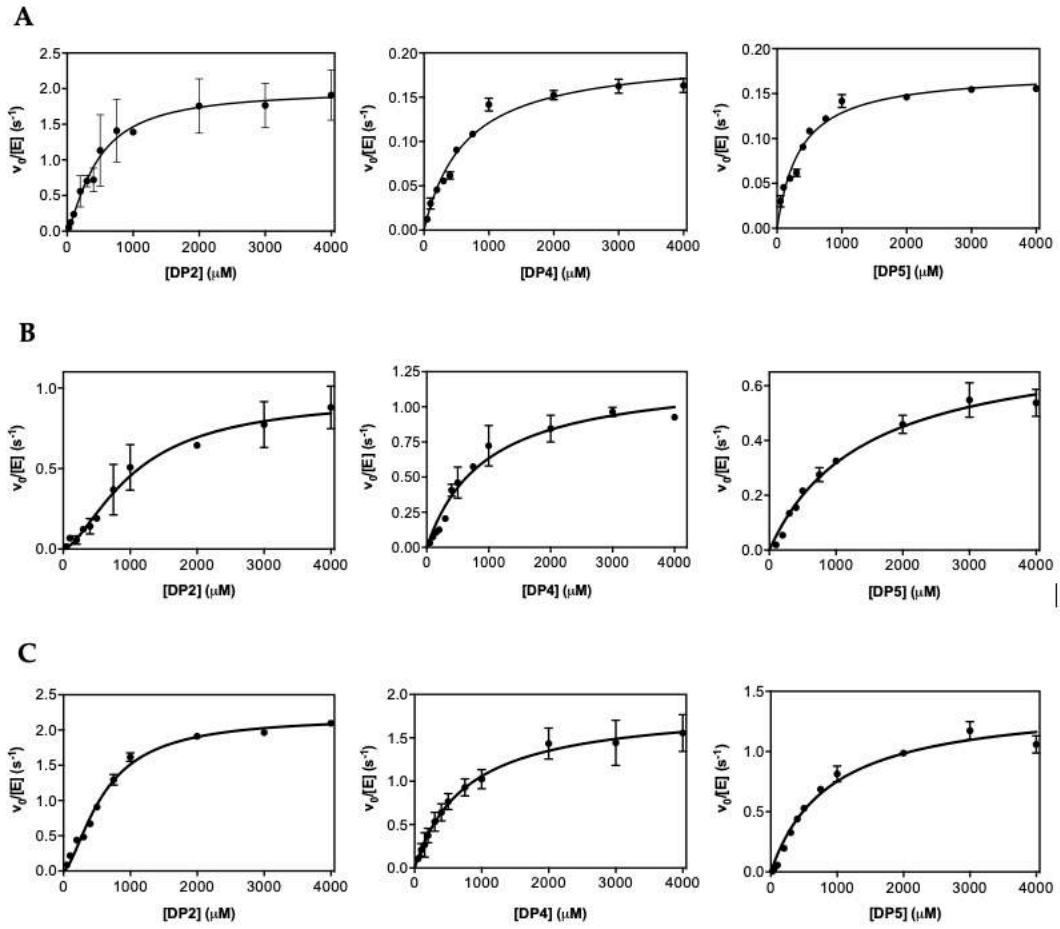


Figure 12.23. Determination of Michaelis-Menten parameters. Kinetics were performed with 3 different substrates, DP2, DP4, and DP5, according to Protocol 4. A) *Vc*CDA WT; B) K275E mutant; C) K275E-H127R mutant. Conditions: 100 μM to 4 mM substrate, 0.05-0.8 μM enzyme, 50 mM phosphate buffer pH 8.5, 300 mM NaCl, 37°C. Initial rates vs. substrate concentration were fitted by non-linear regression to a cooperative Hill model for the DP2 substrate ($v/[E] = k_{cat}\cdot[S]^h / (K_M^h + [S]^h)$), and to a Michaelis-Menten model for DP4 and DP5 substrates. Hill slope value was 1.6 for all the enzymes with DP2 substrate. Data were fitted using Graphpad software (Prism). Kinetic parameters are given in Table 12.1.

Table 12.1. Kinetic parameters of wt and mutant *Vc*CDA enzymes selected from the HTS assay. Conditions: 100 μM to 4 mM substrate, 0.05-0.8 μM enzyme, 50 mM phosphate buffer pH 8.5, 300 mM NaCl, 37°C.

		M&M parameters		
	Substrate	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m ($M^{-1}\cdot s^{-1}$)
WT	DP2	2.02 ± 0.19	0.46 ± 0.10	$3.7\cdot10^3$
	DP4	0.20 ± 0.01	0.64 ± 0.07	$3.1\cdot10^2$
	DP5	0.17 ± 0.01	0.36 ± 0.03	$7.5\cdot10^2$
K275E	DP2	0.94 ± 0.11	1.54 ± 0.24	$5.8\cdot10^2$
	DP4	1.24 ± 0.01	0.95 ± 0.14	$1.3\cdot10^3$
	DP5	0.77 ± 0.04	1.42 ± 0.18	$8.7\cdot10^2$
K275E/H127R	DP2	2.19 ± 0.06	0.60 ± 0.04	$2.9\cdot10^3$
	DP4	1.87 ± 0.10	0.77 ± 0.10	$2.4\cdot10^3$
	DP5	1.42 ± 0.06	0.89 ± 0.10	$1.6\cdot10^3$

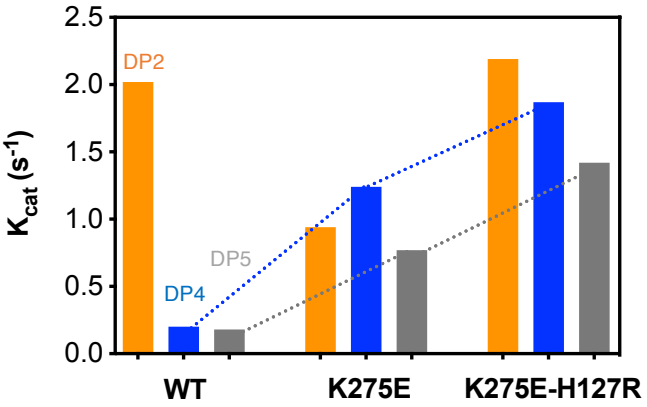


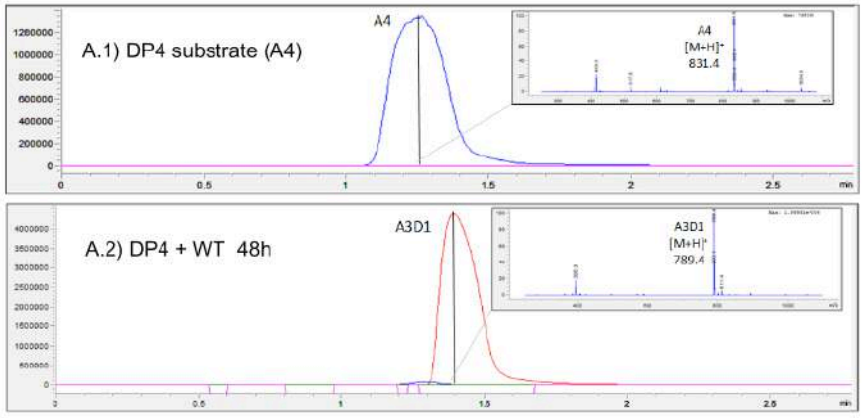
Figure 12.24. k_{cat} values of wt and mutants on DP2, DP4 and DP5 substrates. Activity determined by HPLC-MS method. Orange: DP2, Blue: DP4 and Grey: DP5. Buffer: 50 mM phosphate buffer pH 8.5, 300 mM NaCl, 37°C.

Figure 12.24 plots the k_{cat} values for wt and mutant enzymes with DP2, DP4 and DP5 substrates. The first generation K275E mutant has a slightly reduced k_{cat} for DP2 (2-fold lower) but has significantly higher k_{cat} values for the longer substrates DP4 and DP5 (6 and 4.5-fold relative to the wt enzyme) (Table 12.1.). The K_M values have increased, about 4-fold for the DP2 and DP5 substrates, but less than 2-fold for the DP4 substrate. The main effect achieved with this first generation K275E mutant is an improved activity on longer substrates in terms of kcat without significantly reducing the activity on the natural DP2 substrate of *Vc*CDA, with a significantly modified k_{cat} ratio DP2:DP4:DP5 from 10:1:1 for the wt to 10:13:8 for the mutant (Figure 12.24.). The second round of mutagenesis introduced an additional mutation to afford the K275K/H127R mutant as best hit. Remarkably, k_{cat} values for the DP4 and DP5 substrates have further increased up to 9-fold relative to the initial wt enzyme, but retaining the wt activity

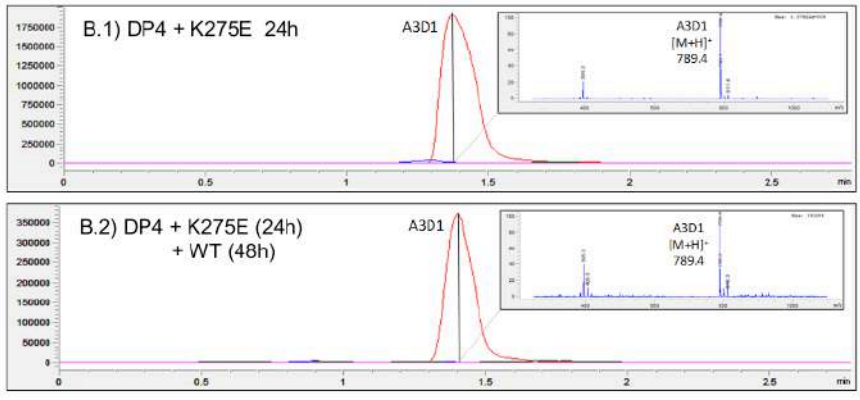
on DP2. The better binding reflected by the lower K_M values relative to the first mutant also resulted in improved k_{cat}/K_M values.

Furthermore, the deacetylation pattern was studied for both mutants. For this purpose, fully acetylated chitotretraose was treated for 24h reaction with *Vc*CDA K275E and *Vc*CDA K275E-H127R mutants. Afterwards, the degree of acetylation of each sample was analyzed by HPLC-MS. Finally, *Vc*CDA wt was added to both samples and let it react for 24h and the degree of acetylation was determined using HPLC-MS (see protocol 21.5).

A) *Vc*CDA wt



B) *Vc*CDA K275E



C) *Vc*CDA K275E/H127R

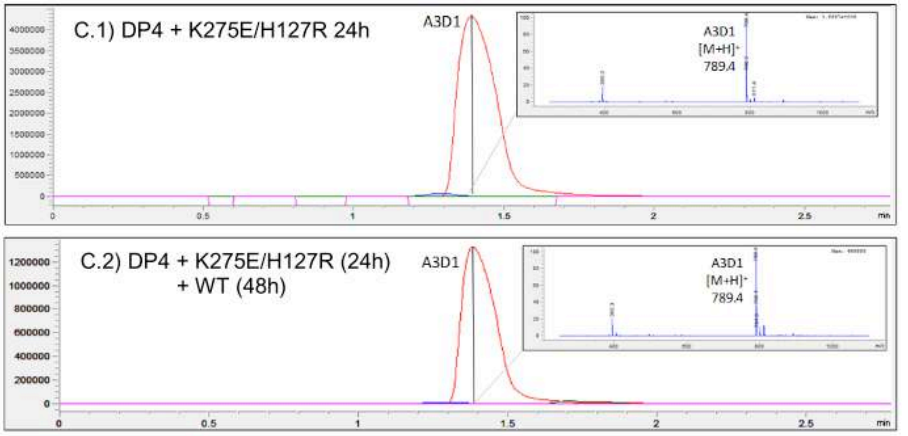


Figure 12.25. Deacetylation specificity of K275E and K275E-H127R mutants with DP4 substrate. A.1) DP4 substrate (A4, tetraacetyl-chitotetraose). A.2) DP4 incubated with WT VcCDA for 48h, giving the monodeacetylated product A3D (ADAA). B.1) K275E + DP4 reaction for 24h. B.2) Reaction B.1 incubated with wt VcCDA for 48h. C.1) K275E/H127R + DP4 reaction for 24h. C.2) Reaction C.1 incubated with wt VcCDA for 48h. Reaction conditions: 8 mM substrate, 1 μM enzyme, PBS buffer (50 mM K_2HPO_4, 300 mM NaCl, pH 8.5), 37 °C.

No dideacetylated products were obtained in B.2 and C.2 reactions (Figure 12.25), confirming the same pattern of deacetylation (ADAA) for wt and the mutants. The MS spectrum of each peak is given as insets. This structural characterization of the reaction products confirmed that the pattern of acetylation using DP4 as a substrate has not been modified in mutants *Vc*CDA K275E and *Vc*CDA K275E-H127R.

12.2.7.3. Temperature profile, Thermostability and Thermotolerance characterization.

As already stated, selection pressure was introduced by increasing temperature of the deacetylation reaction. 50ºC and 60ºC were the temperatures for first and second round of the screening assay. However, kinetic parameters of purified enzymes were determined at 37ºC (standard conditions for kinetic characterization) (Figure 12.27). Notwithstanding, the thermostability, temperature profile and thermotolerance of the each mutant were analyzed in detail to determine if some changes in temperature-related properties had been introduced by this mutations.

Thermostability of each mutant was determined using Sypro Orange (see section 11.3.2). This methodology was applied in the presence and absence of Zn^{+2} (0.1 mM). 5 μL of Sypro Orange solution were added to 20 μl of protein solution (4 μM final concentration) in 50 mM

phosphate buffer pH 8.5, 300 mM NaCl. A thermocycler Rotogene 3000 (Corbett Research) was used to quantify the fluorescence (see protocol 21.1.13.). Tm (°C) values are given in Figure 12.26 and Table 12.2.

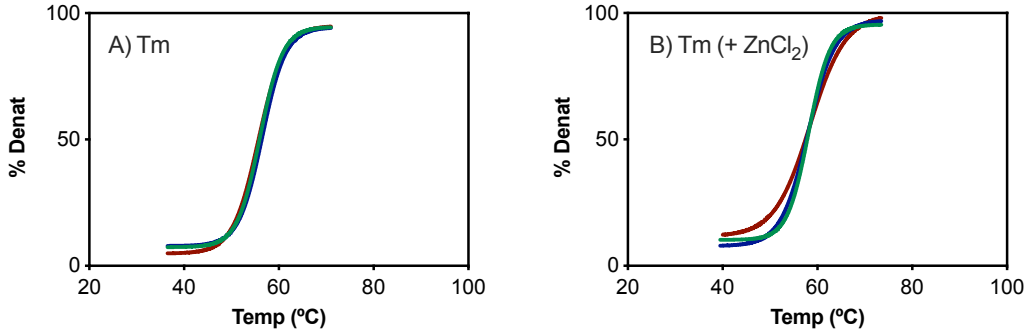


Figure 12.26. Thermal stability of wt and mutant *Vc*CDA proteins in presence or absence of Zn. A) Melting temperature: *Vc*CDA wt (green): 56.0 ± 0.17, *Vc*CDA K275E (red), *Vc*CDA K275E-H127R (blue). B) Melting temperature in presence of $ZnCl_2$: *Vc*CDA wt (green), *Vc*CDA K275E (red), *Vc*CDA K275E-H127R (blue)

Table 12.2. Thermal stability of wt and mutant *Vc*CDA proteins in presence or absence of Zn^{2+}.

Enzyme	Melting temperature	
	Tm (ºC)	Tm (ºC) $ZnCl_2$
***Vc*CDA wt**	**56.0 ± 0.17**	**58.2 ± 0.2**
***Vc*CDA K275E**	**56.5 ± 0.14**	**58.0 ± 0.2**
***Vc*CDA K275E/H127R**	**55.7 ± 0.11**	**58.5 ± 0.4**

In figure Figure 12.26 and Table 12.2., it has been pointed out that presence of Zn^{2+} increases the melting temperature in wt and both mutants. This phenomena could be explained through the role of the catalytic metal structure in the stabilization of the enzyme. In wt and mutants the affinity for Zn^{+2} was reduced and the enzymes were not saturated with the cation resulting a decrease in Tm. On the other hand, in presence of Zn^{2+} the thermal stability of the wt and mutants is similar showing almost no difference. It should be stressed that the mutant *Vc*CDA K275E shows the highest Tm with no presence of Zn^{2+}. To sum up, single mutant K275E and double mutant K275E/H127R has not a significant impact on Tm.

Additionally, temperature profile defines the specific activity of *Vc*CDA wt, *Vc*CDA K275E and *Vc*CDA K275E-H127R mutants for deacetylase reaction using DP4 (2 mM) as a substrate at different temperatures (range between 30ºC to 60ºC) (see section 11.3.1). Specific activities

were determined using HPLC-MS. Temperature profiles determine the optimal temperature for deacetylase activity (Figure 12.27. and Table 12.3.) (see protocol 21.5).

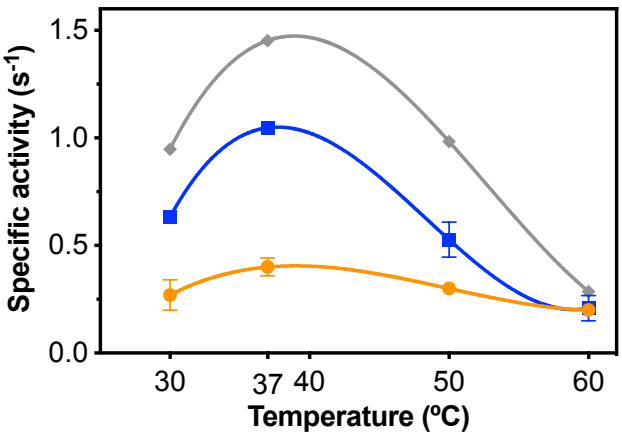


Figure 12.27. Temperature profiles with DP4 substrate. Activity determined by HPLC-MS method. Orange: VcCDA wt, Blue: VcCDA K275E, Grey: VcCDA K275E-H127R. Buffer: 50 mM phosphate buffer pH 8.5, 300 mM NaCl, 37°C. Kinetic parameters are given in Table 4.3.

Table 12.3. Specific activity (s^{-1}) for different mutants at different temperatures. Reaction conditions: Buffer: 50 mM phosphate buffer pH 8.5, 300 mM NaCl, 37°C.

	Specific activity (s^{-1})		
Temperature (Cº)	**WT**	**K275E**	**K275E-H127R**
30	0.27 ± 0.07	0.63 ± 0.01	0.88 ± 0.10
37	0.40 ± 0.04	1.05 ± 0.02	1.56 ± 0.01
50	0.30 ± 0.03	0.53 ± 0.08	0.98 ± 0.01
60	0.24 ± 0.03	0.18 ± 0.06	0.25 ± 0.17

It should be emphasized that temperature profile in terms of specific activity is not affected by the introduction of the mutations (see Figure 12.27. and Table 12.3.). Maximum activity is still obtained at 37ºC for all enzymes but specific activity was greatly increased in single K275E and double K275/H127R mutants. In Figure 12.3, results highlighted that enzymes *Vc*CDA K275E/H12R and *Vc*CDA K275E shows 3-fold and 2-fold specific activity compared with wt at different temperatures. Mutant *Vc*CDA K275E/H127R shows the highest specific activity at each temperature studied.

On the other hand, the mutation has a positive effect on the thermostability of the enzyme. To study this, the thermotolerance of *Vc*CDA WT (green dots in Figure 12.28), K275E (blue squares), and K275-H127R (red diamonds) was undertaken with free enzymes at 50°C, and with CMB-immobilized enzymes at 50ºC and 60ºC (see secton 11.3.3). In figure Figure 12.28, purified enzymes (2 µM) were incubated in PBS pH 8.5 buffer at 50°C. Then, aliquots were taken at different time intervals, diluted 1:10 in PBS pH 8.5, and kept at 37°C for 10 min (final enzyme concentration 0.2 µM). Residual activity was determined using DP4 substrate at 2 mM concentration (see protocol 21.1.14.). For determination of the thermotolerance in CMB-Immobilized enzyme complex (Figure 12.28.B and C) (see protocol 21.1.14).

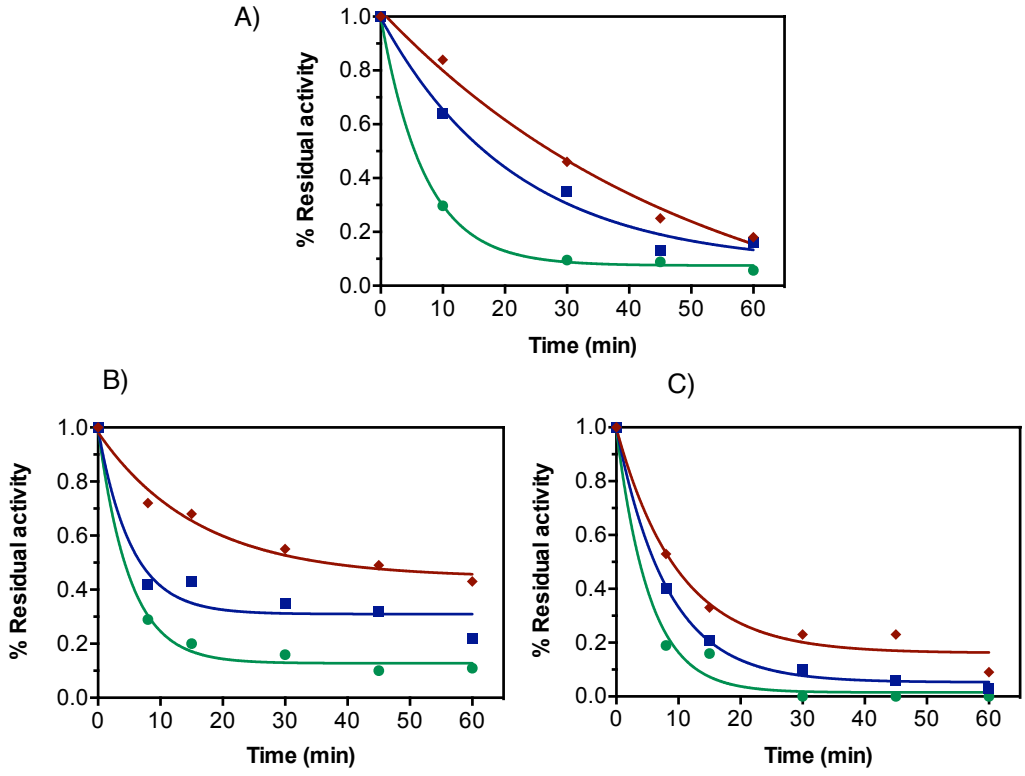


Figure 12.28. Thermotolerance of free and immobilized enzymes. The thermotolerance of *Vc*CDA WT (green dots), K275E (blue squares), and K275-H127R (red diamonds) was determined with free enzymes at 50°C (A), and with CMB-immobilized enzymes at 50ºC (B) and 60ºC (C). Exponential decay equation has been adjusted (Equation 21.1).

Table 12.4. Thermotolerance of free enzymes at 50°C and complex CMB-Enzyme at 50°C and 60°C. inactivation (K_i) constant and half-live (t_{50} min) monitoring residual activity with DP4 substrate (HPLC-MS).

		*Vc*CDA wt	*Vc*CDA K275E	*Vc*CDA K27E/H127R
t_{50} at 50°C (free E)	K_i (min^{-1})	0.14 ± 0.01	0.05 ± 0.01	0.03 ± 0.01
	$t_{1/2}$ (min)	5.4	16.8	24.1
t_{50} at 50°C (CMB-P)	K_i (min^{-1})	0.18 ± 0.04	0.19 ± 0.07	0.06 ± 0.01
	$t_{1/2}$ (min)	3.9	3.8	11.1
t_{50} at 60°C (CMB-P)	K_i (min^{-1})	0.19 ± 0.04	0.12 ± 0.04	0.10 ± 0.02
	$t_{1/2}$ (min)	3.7	5.7	6.8

In our work, in free enzyme state (Figure 12.28. and Table 12.4.), single mutant K275E and double mutant K275E/H127R exhibited more thermotolerance in terms of Ki and half-life time at a 50ºC than the wt enzyme. Mutant K275E/H127R show the highest half-life time with almost 5-fold the wt residual activity at 50ºC yielding a mutant highly thermostable compared with wt enzyme. Mutants obtained clearly have a positive effect on thermotolerance. No thermodynamic properties has been affected by these single and double mutations, only the kinetic parameters are different making an enzyme more productive at different temperatures with higher ability to deacetylate DP4.

Previous articles show that immobilized enzymes have enhanced stability compared to soluble enzymes [204]. Remarkably, comparing the experimental values of $t_{1/2}$, for free and CMB-Enzyme complexs in Table 12.4., unlike expected, the immobilized *Vc*CDA wt and mutants presented half-lives lowers than shown by the free enzymes. This effect is more pronounced in the double mutant since the half-life time is reduced to more than half in the immobilized enzyme with respect to free enzyme. This results could be obtained due to the detrimental effect of the CMB on the protein after the heating step. On the other hand, heating the CMB can release traces of metal or some interference that affect the final enzymatic activity.

Taken altogether, the temperature profile of both mutants showed the same optimum temperature (37°C, in Figure 12.27.) and a similar thermal stability (Tm 56±0.5°C, see figure 4.26 and table 4.2.) than the wt enzyme. However, the mutations conferred higher thermotolerance to the proteins, with an increase of their half-live (t_{50}) at 50°C relative to the wt enzyme (table 4.2.). Therefore, the selective pressure imposed by screening the mutant libraries at high temperatures, where the wt enzyme just retained residual activity, resulted in the selection of mutants with combined higher specific activity and higher thermotolerance.

12.2.7.4. Functional role of the mutations in *Vc*CDA 3D structure

The crystallographic structure of *Vc*CDA in complex with DP3 substrate (PDB 4OUI) is shown in Figure 12.29. [185]. As already stated, Loop5 (green) blocks the access of the substrate to the positive (reducing-end) subsites. Mutation K275E is located in Loop 5 and mutation H127R is located on the surface of the enzyme. The interaction E270···K275 contributes to the rigidity of loop 5.

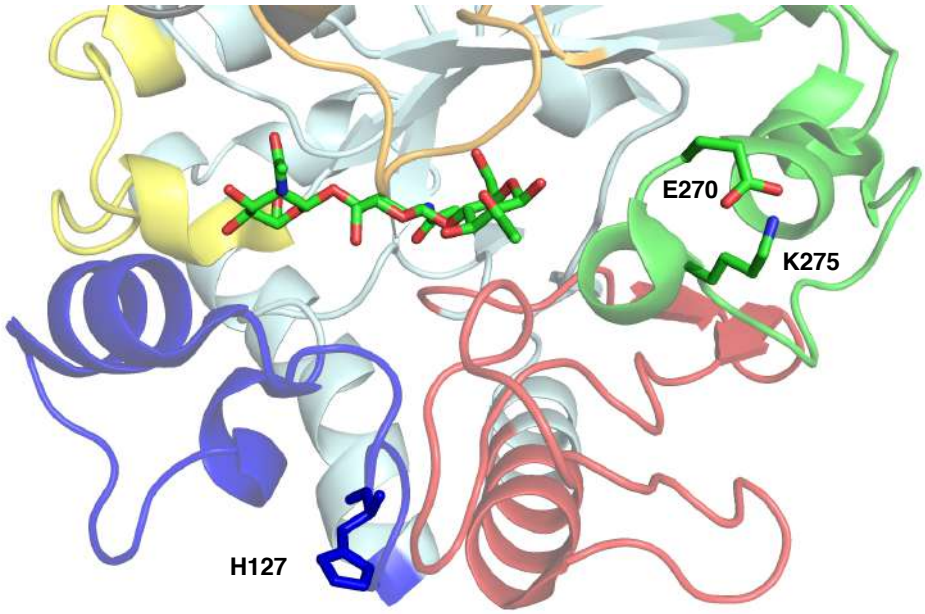


Figure 12.29. Mutated residues in the 3D structure of *Vc*CDA.

The interaction is disrupted upon mutation of K275, enabling the partial opening of the loop to expose additional positive subsites for binding of longer substrates. In further chapters, the role of loop 5 and its dynamics have been elucidated using NMR techniques. On the other hand, H127 is solvent exposed position, but it belongs to Loop2 that also shapes the binding site cleft.

Inspection of the mutants location in the 3D structure of *Vc*CDA (PDB: 4NZ1, 4OUI) [185] suggests that the Glu for Lys replacement at position 275 disrupts a salt bridge with Glu270 that may result in a higher flexibility of an active site loop that blocks the access of longer substrates that DP2 to the reducing-end subsites of the binding side cleft in the wt enzyme. Therefore, the mutation may allow the partial opening of the loop to accommodate the DP4

and DP5 substrates with the concomitant increase of activity (Figure 12.24). The second mutation (H127R), however, is surprising since His127 is solvent-exposed on the surface and farther away from the binding site. This mutation does not significantly alter the specificity but increases the overall activity of the enzyme. Future work (other mutations and molecular dynamics simulations) need to be conducted to understand these effects on specificity and activity.

On the other hand, regarding the mutants obtained, single mutant K275E clearly suggest that the dynamics and mobility of loop 5 plays a pivotal role in the deacetylate activity of the *Vc*CDA towards long substrates such DP4 and DP5. Mutation of K275E lay the foundation for the demonstration of subsite capping model previously proposed where the loops surrounding the active site define the substrate specificity of the enzyme. In next chapters, the role of loop 5 and its dynamics have been elucidated using rational design and NMR studies.

Besides the molecular insight of mutants, we can conclude that we implemented a convenient and sensitive HTS assay to screen directed evolution libraries of CDAs and other CE4 enzymes active on COS. The assay easily allows the adjustment of the dynamic range to different activity levels by modifying the reaction time, temperature and/or volume of the cell extract in the activity step to adapt the signal intensities to the desired positive and negative levels when using different COS substrates or even partially deacetylated substrates.

CHAPTER 3. ENGINEERING LOOP FLEXIBILITY TO TUNE ENZYME SPECIFICITY IN *Vc*CDA

Part of this Chapter has been published in:

Modulation of loop dynamics alters enzyme specificity and switches substrate binding from conformational selection to induced fit mechanisms
Manuscript in preparation
S. Pascual, H.Aragunde, G.Bernardo, J.Jiménez-Barbero, X.Biarnés, O.Millet, A.Planas

13. CHAPTER 3. Engineering loop flexibility to tune enzyme specificity in *Vc*CDA.

13.1. Introduction and Project statement.

The binding of ligands by macromolecules is essential to a multitude of relevant physiological processes. Therefore, considerable effort has been expended in assessing the detailed mechanism by which ligand binding and conformational changes are synchronized. Two defined mechanisms are generally proposed: (i) ‘‘conformational selection,’’ whereby the free enzyme already exist in two conformations under slow equilibrium and the ligand selectively binds to a form of the macromolecule that is present only in small amounts (minor proportion), eventually converting the macromolecule to the ligand-bound conformation; and (ii) ‘‘induced fit,’’ whereby ligand binds to the predominant free conformation followed by a conformational change in the macromolecule to give the preferred ligand bound conformation. Both models describe the allosteric effect as a binding event at one site that induces a conformational change affecting the activity at another site. Alternatively, another model called independent dynamic segments has been proposed, i.e. where protein segments with dynamics distinct from the rest of the protein could be key contributors to binding processes.

The principal challenge is to identify protein binding mechanisms based on experimental data using kinetic and NMR evaluation. Advanced NMR experiments and single-molecule spectroscopy can reveal higher energy conformations that are necessary for conformational-selection or induced-fit binding, permitting us to define a potential mechanism.

In chitin de-*N*-acetylase family, the Subsite Capping Model statement presented in the introduction mentioned that (section 9.3.3.3): "The deacetylation pattern and substrate preference of CE4 family enzymes are governed by the limitations imposed by specific loops and their dynamic". However, it’s based on observations of X-ray structures and on conclusions drawn from the results obtained in the biochemical characterization of *Vc*CDA wt. It must be necessary to provide experimental evidence to support this statement.

The elements, already mentioned in the introduction, on which the Subsite Capping Model was built are the following:

a) The presence of potential subsites that are blocked by the structure of the loops that surround the active site.

b) Loops are not rigid and immobile structures but dynamic structures.
c) The displacement of the loops allows the exposure of subsites necessary for the accommodation of certain substrates.
d) Different deacetylases show very different loops with regard to sequence, size and structure decorating the active sites.

As already stated in section 9.3.3.3, for *Vc*CDA wt enzyme, binding and catalysis events are based on a proposed induced fit mechanism (Figure 9.17). Analyzing the subsites defined with the structure in complex with DP2 (Figure 9.17), it was observed that subsite +2, which is necessary for the accommodation of DP4, is blocked by loop 5 and also for the semi-closed conformation of loop 4 and is not capable of generate enough space for this longer substrate to enter in the active site. Moreover, the crystal structure suggest that the restrictive conditions of the crystallographic ordering of the protein structure have not allowed the movement of loop 5, in a way that the crystallization in complex with DP4 was unsuccessful.

Moreover, in terms of catalytic efficiency, *Vibrio cholerae* chitin deacetylase is highly specific for chitobiose and the catalytic efficiency decreases for longer substrates, while keeping the deacetylation pattern of the products (Figure 9.16). This behavior has been attributed to longer loops surrounding the active site that preclude additional substrate binding pockets in comparison to other chitin deacetylases. In particular, LOOP 3, LOOP 4 and LOOP 5 block the accessibility of chitotetraose, which is deacetylated 40 times slower than chitobiose by VcCDA (Figure 9.16).

Taken altogether, and as already anticipated, these results indicate that flexibility of loops surrounding the active site in chitin deacetylases may also play a key role in substrate specificity and deacetylation pattern of the products. More in-depth approach it has been proposed to understand the role of the loops in *Vc*CDA enzyme.

As a starting study, an in silico approach was carried out on loop 5 suggested that loop 5 can be opened to accommodate a long substrates such as chitotetraose (DP4) (Figure 13.1).

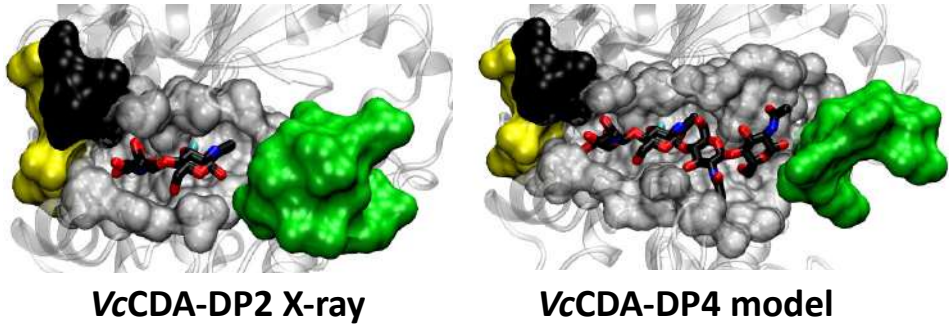


Figure 13.1. Bioinformatic model showing the possible dynamic behaviour of DP5 to allow accommodation of substrates longer than DP3. (Loop 5 in green).

In summary, the bioinformatic model (Figure 13.1) shows how, after forcing the open conformation of loop 5, docking of chitotetrasose is able to produce an energetically stable complex competent for catalysis. Also, wt enzyme show a weak activity for long substrates (DP4 and DP5) (Figure 9.16) suggesting that loop 5 should be partially dynamic.

Taken altogether, the starting hypothesis proposed for further evaluation of the subsite capping model are:

1. Loop 5 in the wt enzyme should have reduced dynamic since long substrates such as DP4 and DP5 are poor substrates compared to the short DP2 substrate that already fits into binding cleft in the closed loop 5 conformation observed in the crystal structures of the free enzyme and E·DP2 complex.
2. Mutant *Vc*CDA FL K275E showed and increased activity towards long subtrates. Since an internal salt bridge in loop 5 was disrupted in the mutant, the higher activity suggests that loop 5 flexibility has increased in line with our working hypothesis.
3. The deacetylation pattern of the enzyme does not depend on the DP of the substrate. *Vc*CDA always deacetylates the second position from the non-reducing end of the oligomer. This result indicates that the subsites that a substrate longer than DP2 occupies are the positive ones, so that loop 5 is the most likely candidate to be dynamic and adaptative depending on the DP of the substrate.
4. Loops 1 and 6 and (to a lesser extent 2) are highly interconnected by an extensive network of interactions of different natures. This fact means that the reorganization capacity of negative subsites is very limited. These loops define the deacetylation pattern.

Different experimental strategies were proposed to demonstrate the functional mechanism of the Subsite Capping Model. These two strategies are complementary and opposite in their conception:

a) Loop 5 blocking: If loop 5 had some, but low, dynamic behavour, blocking loop 5 would result in a loss of deacetylase activity on substrates longer than DP3. Locking loop 5 can be achieved by means of an interloop disulfide bond.
b) Increase the loop 5 flexibility: Loop 5, together with loop 4, is dynamic and would be capable of modifying its spatial conformation to create positive subsites in order to accept longer substrates.

On the other hand, protein dynamics have been studied using NMR techniques in order to elucidate the dynamics of the loops in free enzyme and in complex with substrates.

13.2. Loop 5 blocking by a disulfide bond

In comparison with the crystallographic structure obtained for *Vc*CDA FL in complex with DP2 (PDB:4NZ1), the previous computational model has been able to create two additional positive subsites allowing the accommodation of DP4 in the active site.

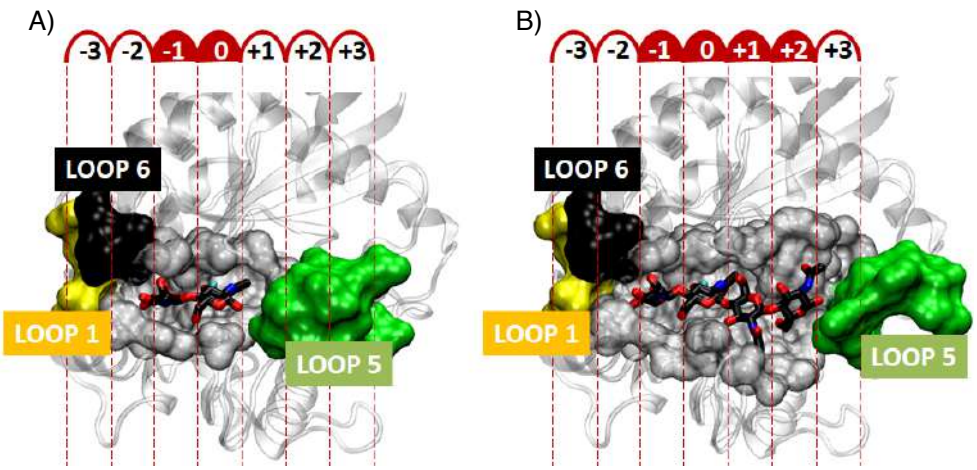


Figure 13.2. Comparison of crystallographic structure (A) and computational model (B). Comparative diagram of the crystallographic structure obtained in complex with DP2 and the computational model in

which loop 5 has been moved and DP4 was docked in the active site. The exposure of two additional positive subsites was observed. Shown as loop 1 and 6 blocking additional negative subsites.

In Figure 13.2, it should be stressed how loops 1 and 6 block the negative subsites. However, all the evidences indicate that the subsites that are generated for the accommodation of substrates longer than DP3 are observed at the reducing end.

To fix loop 5 and limit its dynamics in order to study this hypothesis, it was thought of introducing a disulfide bridge in a suitable position that would fix the loop while not altering the folding of the protein. To this end, the structure of loop 5 and its surroundings were analyzed to identify the optimal point for the creation of this anchor between loops.

The existence of disulfide bridges in the native structure of the enzyme was searched. In the *Vc*CDA structure there are three native disulfide bridges located inside three loops respectively:

a) C106 - C118, Loop 2 (H97 - D131)
b) C185 - C200, Loop 3 (D182 - S206)
c) C264 - C281, Loop 5 (L261 - D288)

On the other hand, within the sequence of the overexpressed VcCDA there are no other cysteines beyond those involved in these disulfide bridges. Figure 13.3. shows the result of the structural analysis to determine the ideal point for the disulfide bond by introducing a double mutation. A single area was identified where the disulfide bridge could be created and this was effectively a determining factor in the movement of the loop. In Figure 13.3. it was observed that the position of this new disulfide bridge is located in the apical zone of loop 5.

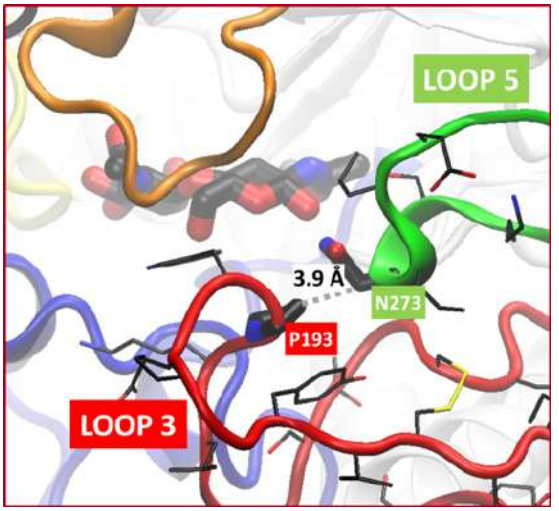


Figure 13.3. Loop 5 blockage location by disulfide bridge. Structural study for the establishment of a disulfide bridge for the loop 5 block. Detail of specific amino acids and distance between them was highlighted.

The two residues to be mutated are P193 and N273, which are at a distance of 3.9 Å between α-carbons (Figure 13.3). This distance is within the range of distances in which native disulfide bridges are found (5.2 Å for the loop 1 bridge, 4.9 Å for loop 3 and 5.4 Å for the loop 5 bridge). The distance of the α-carbon of W273 to C185 is 4.9 Å, and to C200 is 8.7 Å. Once the double mutant was obtained, it should be necessary to experimentally rule out that there is no crosslinking of cysteines in the disulfide bridges that alters the structure and stability of the enzyme. On the other hand, the distances of P193 with the cysteines of the disulfide bridge of loop 3 (option b) are 8.9 Å to C185, and 8.2 Å to C200, so it is unlikely that interference will occur since they are outside the usual distances.

13.2.1. Expression, purification of *Vc*CDA FL P193C / N273C.

Double mutant P193C/N273C was obtained with two consecutive QuikChange® Change PCRs using *Vc*CDA FL wt as a template. Mutant P193C/N273C was expressed and purified following the same procedure described in section 11.1.

In the last purification step, filtration gel chromatography (SEC) was performed (Figure 13.4).

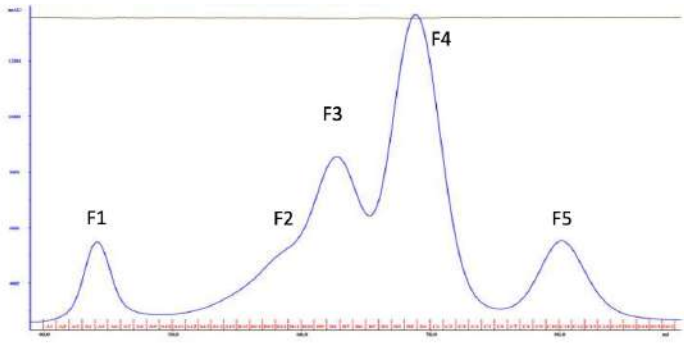


Figure 13.4. GFC chromatogram of VcCDA FL P193C / N273C. Chromatogram obtained on filtration gel chromatography using a Superdex 200 column (16/600). Five multimeric fractions are shown which F5 corresponds to the monomer.

The peak of the monomeric fraction has its maximum close to 80 mL of elution, so using Equation 21.2 a molecular weight of 37.5 kDa is assigned, slightly lower than 46.8 kDa according to the protein sequence. In SDS-PAGE, it was observed that the bands corresponding to VcCDA FL P193C/N273C are slightly above the 45 kDa marker as expected. This differences may reflect a differently hydrodynamic volume of the folded mutant enzyme eluting at different retention time on SEC, but with normal behavior under denaturing SDS-PAGE conditions. Moreover, a high purity of the purified enzyme was observed. The yield of the entire purification protocol results in an average of 3.6 mg/L of culture using the data from two purifications.

13.2.2. Evaluation of intramolecular C193-C273 disulfide bridge.

Cysteines that are not forming a disulfide bond expose a thiol group, whose distinctive reactivity allows its quantification. For the quantification of thiol groups in the enzyme, the Ellman method was used [205]. In this method, Ellman's reagent, 5,5'-dithiobis- (2-nitrobenzoic acid), or DTNB, reacts with free thiol groups to form a mixed disulfide bond between the molecule containing the thiol group and a TNB. TNB released into the aqueous medium ionizes to TNB^{2-} which shows a yellow color that can be followed by an increase in absorbance at 412nm at pH 8.5 (Figure 13.5).

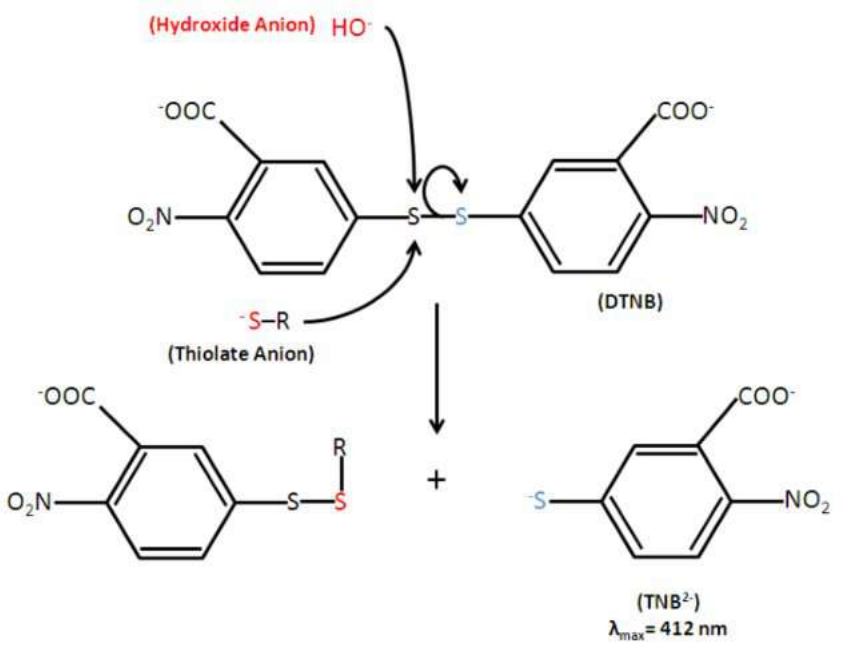


Figure 13.5. Mechanism of Ellman's reaction: DTNB. The disulfide bond reacts in the presence of a thiol releasing 2-nitro-5-thiobenzoate (TNB) which, when ionized in water, acquires a yellow color. Adapted for [205]**.**

To carry out the titration of enzymatic samples, a standard curve was made with L-cysteine hydrochloride as a standard (Figure 13.6).

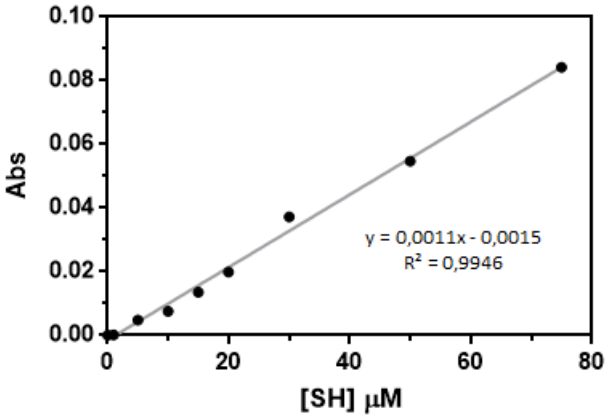


Figure 13.6. Standard curve obtained by the Ellman method for the titration of free cysteines in protein solutions.

Table 13.1. Results of cysteine titration for VcCDA FL and P193C/N273C using Ellman's method.

Protein	[Protein] μM	Mean [SH] μM	[SH]/[protein]
VcCDA FL wt	59.1	6.6 ± 0.9	0.11
VcCDA FL P193C/N273C	14.3	1.7± 0.7	0.12

This methodology was applied to enzyme samples corresponding to *Vc*CDA WT FL and *Vc*CDA FL P193C/N273C, the values shown in Table 13.1. were obtained. It was concluded that there are no free cysteines because the ratio of thiol groups per enzyme is close to 0. These results, together with the fact that the enzyme is active (see section 13.2.3. below), indicate that the new disulfide bond P193C-N273C has been formed (Table 13.1). In addition, the native disulfide bonds of the enzyme have not been disturbed.

13.2.3. Kinetic characterization of *Vc*CDA FL P193C/N273C

This mutant was kinetically characterized using the same substrates (DP2 and DP4) (Figure 13.7. and Table 13.2.).

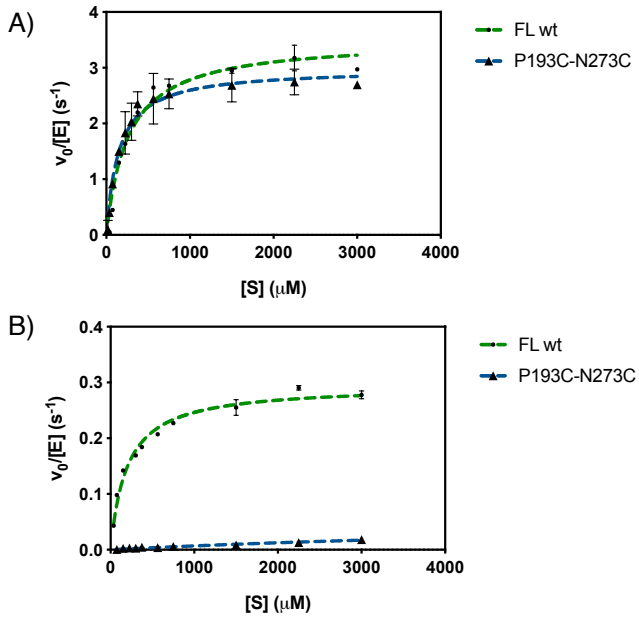


Figure 13.7. Kinetic characterization of *Vc*CDA FL P193C/N273C compared to WT FL. A) Deacetylase activity for DP2 substrate. B) Deacetylase activity for DP4 substrate. All assays were done in 50mM Tris buffer pH 8.5 300mM NaCl and at 37°C.

Table 13.2. Kinetic parameters of the characterization of *Vc*CDA FL P193C/N273C compared to those of *Vc*CDA WT FL parameters.

Substrate	Protein	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M ($M^{-1} \cdot s^{-1}$)
$(GlcNAc)_2$	**WT FL**	3.5 ± 0.12	0.26 ± 0.03	$1.3 \cdot 10^4$
	P193C/N273C	2.97 ± 0.11	0.15 ± 0.02	$2.0 \cdot 10^5$
$(GlcNAc)_4$	**WT FL**	0.29 ± 0.007	0.52 ± 0.05	$5.6 \cdot 10^2$
	P193C/N273C	0.06 ± 0.04	8.7 ± 1.2	0.7

The k_{cat} and K_M values for DP2 vary less than 10% of the value obtained for the WT protein. This is because this substrate does not need the creation of additional subsites that the movement of loop 5 would provide. On the other hand, for long substrates such DP4, the k_{cat} value for P193C-N273C mutant clearly exemplifies that the enzyme has lost the enzymatic activity towards long substrates.

Blocking of loop 5 by creating a disulfide bridge has significantly reduced the activity of the enzyme with substrates that require the access to positive subsites such DP4 (especially +2) while those substrates that do not need the access to a positives subsites are not affected (DP2). To sum up, it should be stressed that loop 5 could be partially dynamic for the accommodation of long substrates. Further experiments should be done to confirm this hypothesis. In addition, to enhance activity on long substrates, loop 5 should be engineering to increase flexibility.

13.3. Flexibilization of Loop 5: Proline to Glycine variants.

13.3.1. Introduction: Loop 5 flexibilization

The Subsite Capping Model, whose experimental demonstration has been initially provided in the previous section, is not only a theory to explain the deacetylation pattern of CE4 family enzymes. The statement also lays the foundation for engineering the substrate specificity of this family of enzymes. This idea has already been presented in previous chapter with the mutant K275E.

Additionally, since loop 5 was shown to be mobile, the flexibility of the different loops in the crystallographic structure of the *Vc*CDA was studied. To analyze this factor, the distribution of structural β factor along the sequence of this deacetylase was investigated. In general, this parameter measures the dispersion of atoms from their average coordinates in a crystallographic structure. The higher the value of the β factor, the greater the dispersion and therefore the more flexible is that area of the protein structure in question. It should be stressed that this interpretation should be done with caution, since the crystallization conditions of a protein do not represent 100% the state of the enzyme in solution.

Figure 13.8 shows the structure of the colored *Vc*CDA as a function of crystallographic β factor (PDB:4NY2). It was observed that the entire conserved nucleus (the part of the sequence that does not include the loops) shows a generalized blue color (low β factor), something to be expected since no significant movements are predicted at the structural level in this area. On the other hand, loop 4 is the one that shows the most intense red color (high β factor). This is because this loop is dynamic and is the one that closes after the binding of the substrate in the active site to generate a perfect catalytic environment. This β factor study indicates that loop 5 is apparently more rigid, hence the efficiency when deacetylating substrate such as DP4 is significantly penalized compared to DP2, which does not require the displacement of this loop.

Previous work in our work was done to determine the role of the flexibilization of loop 5 for the acceptance of longer substrates and which is the impact in enzyme specificity [104]. As examples, complete removal of loop 5 to further create new positive subsites, increase the flexibility of loop 5 by exchanging proline for glycine's residues or reduction of the size of loop 5 by eliminating specific structural motifs were previously evaluated [104] yielding a group of

mutants with different enzyme specificity and increasing the activity towards longer substrates in detrimental of the activity for short substrates (DP2).

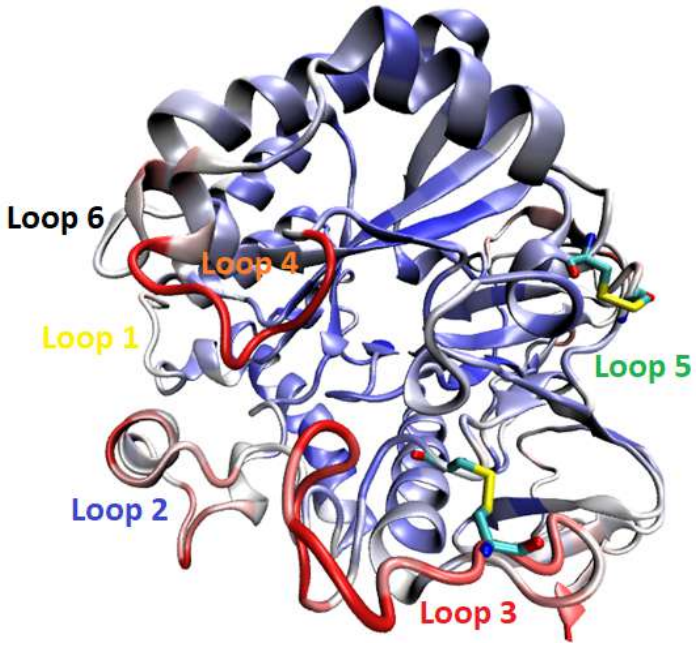


Figure 13.8. β factor in the structure of the *Vc*CDA. Structure of the *Vc*CDA represented based on the different values of β factors throughout its sequence. Gradual scale colouring: blue, minimum values - Red maximum values. Access code PDB: 4NY2.

One of the most impressive strategy to increase the flexibility of loop 5 was the mutation prolines to glycines. The β factor showed us that loop 5 has a higher stiffness than other loops. Close analysis of the structure and sequence of this area of the protein revealed, the presence of three prolines very close to each other in the central part of loop 5 delimited between the cysteines of the internal disulfide bond (Figure 13.8). These prolines are responsible for greatly limiting the rotation capacity of the linkages in which they are involved. This limitation translates into a stiffness that is transmitted to the structure in which they are located. To increase the mobility of the loop, the mutation of these three residues to glycine (P266G, P271G and P280G) was proposed and studied (Figure 13.9) [104]. Mutant was called VcCDA CD PG. Unlike prolines, glycines do not have any type of side chain. These mutations also alter the backbone parameters and topology. Prolines to glycines mutant will allow the loop to be flexible, which may mean that the movement of loop 5 is more efficient and the accommodation of long substrates in further positives subsites is favored, or that by increasing flexibility, loop 5 changes its configuration or/and conformation and adopts a

permanently open conformation in which subsites towards the reducing end are always available.

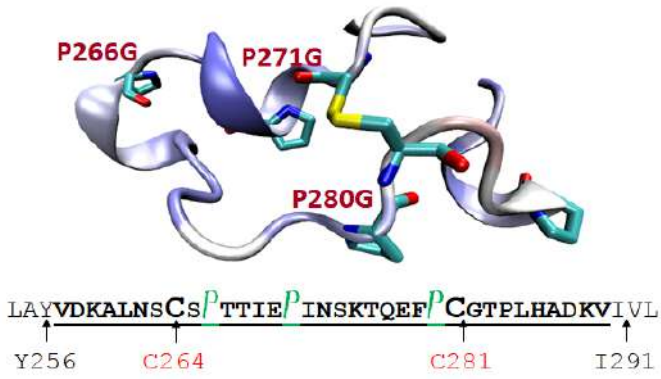


Figure 13.9. Increased flexibility by prolines to glycine's mutant. Increasing the flexibility of loop 5. The three prolines are mutated by glycines, a type of amino acid that provides more flexibility to the backbone of a protein.

In previous work [104], P-G mutant was prepared using QuikChange® protocol. Protein expression and purification was done according the section 11.1. Gel filtration profile obtained in previous work is shown in Figure 13.10 (see protocol 21.4.6.1).

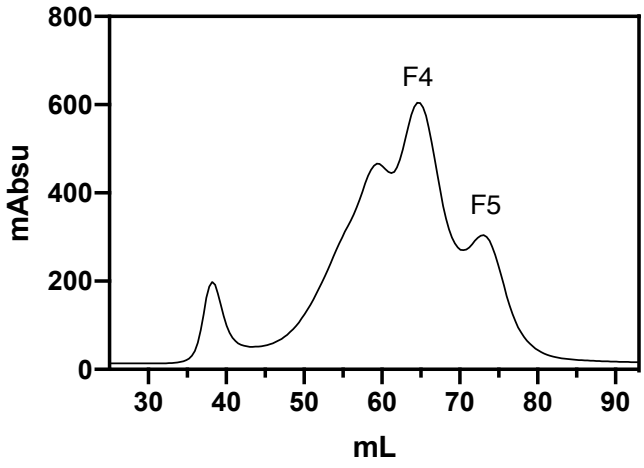


Figure 13.10. Size exclusion chromatography of VcCDA CD PG. Superdex 200. Different fractions were assigned. F4: Dimeric fraction, F5: Monomeric fraction.

Five fractions were observed similar to *Vc*CDA CD but with a very low presence of the monomeric fraction F5 (Figure 13.10). If the molecular masses are calculated for fraction F4, a value of 62.6 kDa is obtained while for fraction F5 it is 31.2 kDa.

The kinetic characterization of this mutant (monomeric fraction) was carried out using the substrates DP2 and DP4 by HPLC-MS (Table 13.3).

Table 13.3. Kinetic parameters for *Vc*CDA CD and PG. All assays were done in 50 mM Tris buffer pH 8.5, 300 mM NaCl and at 37°C.

Substrate	Protein	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M ($M^{-1}\cdot s^{-1}$)	% k_{cat} (Mut/CD)
$(GlcNAc)_2$	***Vc*CDA CD wt**	2.5 ± 0.2	1.3 ± 0.3	$1.9 \cdot 10^3$	
	***Vc*CDA CD PG**	0.10 ± 0.02	6.1 ± 1.9	$1.6 \cdot 10^1$	4
$(GlcNAc)_4$	***Vc*CDA CD wt**	0.5 ± 0.03	0.9 ± 0.1	$5.6 \cdot 10^2$	
	***Vc*CDA CD PG**	2.9 ± 0.4	4.9 ± 0.7	$5.9 \cdot 10^2$	580

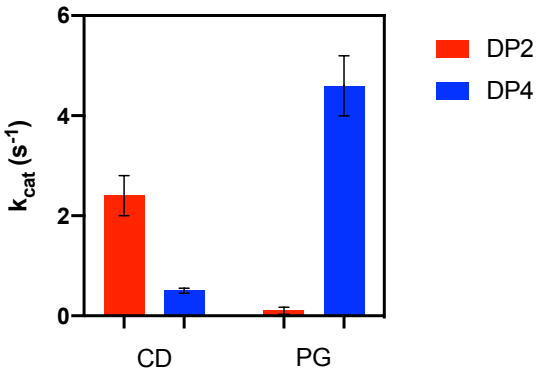


Figure 13.11. Evolution of activities (k cat) depending on the DP of the substrate. All assays were done in 50mM Tris buffer pH8.5 300mM NaCl and at 37°C.

For PG mutant, of which the monomeric fraction was characterized, the catalytic behaviour for deacetylation of DP2 was lost but catalytic efficiency increases with long substrates. For long substrates, the k_{cat} value has increased 580%, a very significant increase at expenses of K_M, which has been drastically reduced meaning that binding of the substrate has been strongly affected. It was proposed that these longer substrates can be accommodated on an additional number of subsites through the greater flexibility of loop 5 maintaining a linear arrangement over the groove of the active center and allowing loop 4 to close completely and develop a greater catalytic activity. In next sections, an in-depth analysis of loop protein dynamics will be done using NMR techniques (see section 13.4). The PG mutant maintain the

deacetylation pattern of the original enzyme regardless of the substrate used (from DP2 to DP5) [104].

13.3.2. P-G mutant deconvolution: Evaluation of individual and double mutations.

Glycine and proline residues have a major influence on the kinetics of loop formation in proteins. Glycine accelerates loop formation by decreasing the activation energy, whereas trans prolyl bonds slow loop formation by increasing the barrier height [206]. A single prolyl or glycyl residue influences chain dynamics only locally. Mutation of proline to glycine unequivocally plays a key role in loop formation and final conformation. Here, mutation of the three identified prolines to glycine might be a drastic modification in loop 5 topology that induces a conformational change reducing the affinity towards natural substrates (chitobiose) as it has been previously observed in K_M values (Table 13.3). On the other hand, mutant *Vc*CDA K275E obtained and characterized in section **¡Error! No se encuentra el origen de la referencia.** suggests that, instead of a drastic flexibilization as with proline to glycine mutations, slight flexibilization may allow the partial opening of the loop to accommodate DP4 and DP5 substrates with the concomitant increase of activity without completely disturbing the loop conformation affecting the proper binding of the substrates (K_M) (Table 12.1).

The impressive results obtained in kinetic characterization of *Vc*CDA PG mutant raising different questions to elucidate the role of individual prolines in the wt enzyme:

- Is the effect of proline mutations additive?
- Each mutation contributes equally to the final triple mutant? Do all Pro → Gly mutations contribute to the increased activity on long substrates of the triple mutant?
- Is the low activity observed for DP2 on the triple mutant due to a particular Pro mutation?
- Is there a correlation between Pro position and loop dynamics? Could single mutations enhance enough the loop dynamics?

We next sought to clarify the molecular insights previously proposed by the deconvolution of *Vc*CDA PG mutant to perform the biochemical characterization of the individual mutants. The molecular effect of each proline mutation (P266, P271 and P280G) will be analyzed, individual

mutants were prepared and biochemically characterized to obtain the M&M kinetic parameters.

In previous work [104], the activity profile of the different oligomeric fractions of the PG mutant were analyzed with the AcOMU substrate. It should be stressed that the dimeric state was the most active fraction and almost no activity for the monomeric fraction. As a general rule, the monomer had always been the fraction with the highest activity in all other variants. Since AcOMU is an artificial substrate (although unspecific), the specific activities of the dimeric and monomeric fraction of PG mutant were evaluated using DP2 and DP4 prior to the expression and characterization of individual mutants. Here, detailed activity profile of dimeric and monomeric fraction have been studied (see Figure 13.12 and Table 13.4).

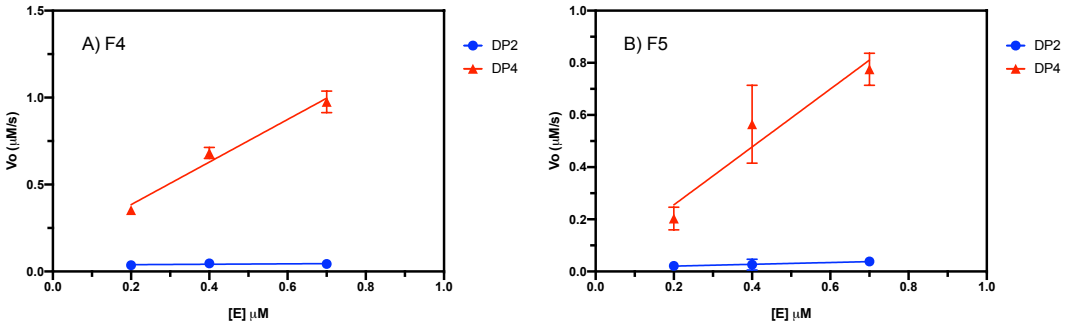


Figure 13.12. Comparison of the monomeric and dimeric fraction of the PG mutant. Graphs showing the comparison of the specific rates for the monomeric (B) (F5) and dimeric fraction (A) of the PG for DP2 and DP4. All assays were done in 50mM Tris buffer pH 8.5 300 mM NaCl and at 37°C.

Table 13.4. Specific activities of dimeric (F4) and monomeric (F5) form using DP2 and DP4 substrates (data obtained from Figure 13.12.).

Protein	Substrate	Specific activity (s^{-1})
VcCDA PG (F4)	**$(GlcNAc)_2$**	0.01 ± 0.002
	$(GlcNAc)_4$	1.22 ± 0.12
VcCDA PG (F5)	**$(GlcNAc)_2$**	0.03 ± 0.02
	$(GlcNAc)_4$	1.11 ± 0.22

Figure 13.12 shows that the monomeric fraction F5 is actually highly active. Moreover, if we pay attention to the specific activity of each (Table 13.4), it can be concluded that the monomeric fraction have similar activity for both substrates DP2 and DP4 compared with the

dimeric fraction (F4). Taken altogether and considering that further NMR studies need to be performed using the monomeric fraction, all studies were done with monomeric fraction.

13.3.2.1. Expression and purification of individual mutants

Individual mutants were obtained by QuikChange® PCR (see protocol 21.3.1). Mutants expression and purification was done as reported in section 11.1. Strep-tag and Gel filtration profiles for each individual mutant are shown in Figure 13.13:

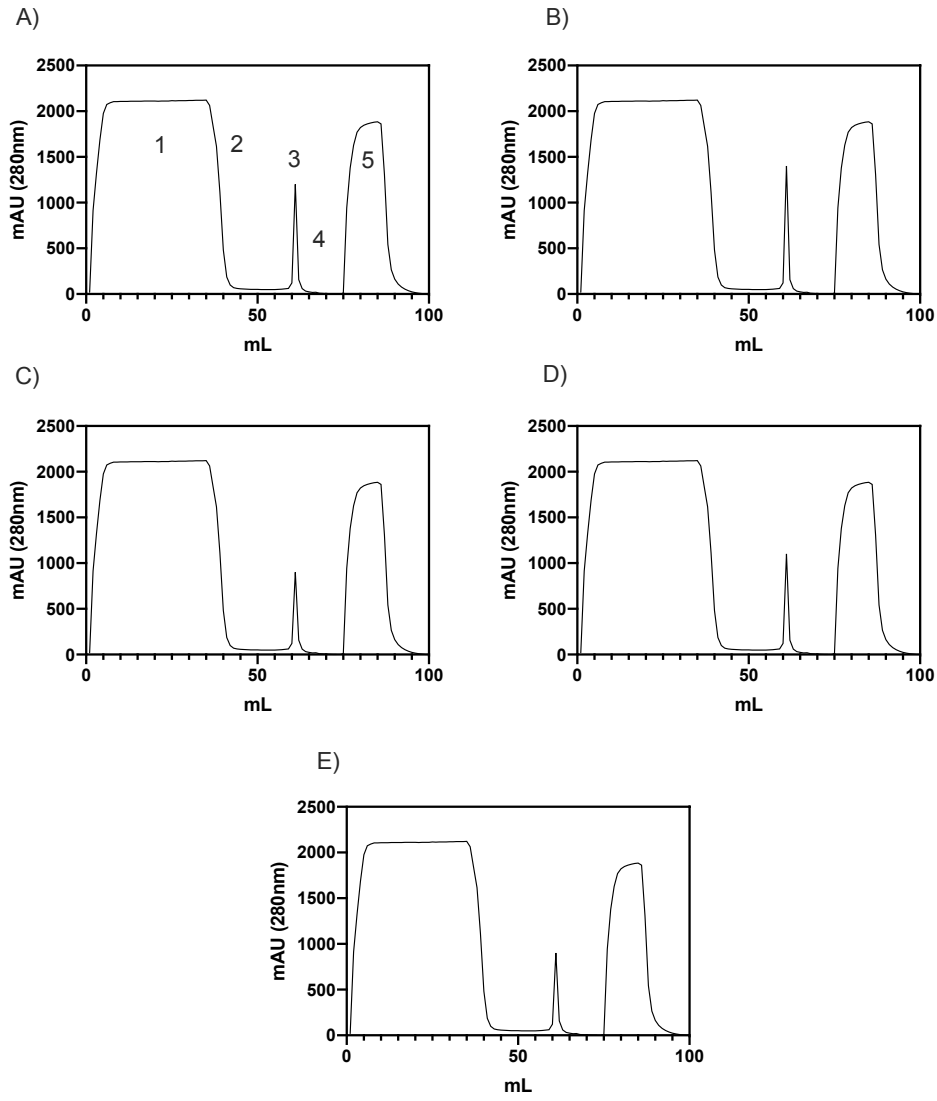


Figure 13.13. Strep-Trap purification chromatogram of individual mutants. A) *Vc*CDA CD (catalytic domain) wt B) *Vc*CDA CD PG (contains three prolines mutated to glycine; P266G, P271G, P280G. C) *Vc*CDA CD P266G D) *Vc*CDA CD P271G E) *Vc*CDA CD P280G. 3 and 4 column loads were carried out in which three

column loads were carried out and the usual protocol of loads washes and elutions was followed. The numbers indicate the stages of a charge/elution cycle: 1. Charge, 2. Wash, 3. Elution, 4. Wash, 5. Regeneration

Gel filtration elution profiles of the three individual mutants together with *Vc*CDA CD WT and *Vc*CDA CD PG are shown in Figure 13.14. Considerably different results are observed in terms of the population distribution of the oligomeric fractions.

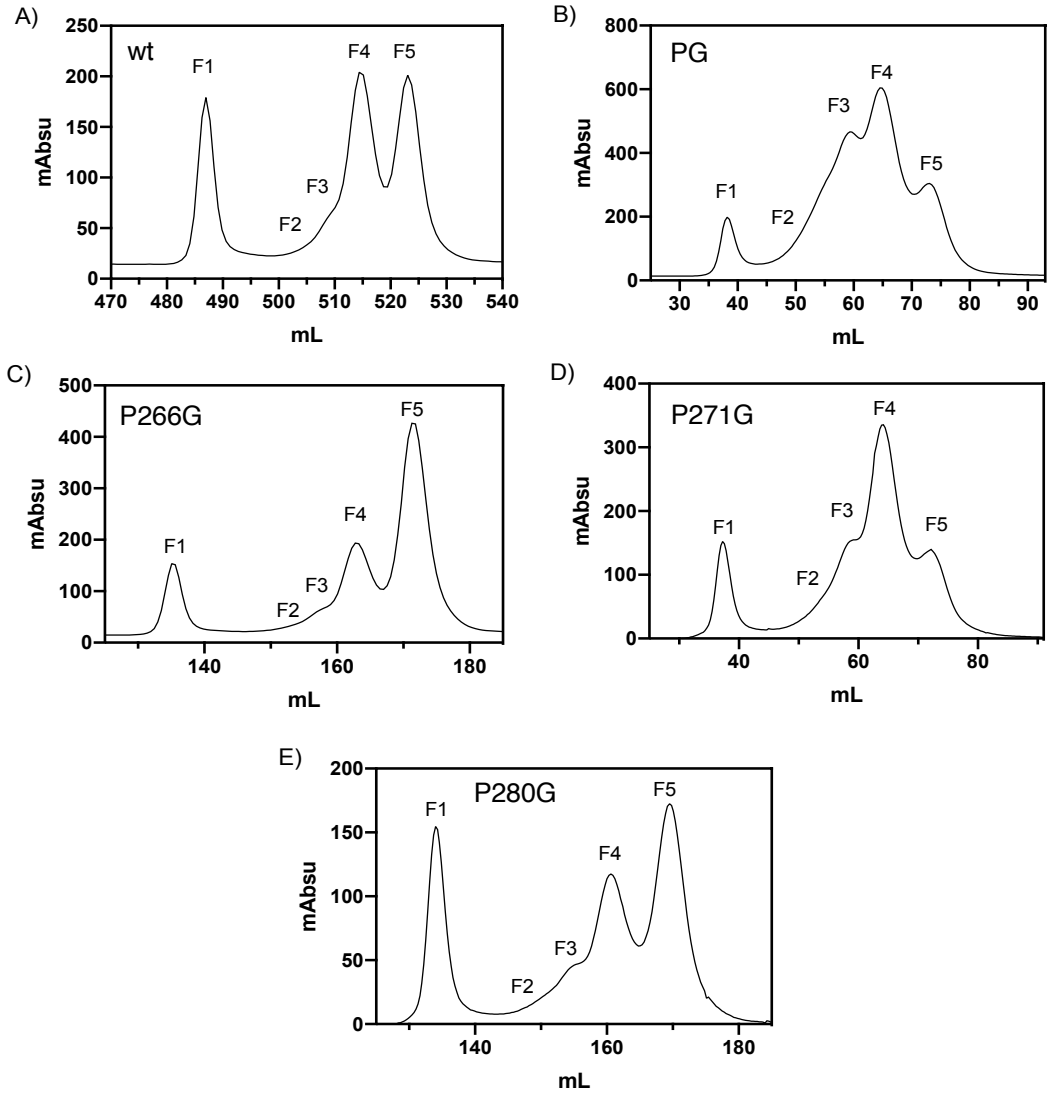


Figure 13.14. Chromatogram obtained in gel chromatography filtration using SuperDex 200 column (16/600). A) *Vc*CDA CD (catalytic domain) wt B) *Vc*CDA CD PG (contains three prolines mutated to glycine;

P266G, P271G, P280G. C) *Vc*CDA CD P266G D) *Vc*CDA CD P271G E) *Vc*CDA CD P280G. All purificationswere carried out in 50 mM Tris buffer pH 8.5, 300 mM NaCl and at a temperature of 37ºC.

The elution profile of the proteins can be summarized in that proline to glycine mutant entails a displacement of the population profile of oligomers towards the dimeric or monomeric fraction compared with CD wt. For the PG mutant (Figure 13.14.B) a profile of 5 fractions is observed similar to that of the VcCDA CD but with a very low presence of the monomeric fraction F5. For individual mutants, P266G and P280G how a similar gel filtration profiles as CD wt enzyme, on the other hand, the mutant P271G shows a profile closely related to the PG enzyme.

Samples from all purifications, from lysis and clarification to the different fractions of the gel filtration elution were analyzed by SDS-page (Figure 13.15. for P271G as an example). Same profile of SDS-PAGE was obtained for all mutants.

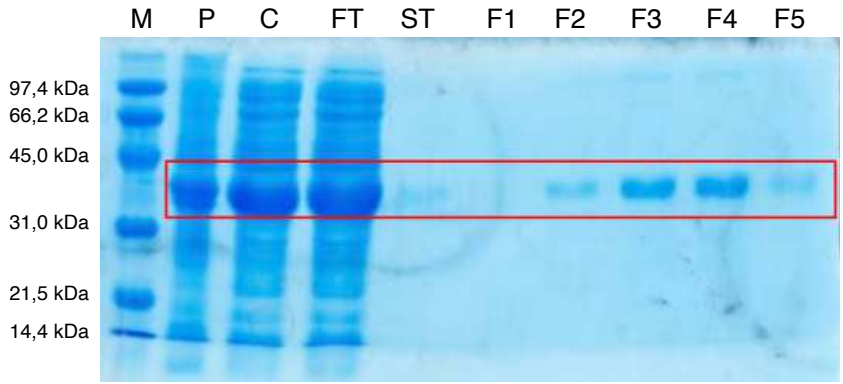


Figure 13.15. SDS-PAGE of the *Vc*CDA CD P271G individual mutant as example. M: Marker Low Range, P: Pellet from the centrifugation after the lysate, C: crude, soluble fraction of the lysate, FT: Flow through the column, ST: Protein purified by the Strep-Trap, F1-F5: fractions of the 1 to 5.

SDS-PAGE gel shows high purity and the fractions of different elution peaks in the SEC as single bands confirming the no-covalent nature of the oligomers (fractions F1 to F4) that are denatured under SDS-PAGE conditions. The experimental molecular mass (around 35 kDa) corresponds to the expected theoretical values. Protein concentrations were quantified by the BCA method using a BSA standard, and the yields per liter of culture of each of the proteins were calculated (Table 13.5).

Table 13.5. Purification yield of *Vc*CDA CD wt, PG triple mutants and individual mutants.

Mutant	Yield (mg protein/L culture)
***Vc*CDA CD wt**	**3.5-7**

***Vc*CDA CD PG**	**4.5**
***Vc*CDA CD P266G**	**3.1**
***Vc*CDA CD P271G**	**3.8**
***Vc*CDA CD P280G**	**4.1**

All expressed and purified mutants show a similar yield of expression and purification.

13.3.2.2. Kinetic characterization of individual mutants and double mutant

Enzyme kinetics were analysed using the reference substrates DP2 and DP4. On the other hand, specificity in terms of degree of deacetylation of the substrate was studied. The protein used as a reference was the isolated catalytic domain (*Vc*CDA CD wt). In the set of graphs in Figure 13.16, activity profiles were obtained by representing v_0/[E] against substrate concentration to calculate the kinetic parameters by fitting the data to the Michaelis-Menten equation (see protocol 21.5.4). Visual inspection already suggests that two individual mutants together with PG triple mutant show higher activity for long substrates than for short substrates.

Table 13.6 summarizes k_{cat} and K_M values and other derived parameters obtained for each of the three individual mutants in comparison with the isolated catalytic domain and PG triple mutant. As with the WT CD enzyme, no cooperative behaviour was observed for any substrate.

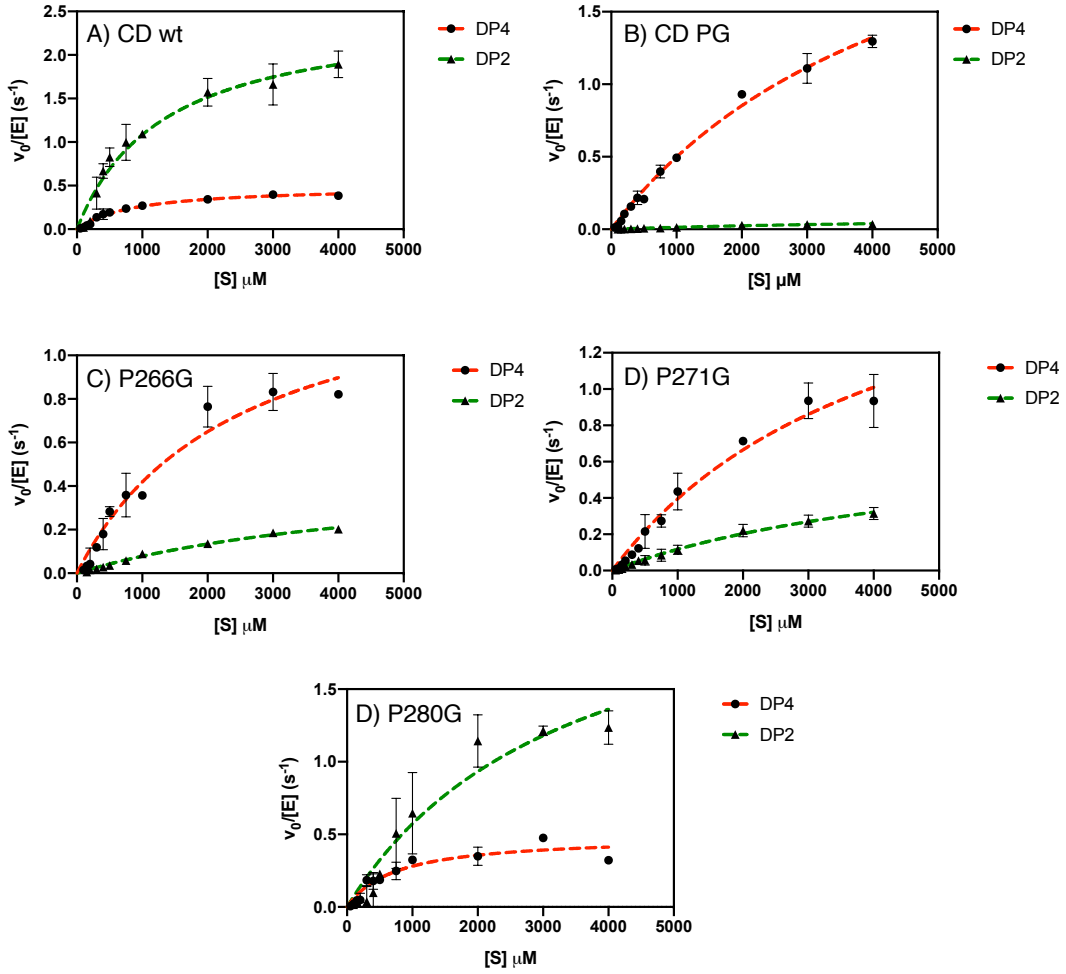


Figure 13.16. Kinetic characterization of CD WT, PG and individual mutants. A) *Vc*CDA CD wt, B) *Vc*CDA CD PG (triple mutant), C) *Vc*CDA CD P266G, D) *Vc*CDA CD P271G, E) *Vc*CDA CD P280G. DP4 substrate (red) and DP2 substrate (green). All the assays were carried out in 50 mM Tris buffer pH 8.5, 300 mM NaCl and at a temperature of 37°C

Table 13.6. Kinetic parameters of CD WT, PG and individual mutants. As cooperative behavior was not observed, the Michaelis Menten equation was used in all cases. All the assays were carried out in 50 mM Tris buffer pH 8.5, 300 mM NaCl and at a temperature of 37ºC.

Substrate	Mutant	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M ($M^{-1}\cdot s^{-1}$)	% k_{cat} (Mut/CD)
$(GlcNAc)_2$	CD	2.5 ± 0.2	1.3 ± 0.3	$1.9 \cdot 10^3$	
	PG	0.1 ± 0.02	6.1 ± 1.9	$1.6 \cdot 10^1$	4
	P266G	0.5 ± 0.1	5.2 ± 1.1	$9.2 \cdot 10^1$	19
	P271G	0.8 ± 0.1	5.6 ± 1.5	$1.4 \cdot 10^2$	32
	P280G	2.5 ± 0.1	3.4 ± 1.3	$7.4 \cdot 10^2$	100
$(GlcNAc)_4$	CD	0.5 ± 0.03	0.9 ± 0.1	$5.6 \cdot 10^2$	
	PG	2.9 ± 0.4	4.9 ± 0.7	$5.9 \cdot 10^2$	580
	P266G	1.5 ± 0.2	2.5 ± 0.5	$5.8 \cdot 10^2$	290
	P271G	2.1 ± 0.4	4.3 ± 1.2	$4.9 \cdot 10^2$	420
	P280G	0.5 ± 0.1	0.7 ± 0.2	$7.1 \cdot 10^2$	100

Broadly, the PG mutant highly increases the activity towards longer substrates such as DP4 although the activity for natural substrate chitobiose has been reduced (Table 13.6). It mutant shows an activity increase (k_{cat}) towards DP4 of almost 600% with regard to CD wt enzyme. The K_M value is also affected, being significantly higher than that for the isolated catalytic domain showing lower affinity for the substrates (Table 13.6). In addition, in terms of catalytic efficiency (k_{cat}/K_M) for DP4, PG mutant and CD show a similar values due to the increase of the turnover number (k_{cat}) and Michaelis constant K_M simultaneously. On the other hand, the activity of PG mutant towards DP2 (k_{cat}) has been decreased to 4% compared with CD wt, likewise catalytic efficiency (k_{cat}/K_M) for DP2 is notably decreased in PG mutant. This loss of affinity (K_M) towards DP2 and DP4 substrates could be related to an increase of loop 5 flexibility. The higher flexibility could cause the substrate to have a lower capacity to stick and interact with the active site, thus loosing affinity. A deeper study using NMR tools will be performed in further sections to understand and define the real molecular insights (see chapter 13.4).

Regarding the individual mutants it should be stressed that P266G and P271G mutants show higher activity in terms of k_{cat} for long substrate (DP4) but lower activity towards DP2 substrate compared with CD wt enzyme as it has been obtained for PG triple mutant (Table 13.6). In short, P266G and P271G mutants show an increase of 290% and 420% in k_{cat} for DP4. For K_M values, P271G exhibit an affinity constant for DP4 similar to the PG mutant and it is almost two-fold higher than the P266G value. P271G and P266G enzymes lost the affinity (K_M) for DP2 substrate. The individual mutation P271G has the highest contribution to the increased activity towards longer substrates shown in PG triple mutant.

For P280G mutant, it must be underlined that it shows a similar behaviour than CD wt enzyme maintaining the higher activity for short substrates and lower activity for long substrates (DP4). Based on the results presented in Table 13.6, it can be concluded that P280G mutant has almost no contribution to the flexibilization of loop 5 by introduction of proline to glycine mutations.

Taken altogether, the results strongly suggest that double mutant P266G-P271G may have the same effect on increased activity for chitotetraose than PG triple mutant and that P280G mutant show the same activity on DP2 and DP4 than CD wt enzyme thus discarding its involvement in the flexibilization of loop 5.

The double mutant P266G-P271G was prepared by QuikChange® PCR (see protocol 21.3.1). The double mutant was expressed using autoinduction media and purified using Strep-tag affinity chromatography and subsequent gel filtration Superdex 200 column (see protocol 0).

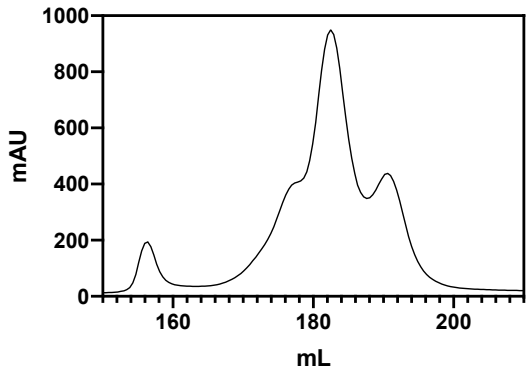


Figure 13.17. Chromatogram obtained in gel chromatography filtration using SuperDex 200 column (16/600) for *Vc*CDA CD P266G-P271G.

The same gel filtration profile (SEC) has been obtained for the double mutant P266G-P271G (Figure 13.17). The yield of the purification was 3.8 mg of protein per liter of culture. Biochemical characterization is shown in Figure 13.18 and Table 13.7 in order to determine the kinetic parameters using DP2 and DP4 as a substrates.

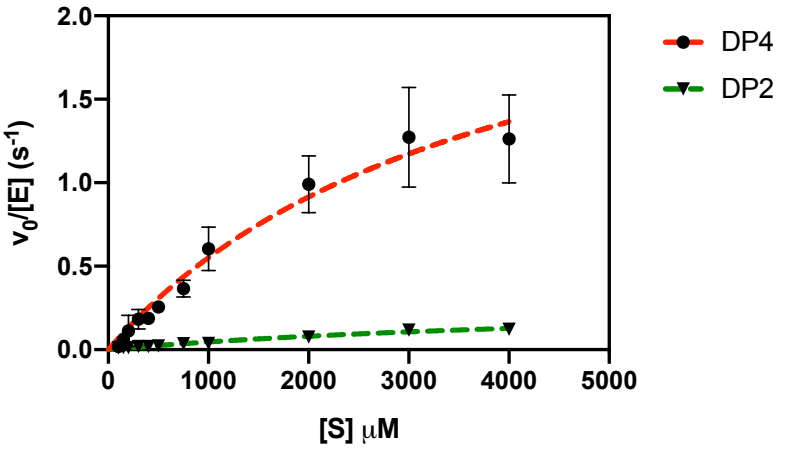


Figure 13.18. Kinetic characterization of VcCDA CD P266G+P271G. DP4 substrate (red) and DP2 substrate (green). All the assays were carried out in 50 mM Tris buffer pH 8.5, 300 mM NaCl and at a temperature of 37°C.

Table 13.7. Kinetic parameters of VcCDA CD P266G+P271G. k_{cat} and K_M. All the assays were carried out in 50 mM Tris buffer pH 8.5, 300 mM NaCl and at a temperature of 37°C.

Substrate	k_{cat} (s^{-1})	K_M (mM)	K_{cat}/K_M ($M^{-1}\cdot s^{-1}$)	% k_{cat} (Mut/CD)
$(GlcNAc)_2$	0.3 ± 0.04	3.9 ± 1.3	$7.6 \cdot 10^1$	12
$(GlcNAc)_4$	2.7 ± 0.5	5.7 ± 1.1	$4.7 \cdot 10^2$	540

The double mutant shows an activity profile highly resembling to that of the PG mutant. As a summary, in the following figures, the evolution of deacetylase activities in terms of k_{cat} and substrates affinities (K_M) depending on the DP of the substrate are displayed (Figure 13.19 and Figure 13.20).

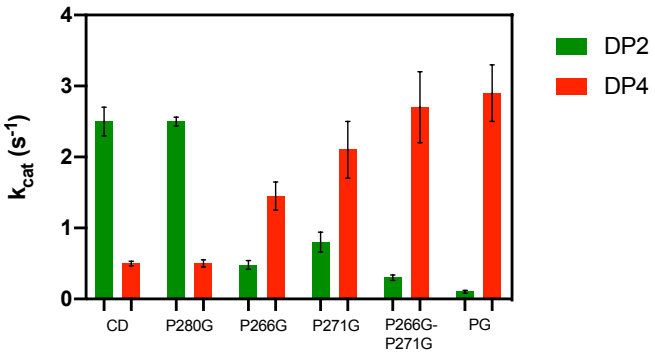


Figure 13.19. Evolution of activities in terms of k_{cat} depending on the DP of the substrate for CD wt , PG triple mutant and individual mutants (P280G, P266G and P280G) and double mutant P266G-P271G.

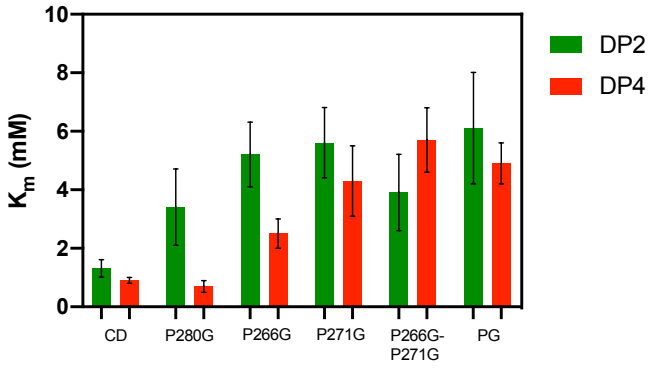


Figure 13.20. Evolution of substrate affinities in terms of K_M depending on the DP of the substrate for CD wt, PG triple mutant and individual mutants (P280G, P266G and P280G) and double mutant P266G-P271G.

Taken altogether, an increment of k_{cat} values for long substrates in both individual mutants (P266G and P271G) and double mutant P266G-P271G clearly is at expenses of lower activity for short substrates (Figure 13.19). In terms of k_{cat}, most of contributions to increase the activity in the triple mutant PG comes from the double mutation P266G-P271G, and mutant P280G does not contribute since it behave almost the same as the enzyme wt.

On the other hand, all mutants have a negative impact in terms of substrate affinity (Figure 13.20). K_M values for DP2 suggest a loss of affinity for short substrates in all mutants and K_M values for DP4 are also higher in double mutant P266G-P271G and mutant P271G. For mutant P266G the impact in K_M is not as marked.

Kinetic results suggest that proline mutation effects is additive for P266G and P271G mutants. P271G mutant shows the highest activity towards long substrates but that contribution of both mutations is not equal. P271G mutant is located in the apical zone of loop 5 and P266G is oriented to the active site (Figure 13.21). On the other hand, P280G mutant is located behind the loop and slightly far of the active site. After herein, mutants located in the part of the loop 5 facing the active site seem to induce a change in dynamics and conformation to further create an enabling environment for catalysis of long substrates.

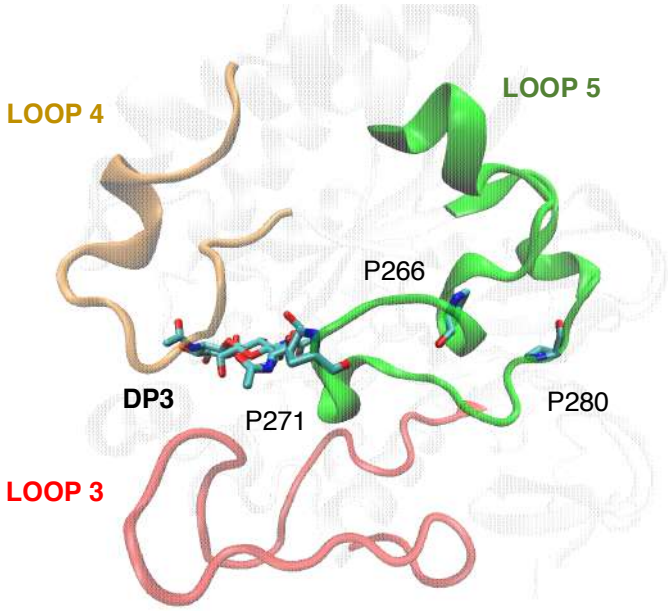


Figure 13.21. Increased flexibility by prolines to glycine's mutant. P266, P271 and P280 positions were highlighted. PDB: 3OUI: *Vc*CDA wt in complex with DP3.

This restructuring of loop 5 also implies a noteworthy impact in substrate affinity. Kinetic results suggested that dynamics or conformation of loop 5 changes in PG mutants result in an increase of activity towards long substrates at the expenses of a loss of affinity.

To elucidate the dynamics and conformational changes of loop 5 in the PG mutants, NMR experiments will be presented in next section 13.4.

13.4. Protein dynamics by NMR.

13.4.1. Introduction: NMR techniques for protein dynamics.

Solution nuclear magnetic resonance (NMR) spectroscopy is unique in its ability to elucidate the details of atomic-level structural and dynamic properties of biological macromolecules under native-like conditions. The information obtained from NMR dynamics experiments provides insights into specific structural changes or configurational energetics associated with function. A variety of applications illustrate that studies of protein dynamics provide insights into protein–protein interactions, target recognition, ligand binding, and enzyme function.

Isotopic labelled samples are generally required for optimizing NMR solution conditions [207], NMR assignments and structure determination [208], as well as performing heteronuclear spin relaxation studies [209], protein dynamics studies [210], biological screening applications such as 'SAR by NMR' [211], and high throughput structural genomics studies [212].

13.4.2. Protein expression and purification in minimal media for isotopic labelling.

Most of the biomolecular NMR techniques for protein studies require isotopic labelling with different probes (^{2}H, ^{13}C, and/or ^{15}N). Currently, most isotopically labelled recombinant proteins are expressed in the bacterial host *E. coli* and large protein quantities are needed for most of these techniques. Methods of generating heteronuclear labelled samples in *E. coli* commonly use standard or modified versions of M9 minimal media [213] employing ^{13}C glucose for carbon labelling, ^{15}N ammonium sulphate or ^{15}N ammonium chloride for nitrogen labelling, and deuterium oxide for deuteration. Other labelling strategies such as methyl-labelling are commonly used in TROSY experiments in which isotope-labelled amino acids or amino acid precursors are added to the growth medium of the cells. To ensure that these labelled compounds only end up in the target sites and are not ''scrambled" into other amino acids, it may be necessary to add additional metabolites or to genetically modify the expression host [214,215]. For instance, in Methyl-TROSY experiments (see section 13.4.3.4), methionine labelling can be conveniently achieved by the direct addition of methyl-^{13}C-methionine to the growth medium, which efficiently represses endogenous methionine synthesis.

A variety of different strategies for enhancing *E. coli* growth and protein expression have been employed. Minimal media supplements such as trace ion metal mixtures, vitamin cocktails as well as commercially available algal and microbial hydrolysates have shown enhancements in growth and expression [213,216].

An initial barrier to NMR studies can be the sub-optimal quantities of soluble protein produced from heterologous expression in *E. coli* bacteria. This problem is more acute when considering the loss in yield and the increase in cost associated with producing the highly deuterated samples that are required for TROSY-based NMR studies of larger (>30 kDa) macromolecules [217]. Moreover, the heterogeneous expression of recombinant proteins in *E. coli* is often hampered by protein aggregation into inclusion bodies. Protein aggregation is governed by physicochemical and structural features of the proteins themselves (MW, hydrophobic residues, low complexity) [218]. Recently, several reports have appeared describing new methods for increasing protein yield using high cell densities, fermentation techniques, or amino acid algae lysate supplements [219,220].

Different steps need to be undertaken on a protein expression system using minimal media, such as the concentration of IPTG for transcription induction of heterologous expression and/or the temperature and duration for protein expression, to maximize protein yield. Other optimizations may include the strain used for protein expression (e.g., CodonPlus™ or Origami™), changing the plasmid backbone, or codon optimization of the gene of interest.

Here, we described a general and simple protocol using the standard M9 minimal medium recipe that gives enough yield of heterologous expressed proteins in *Escherichia coli* BL21(DE3) bacteria for protein labelling. Protein expression and purification protocol were optimized using *Vc*CDA CD D39S and *Vc*CDA CD PG D39S. Inactive mutants (D39S) of both enzymes were selected for NMR studies to prevent substrate turnover and focus on substrate binding.

In our final protocol system, *E. coli* BL21 (DE3) pLyss has been selected for protein expression [221] to control protein labelling. This strain contains the pLysS plasmid that expresses the T7 lysozyme gene to reduce the background of the target protein expression without disturbing IPTG induction. Two labelling strategies were selected in our final protocol for further NMR studies: nitrogen labelling using $^{15}NH_4Cl$ and Methionine and Isoleucine labelling using L-METHIONINE (METHYL-13C, 99%) and ALPHA-KETOBUTYRIC ACID, SODIUM SALT(METHYL-13C, 99%; 3,3-D2, 98%) labels, respectably.

Following the classical protocol for protein expression several variations in the expression conditions methods were tested in order to improve soluble protein expression. Table 21.9 lists the composition of final media used for protein expression [222].

After cell growth to late exponential phase at 37ºC, the temperature was lowered to 24ºC in protein expression induced with IPTG. Samples were withdrawn at different time intervals (0h, 3h, 6h and 24h after induction). After lysis, soluble and insoluble fractions were analyzed by SDS-PAGE.

Two different carbon sources supplementing the M9 medias were tested: glycerol and glucose (Figure 13.22 and Figure 13.23)

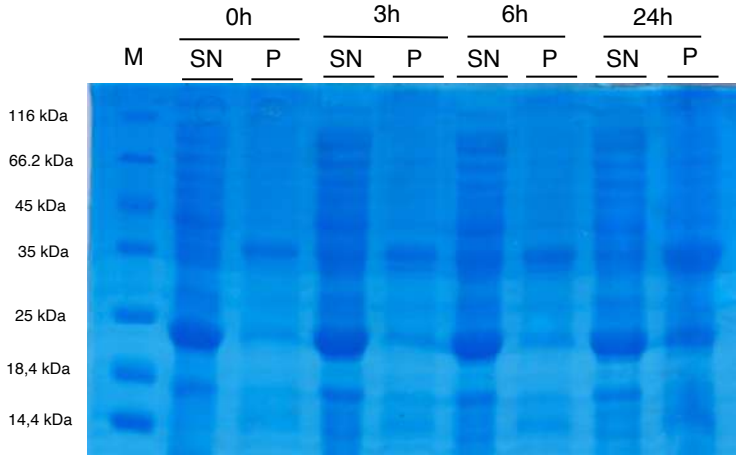


Figure 13.22. VcCDA CD D39S protein expression using glycerol as carbon source (40 g/L) following the protocol . Expected MW: 35 kDa.

As shown in Figure 13.22, not enough protein (35 kDa band) was observed at different time intervals (from 0h to 24 h). This might be because the slow rate of cell growth using glycerol. An unknown band around 25 kDa appears and it cannot be assigned to any specific protein. Protein expression using glycerol as carbon source was discarded.

The same protocol was done using glucose as a carbon source. Different samples of the culture were withdrawn at several time intervals (0h, 3h, 6h and 24h after induction) and soluble and insoluble fractions were analysed by SDS-PAGE for *Vc*CDA CD D39S and *Vc*CDA CD PG D39S (Figure 13.23).

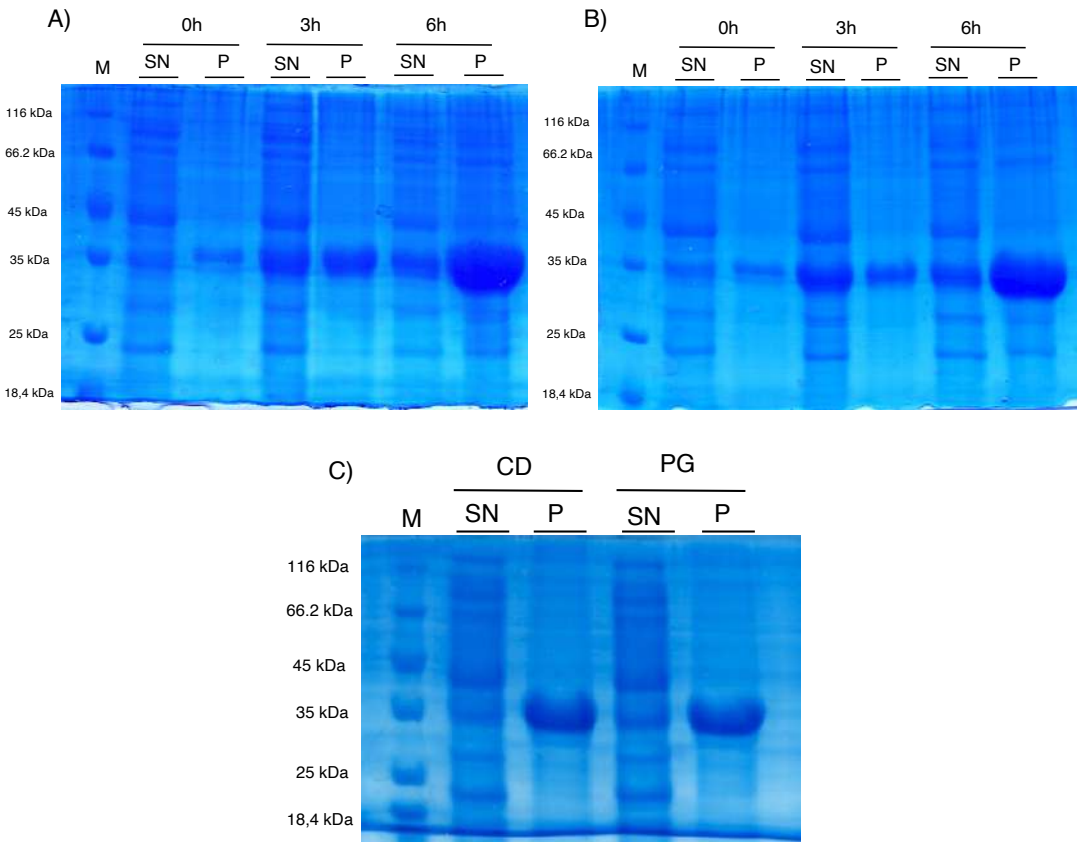


Figure 13.23. SDS-PAGE 14% acrylamide of expression tests of *Vc*CDA. Molecular weight of *Vc*CDA CD wt and PG is 35 kDa. M: Marker, SN: Soluble fraction, P: Pellet. A) Protein expression of *Vc*CDA CD D39S at different times. B) Protein expression of *Vc*CDA PG D39S at different times. C) Protein expression after 24h at 24ºC and 250 rpms. CD: *Vc*CDA CD D39S, PG: *Vc*CDA PG D39S.

Maximum soluble protein after lysis was achieved after 3h of protein expression at 24ºC and 180 rpm's (Figure 13.23.A and B). After 6h of protein expression, most of the protein has been observed in the insoluble fraction for both proteins. After 24h, no protein was observed in the soluble fraction. It was concluded that 3h of protein expression at 24ºC and 180 rpms are the exact time for soluble protein expression.

Additionally, the reduction of protein expression temperature is one of the common approaches to prevent formation of inclusion bodies by correct folding of the molecule. Unfortunately, this comes along with prolonged cultivation times. In order to maximize protein expression in the soluble fraction, a decrease in temperature to 20ºC for IPTG induction was tested (Figure 13.24).

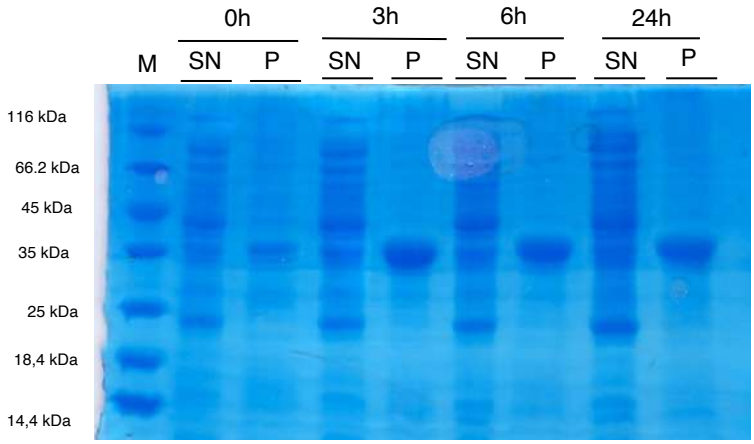


Figure 13.24. Study of protein expression at different intervals of time at 20ºC using *Vc*CDA CD D39S enzyme. M: Marker, SN: Soluble fraction, P: Pellet.

However, most of the expressed protein went to the insoluble fraction in all time intervals at 20ºC. Reduction of temperature for protein expression was then discarded.

As a summary, different expression conditions were tested, and several methods designed to enhance the solubility of recombinant proteins in bacterial hosts were used. The final protocol for protein expression in minimal media is detailed in section 21.4.3. Protein expression at 24ºC and 180 rpms for 3h has been selected as final optimal conditions (Figure 13.25).

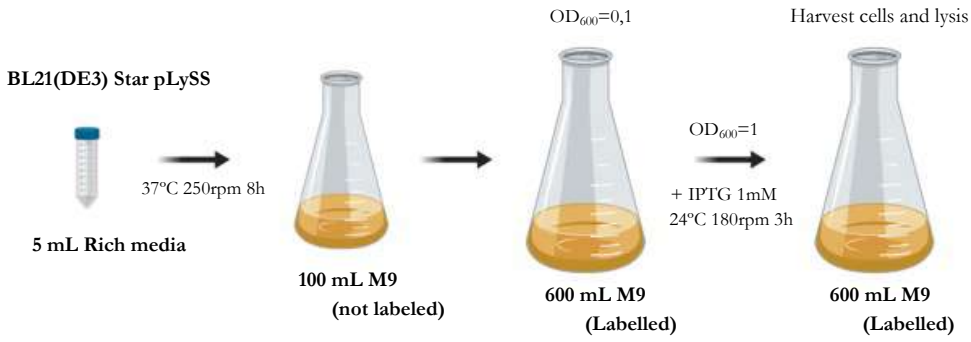


Figure 13.25. Scheme of final protocol for protein expression in minimal media. M9: minimal media (see section 21.4.3)

Two labelling strategies were used; Nitrogen labelling using $^{15}NH_4Cl$ formulated in the initial M9 medium added in 600 mL cultures, and methinonine and isoleucine labelling by adding 70 mg/L of L-methionine (methyl-^{13}C, 99%) and alpha-ketobutyric acid, sodium salt(methyl-^{13}C, 99%; 3,3-d2, 98%) to 600 mL culture when the OD_{600} reaches 0.4. Afterwards, for protein purification, isolation of the recombinant protein was achieved by Strep-Tag affinity

chromatography of the soluble fraction. In order to identify and separate the oligomeric forms of the expressed protein, a second purification step was performed using Size exclusion chromatography (SEC).

In previous section, proteins were purified and kinetics were determined using Tris 50 mM at pH 8.5 and NaCl 0.3 M as a buffer. For NMR studies pH was decreased to 7.5 to permit a more clear NMR elucidation.

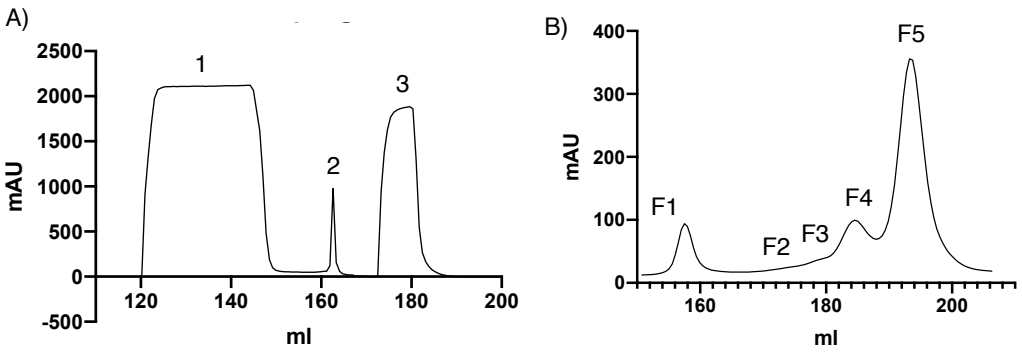


Figure 13.26. A) Chromatogram of Strep-trap purification of *Vc*CDA CD D39S. Resulting chromatogram from affinity chromatography in which three cycles of loading and elution were performed. 1) loading and washing, 2) elution; 3) regeneration with HABA and washing. mAU: 280nm. B) SEC chromatogram of VcCDA CD D39S. Chromatogram obtained on filtration gel chromatography using a Superdex 200 column (16/600). Five multimeric fractions are shown which F5 corresponds to the monomer.

All purified enzymes show the same profile in affinity and Gel filtration chromatographies (*Vc*CDA CD D39S, *Vc*CDA PG D39S and *Vc*CDA CD wt) (Figure 13.26). Expression in minimal media shows that the monomeric fraction (fraction 5) is the most abundant dislike the expression in rich media where dimeric fraction and oligomeric fraction are highly abundant. The expression yields obtained with the mutants were similar to those previously reported for the wild type enzyme in rich media, obtaining between 2 and 8 mg of total monomeric fraction per liter of culture.

On the other hand, the study of oligomeric systems in solution is frequently hindered by their large molecular weight, which limits the resolution of NMR spectra due to fast transverse spin–spin relaxation rates, and complicates the application of NOE methods for the derivation of interface restraints. Moreover, the symmetry inherent to such protein complexes gives rise to spectral degeneracy, as the equivalent spin sites experience the same chemical environment among the protein subunits. Despite the spectral simplification resulting from symmetry, analysis of corresponding NOE spectra can be more complicated as cross peaks can represent intra- or intermolecular interactions [210,223]. For further NMR studies is

essential to confirm the purity of the monomeric fraction. Analysis of oligomers in pure samples was done using HPLC with a SHODEX column [224]. Shodex, silica-based, diol-coated, size exclusion columns are designed for the separation and analysis of proteins and other biocompounds. Calibration curve was obtained using standard proteins of known weight. Single peak appear at retention time of 12.6 indicating the high purity of the sample as monomer for both proteins.

Having isolated the monomeric fraction for subsequent NMR analysis, the specific activity of two different samples of active and unlabeled *Vc*CDA CD wt was determined, one expressed in rich media and in minimal media to detect possible changes in protein topology which can affect the protein activity. As presented in Table 13.8, enzymes expressed in rich media or in minimal media showed the specific activity using DP4 as a substrate. Specific activity was determined by HPLC-MS.

Table 13.8. Specific activity of *Vc*CDA Cd from different expressions systems using DP4 as a substrate. Buffer: Tris 50 mM pH 7 5, NaCl 0.3 M.

Protein	Specific activity (s^{-1})
***Vc*CDA CD wt (rich medium)**	0.97 ± 0.2
***Vc*CDA CD wt (M9 medium)**	0.81 ± 0.3

For NMR studies, enzymes V*c*CDA CD D39S and *Vc*CDA PG D39S were subsequently expressed and purified using previously optimized protocol with previously selected labelling.

13.4.3. Elucidation of loop 5 dynamics: *Vc*CDA CD wt and *Vc*CDA CD PG

NMR experiments were done in collaboration with Dr. Óscar Millet, Dr. Jesús Jimenez-Barbero and Dr. Ganeko Bernardo, CIC BioGUNE, Bilbao.

In this chapter, we sought to determine and characterize the dynamics at multiple time scales of chitin deacetylase *Vc*CDA using NMR spectroscopy. The main objective is unravelling the molecular mechanism of loop 5 dynamics in PG mutant compared with wt enzyme. To solve this, relaxation dispersion techniques were employed.

Relaxation dispersion (RD) experiments resolve motions on micro- to millisecond timescales, encompassing biomolecular motions associated with ligand binding, enzymatic catalysis, and domain-domain opening. These experiments provide structural, kinetic, and thermodynamic

information on “invisible” (lowly populated) excited conformational states. Relaxation dispersion can be applied not only to single biomolecules but also to protein-ligand complexes to study the kinetics and thermodynamics of association and dissociation. Importantly, RD experiments are able to sense exchange events with very skewed populations, where only the major species is apparently observable in the NMR spectrum.

A simple explanation of the experiment is illustrated in Figure 13.27. The essence of the experiment is the spin-echo pulse sequence (Figure 13.27.A) where, after a 90º excitation pulse, a total τCP evolution time is followed by FID registration. The 180º pulses are required to refocus any putative evolution such as chemical shift and heteronuclear J-coupling. In the absence of exchange, the intensity of the signal right before the FID registration will be proportional to the maximum intensity (that would be right after the 90º pulse minus the intensity losses due to T2 relaxation during the τCP period. Moreover, in the absence of exchange, the measured intensity will be invariant to any variation in the number of 180º pulses that are allocated during τCP (i. e., all the experiments in Figure 13.27.A would render the same result). Yet, in the presence of chemical-conformational exchange in the microsecond-millisecond time scale, the chemical shift experienced by the nuclei is dependent on the probability of shifting by a (stochastic) chemical exchange event. Because the defocusing time is different depending on the number of 180º pulses, the several schemes shown in Figure 13.27 will produce different responses as a function of the exchanges events, encoded as an extra loss of intensity. Specifically, the fewer the number of 180º pulses, the more intensity loss due to the exchange process. Of note, the number of 180º pulses in the sequence in normally expressed as a B1 frequency in Hz according to the expression B1(Hz) = 1/(4·np), where np is the number of 180º pulses.

A typical experiment monitors 16 or 18 different pulse schemes that provide a relaxation dispersion profile (RD). The experiment is run in a 2D mode, based on a $^{1}H,^{15}N$-HSQC (for backbone amide) or a $^{1}H,^{13}C$-HMQC (for methyl groups) so the signals are separated and can be analyzed independently. An analytical equation is available to fit the experimental data, assuming a two-state dynamic equilibrium. Such fitting of the RD profile provides information about the exchange rate (k_{ex}), the populations of the two different states (p_A and p_B) and the chemical shift difference between the states (Δw). This information is obtained independently for each of the measured probes (i. e. an amide or methyl group). Yet, often the RD profiles of different probes can be attained to be fitted collectively when collective motion is suspected.

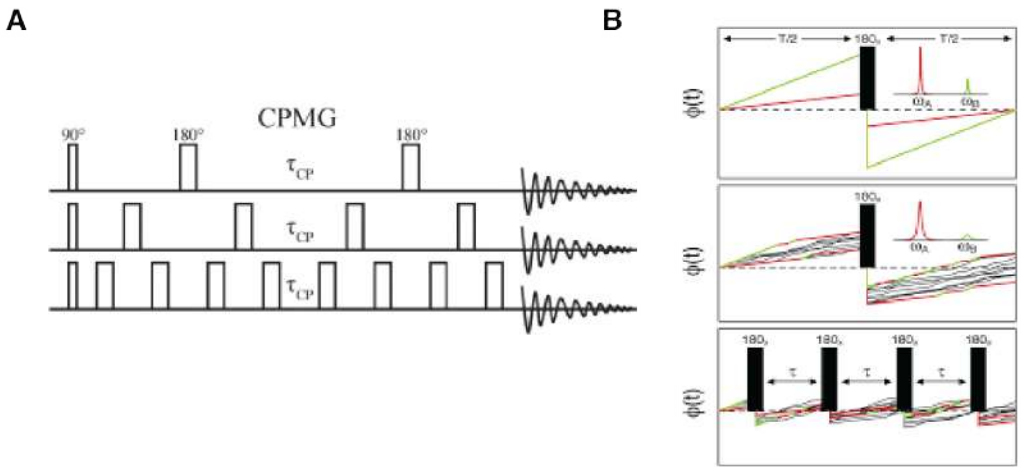


Figure 13.27. A) Basic RD experimental scheme. The total echo time is τCP. B) The chemical exchange is sensittive to a variation in the number of 180° pulses allocated during τCP.

Finally, it is important to emphasize that the RD profile fitting benefits from the use of relaxation data at multiple fields. The reason is because the chemical shift difference between two exchanging sites ($\Delta\omega$) is a function of the field strength, while the population of the substates (p_A and p_B) are not. The fitting equation only contains the $\Delta\omega \cdot p_B$ term, that needs field dependence to be adequately resolved.

Methyl TROSY experiments enable the possibility of recording high-quality data on complexes that are far over 100 kDa in molecular weight, thanks to a relaxation compensation of relaxation mechanisms that produce sharp lines for the methyl group. Different approaches can be used to assign methyl resonances in these spectra, including site-directed mutagenesis.

13.4.3.1. Evaluation of protein backbone amide relaxation dispersion

First, we explored the possibility of measuring backbone amide relaxation dispersion (RD) experiments but the quality of the $^{1}H,^{15}N$-HSQC spectrum (Figure 13.28) did not allow to do it. There are several possible reasons for the chemical shift heterogeneity observed in the spectrum, but later experiments with methyl groups confirmed that this is due to pervasive dynamics. $^{1}H,^{15}N$-HSQC spectrum experiments suggested a global dynamics of *Vc*CDA that disguise the dynamics of loop 5.

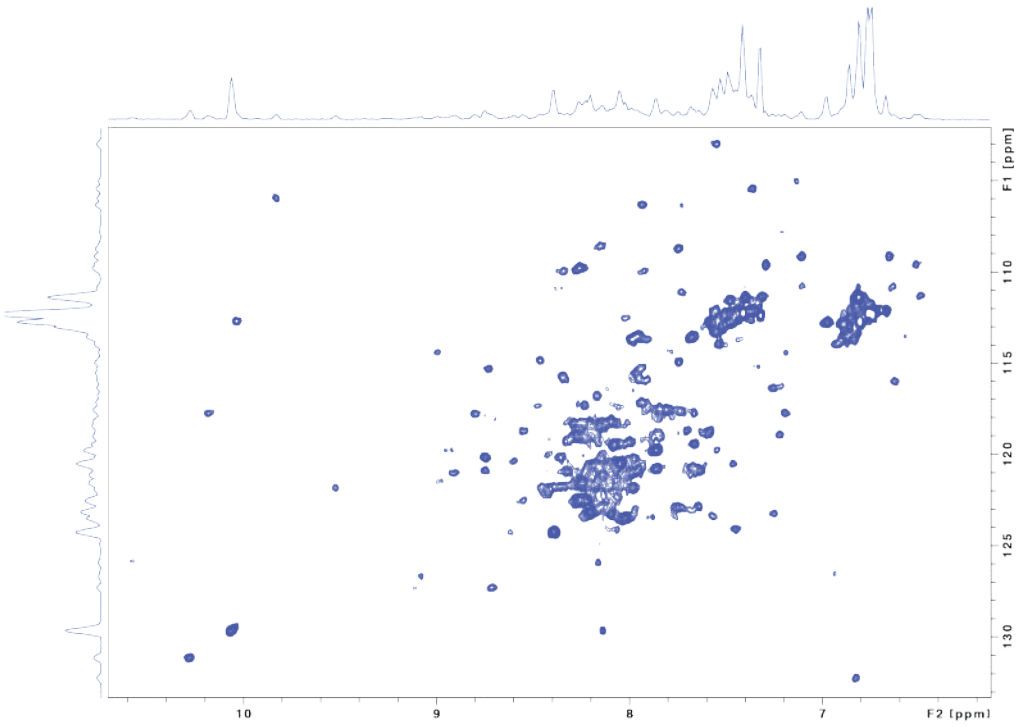


Figure 13.28. $^{1}H,^{15}N$-HSQC spectrum of *Vc*CDA (800 MHz, 298K). Only the signals from flexible regions are observed. The heterogeneous intensity patter is compatible with the protein undergoing pervasive motions.

Based on these results, we explored the potential of the methyl group as a probe to investigate protein dynamics. To that end, we used a ^{13}C-Met Met/Ileδ *Vc*CDA samples spectrum were to monitor all the dynamic events. We investigated two samples: *Vc*CDA CD/D39S (referred as to CD wt enzyme) and *Vc*CDA CD PG /D39S (referred as to CD PG). Samples were prepared according to section 13.4.2. This strategy turned out useful since the methyl TROSY experiment renders a good signal dispersion and optimal signal linewidths. Based on this, Table 21.12 summarizes the experiments collected.

13.4.3.2. Identification and analysis of a second population in CD wt

Due to the ambiguous results obtained in $^{1}H,^{15}N$-HSQC spectrum, Methyl TROSY experiments were proposed to elucidate the conformational states of *Vc*CDA enzyme. The labelling strategy was designed to monitor all the Met signals (5 in total), and all the δ-methyls in the Ile residues (18 in total). As expected, the methyl TROSY spectra shows a good signal dispersion for most of the Met residues, showing a bit more of overlap in the Ile region. Methionine's and isoleucine's of *Vc*CDA are highlighted in Table 13.9.

Table 13.9. Methionine's and isoleucine's in CD wt and CD PG. Nt: N-terminal methionine.

	Total	Loop 1	Loop 2	Loop 3	Loop 4	Loop 5	Loop 6	Others
Met	**5 (4 + Nt)**	M72	M98,	0	**M242**	0	0	Nt, M332
Ile	**18**	0	I99, I122	0	**I240**	I270, I273	M306	I34, I50, I93, I103, I126, I144, I153, I213; I218, I269, I319, I321

Methyl TROSY spectra for CD wt and CD PG enzymes were done with and without presence of DP4 substrates (Figure 13.29 and Figure 13.30). As shown in Figure 13.29 and Figure 13.30.A-B, the number of peaks for CD PG (red peaks) roughly corresponds to the expected number considering the sequence but, for CD wt (blue peaks), a second population clearly emerges (marked with a box in Figure 13.29 and with an asterisk Figure 13.30). CD wt spectra is compatible with equilibrium of two species (major and minor in a slow exchange).

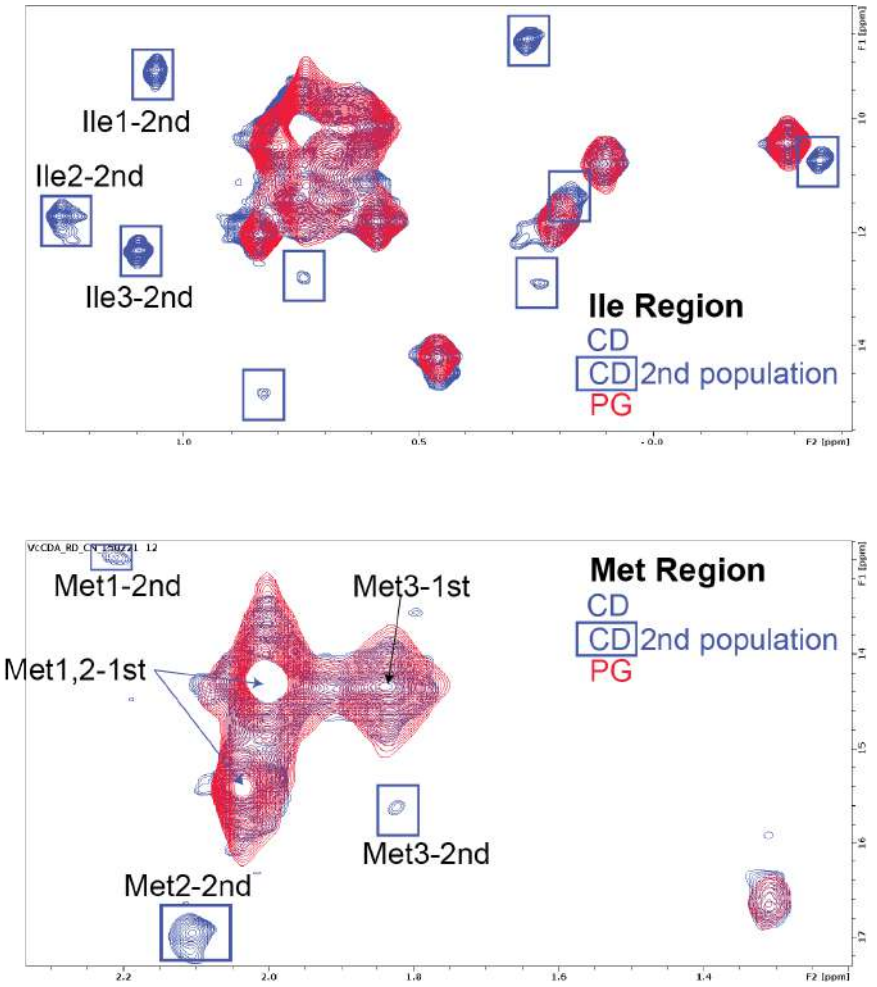


Figure 13.29. Methyl TROSY spectra of ^{13}C-Met Met/Ileδ (800 MHz, 298K) for CD wt (blue) and CD PG (red). CD wt shows a second set of signals (marked in a box).

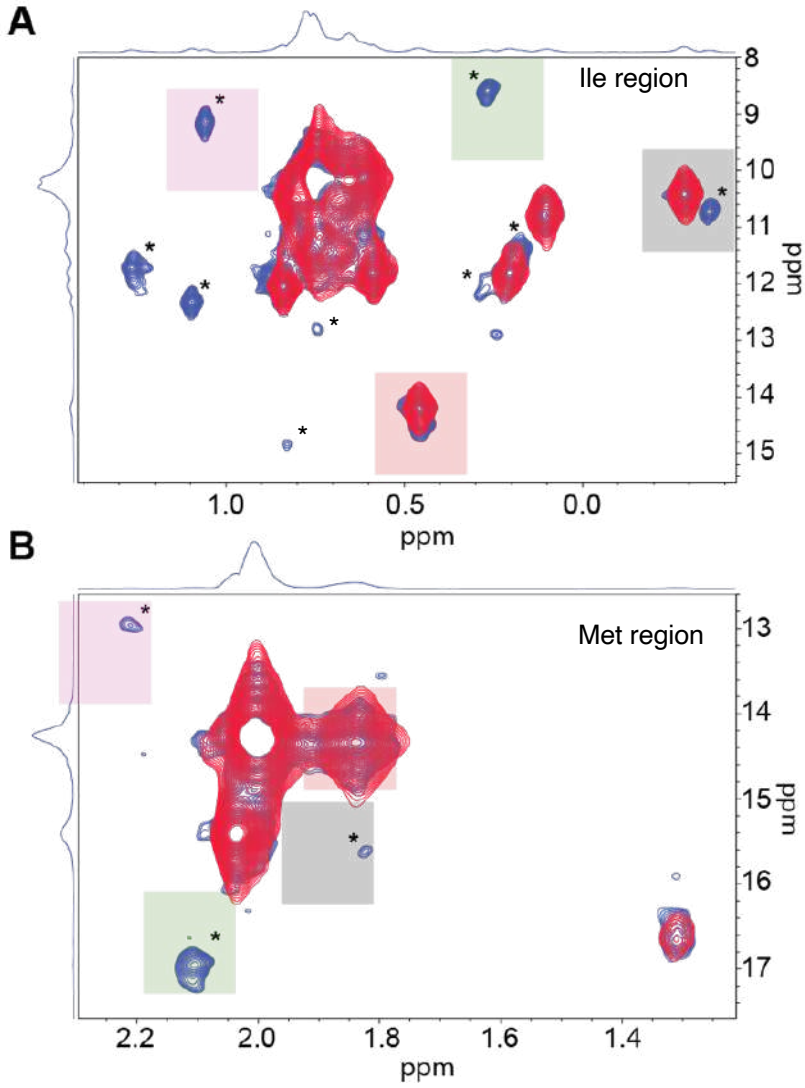


Figure 13.30. A-B) Methyl TROSY spectra of ^{13}C-Met Met/Ileδ (800 MHz, 298K) for CD wt (blue) and CD PG red). CD wt shows a second set of signals. The regions for the δ-Ile and Met residues are shown in panels A and B respectively. Peaks assigned to the minor species are identified with an asterisk. The bound signals follow the same color than the regions, while peaks from CD wt and CD PG are colored in blue and red respectively.

Several observations can be extracted for the joint analysis of both spectra: the spectrum of CD wt is compatible with an equilibrium between two species, one major and a second minor species, that are in slow exchange (ms to seconds). Furthermore, according to the average intensities, the minor population would represent 20-25% of the total population of CD wt. Finally, the spectrum of CD PG only shows a single species, chemically equivalent to the major species of the CD wt spectrum. This approach permits us to clearly identify and analysis a second population in CD wt that is absent in CD PG.

To demonstrate that both species are in slow exchange we carried out a NOESY spectrum that should also show the exchange peaks between the different species (EXSY spectroscopy). As shown in Figure 13.31, two of the peaks connect chemical shifts of peaks belonging to two different species and are exchange peaks. The NOESY mixing time was 500 ms, suggesting that the slow equilibrium is in the order of the hundreds of milliseconds timescale.

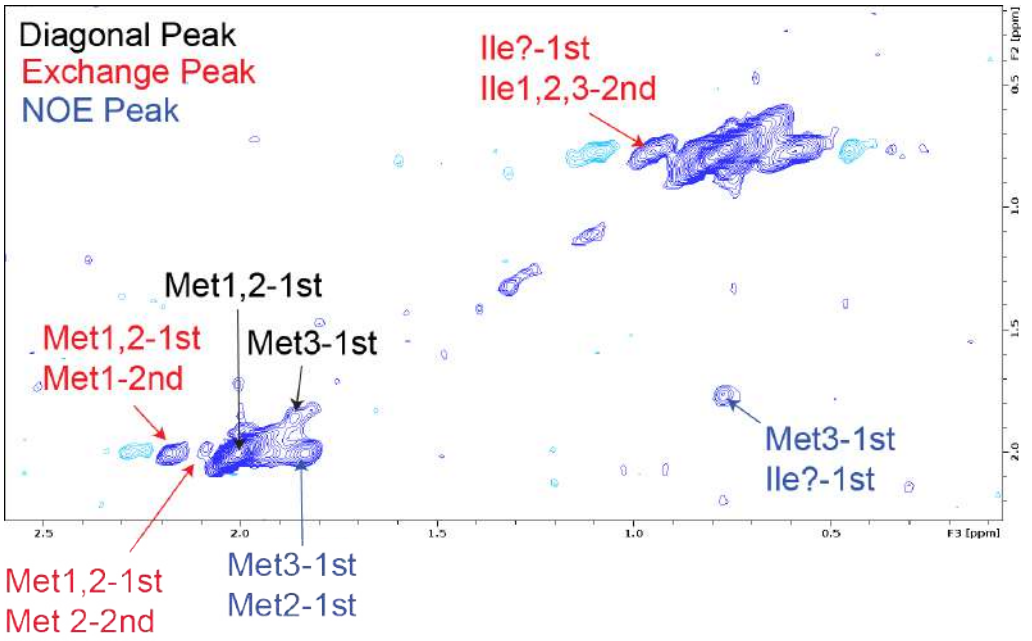


Figure 13.31. NOESY/EXSY spectrum of CD wt. The square-highlighted peaks connect minor and major species peaks and, therefore, are exchange peaks. 1st species corresponds to major species and 2nd to minor species.

13.4.3.3. Binding of substrates to CD wt and CD PG

The binding of the substrates DP2 and DP4 to CD wt and CD PG was then investigated. Due to the low affinity (K_M values in the range of 1-6 mM, see Table 13.3), substrates were added in a large excess, only limited by the system's solubility.

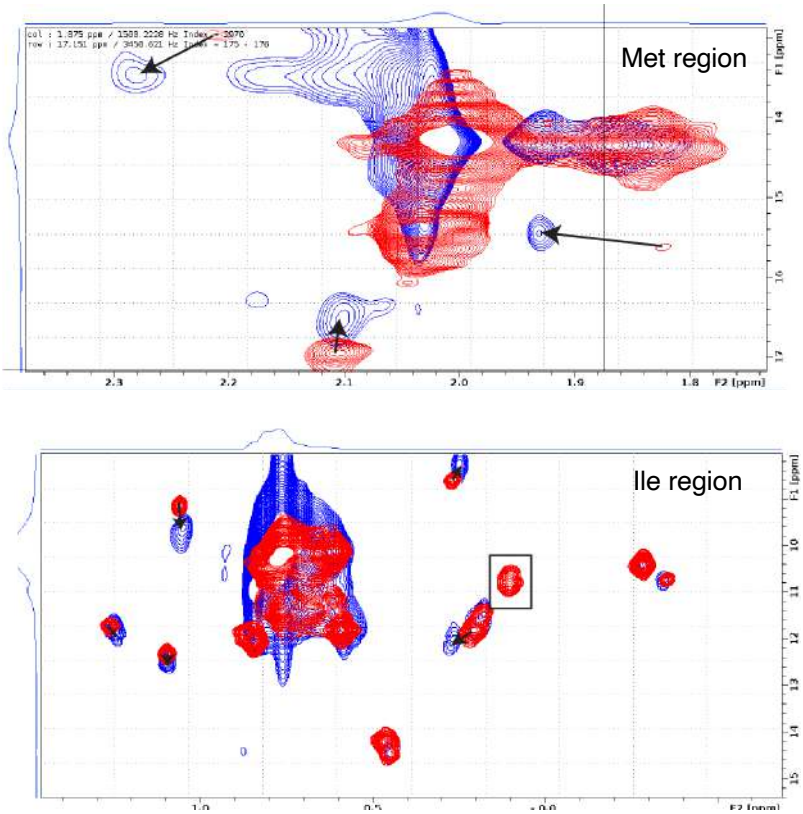


Figure 13.32. Methyl TROSY spectra of ^{13}C-Met Met/Ileδ (800 MHz, 298K) for CD wt (red) and CD wt + DP4 (red) in presence of DP4 substrate. Notice that the shifts induced upon substrate binding (black arrows) are mostly produced in the minor species.

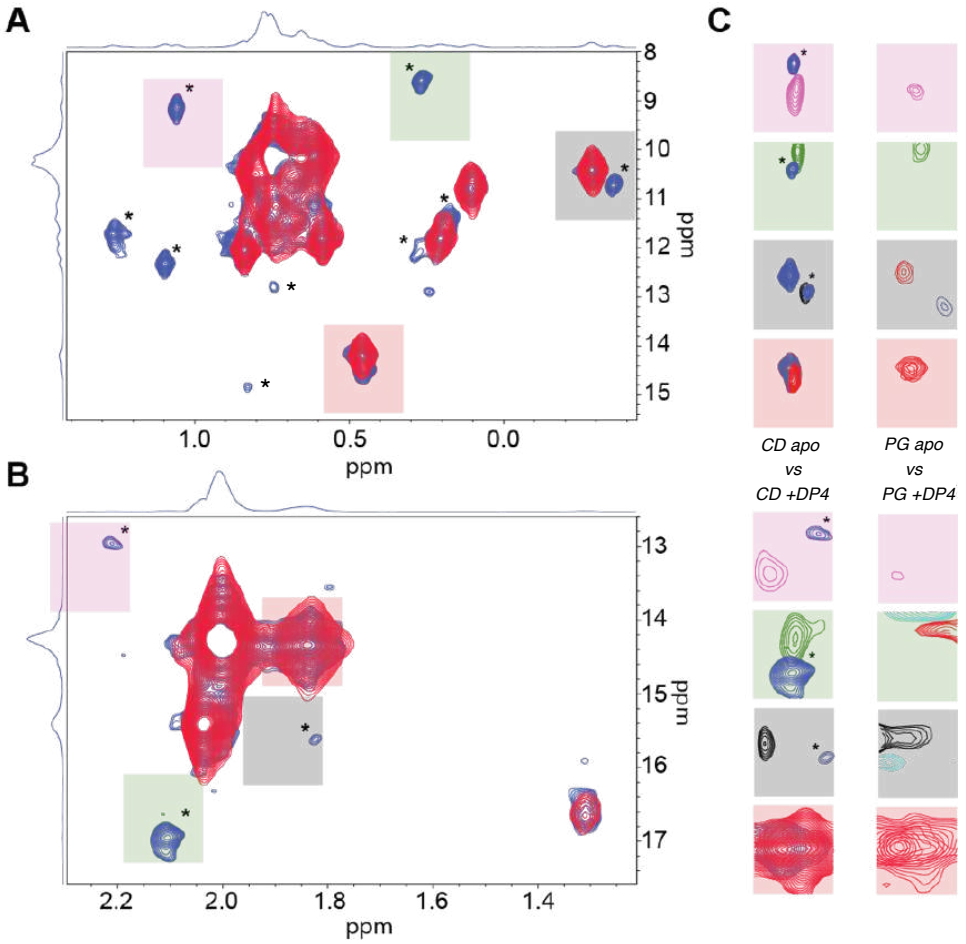


Figure 13.33. A-B) Methyl TROSY spectra of ^{13}C-Met Met/Ileδ (800 MHz, 298K) for CD wt (blue) and CD PG (red). CD wt shows a second set of signals. The regions for the δ-Ile and Met residues are shown in panels A and B respectively. Peaks assigned to the minor species are identified with an asterisk. The bound signals follow the same color than the regions, while peaks from CD wt and CD PG are colored in blue and red respectively. C) Selected regions to show the overlap between apo and DP4 bound CD wt (left) or CD PG (right). Blue color peaks indicates apo enzyme and coloured peaks indicated enzyme-DP4 complex. The bound signals follow the same color than the regions, while peaks from CD wt and CD PGare colored in blue and red respectively.

Results for DP2 and DP4 are very similar, but we analyze in more detail results for DP4 because, even though it is not the natural substrate for the enzyme, it is larger and the effects are more apparent.

Figure 13.32 (marked with arrows) and Figure 13.33.C show (insets left) selected regions of the spectrum, showing CD wt (apo form, blue) and CD + substrate (DP4, colored). Remarkably the bound spectrum produces shifts in the minor species, indicating that this species is conformationally closer to the bound state. These shifts were also observed for DP2 substrate. Consistently, the peaks in the major species seem to be largely insensitive to

the substrate, but a quantitative analysis of the intensities is not possible due to the signal overlap. A closer view of unbound vs bound species in the CD wt spectrum is shown in Figure 13.33 where, as example, where the peak at δ 2.1/17 ppm in Met region of the minor unbound species shifts to the bound species. This peak is later assigned to Met 242 of loop 4. Similar shifts are observed for a number of other peaks as shown in Figure 13.33.C.

Likewise, Figure 13.33.C (same colour code was used) shows the insets (right) for the same regions for CD PG titrated with DP4. Binding is in slow exchange with the apo form and, remarkably, the bound peaks appear at the same chemical shifts than CD wt, indicating a similar structural complex (as expected). The comparison between the two bound forms clearly reflect this fact.

Analysis of NMR spectra at ms-s scale (in the chemical shift time scale) show a slow equilibrium in CD wt enzyme between two species: a major species and a minor species which is structurally close to the bound form. These two species are conformationally different. This results clearly suggest a ‘’conformational selection’’ mechanism where the substrate binds biasedly the CD wt minor conformation and drive the equilibrium towards the catalytic event. On the other hand, CD PG mutation causes a change in binding towards an ‘’induce fit’’ mechanism since there is no presence of any minor conformation and binding events involve the unique major conformation (Figure 13.34).

Figure 13.34. a) In induced-fit binding, the change between the conformations P_1 and P_2 of the protein occurs after binding of the ligand L. The intermediate state P_1L relaxes into the bound ground state P_2L with rate k_r, and is excited from the ground state with rate k_e. (b) In conformational-selection binding, the conformational change of the protein occurs prior ligand binding. The intermediate state P_2 is excited from the unbound ground state P_1 with rate k_e, and relaxes back into the ground state with rate k_r.

According to the number of peaks in the minor species in the CD wt species, most of the Met and Ile residues participate of this slow equilibrium, suggesting a major global rearrangement of the enzyme. It should be stressed that this major global dynamics of the protein suggest a synchronization of loops in terms of dynamics (especially loops 4 and 5).

Finally, we hypothesize that the minor species is an advanced conformational state in the enzymatic reaction pathway, explaining the higher affinity of CD wt for substrates. Conversely,

because of the induce fit mechanism, CD PG elicits lower affinity for substrates (higher K_M values relative to the CD wt enzyme, see Table 13.3).

13.4.3.4. Analysis of fast dynamics by relaxation-dispersion experiments

Based on the open-close conformation of loop 4 between apo and bound CD wt enzyme by x-ray crystallography, we expect that at least loop 4 will be involved in the conformational equilibrium of the apo enzyme as observed by methyl TROSY NMR experiments. Site directed mutagenesis.was used to identify relevant peaks. Met242 and Ile240, located in loop 4, were mutated to alanine and leucine (M242A and I240L). (see protocol 21.3.2.).

First, the assignment by site directed mutagenesis of the relevant residues in loop 4 allowed identifying Met242 as an isolated peak, while Ile240 remains in the overlapped section of the Ile region of the spectrum, and it cannot be unequivocally assigned (Figure 13.35). mutants were prepared according protocol 21.3.2.

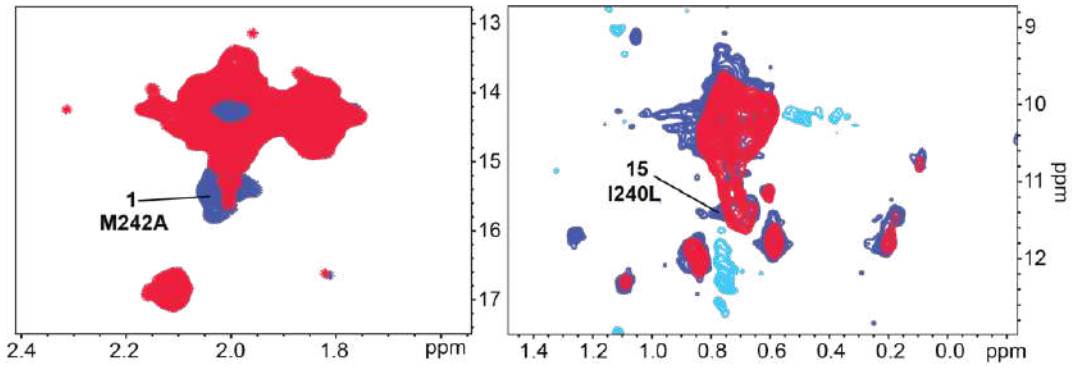


Figure 13.35. Methyl TROSY spectra of 13C-Met Met/Ileδ (800 MHz, 298K) for mutants M242A (left, red) and I240L (right, red) of CD VcCDA and compared to the apo form (blue). Peak numbers (1-M242 and 15-I240) correspond to peaks in Figure 13.36.

The dynamics of the system using relaxation dispersion experiments and using the methyl groups as a relaxation probe were investigated. This experiment is sensitive to motions in the μs-ms time scales. Figure 13.36 shows the peak picking used in the different experiments to investigate relaxation dispersion. A total of 27 peaks were monitored for CD wt and 20 peaks for CD PG (Figure 13.36).

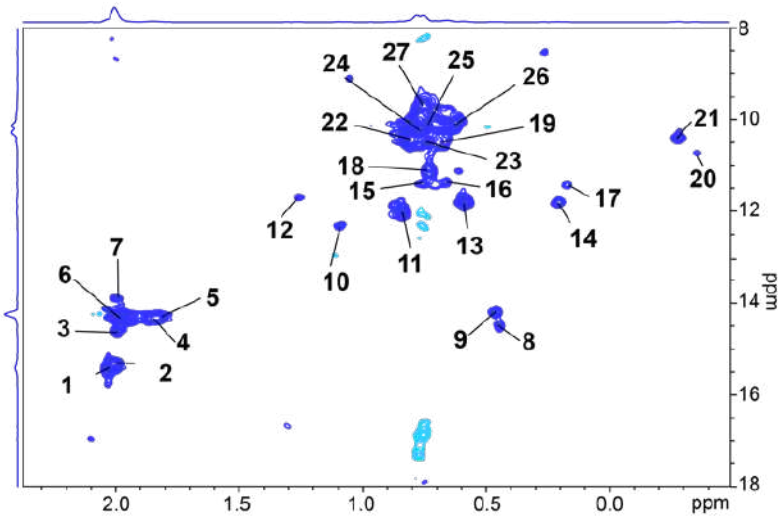


Figure 13.36. Spectrum of CD wt with all the peaks investigated by RD. The peaks M242 and I240 correspond to peaks 1 and 15 respectively.

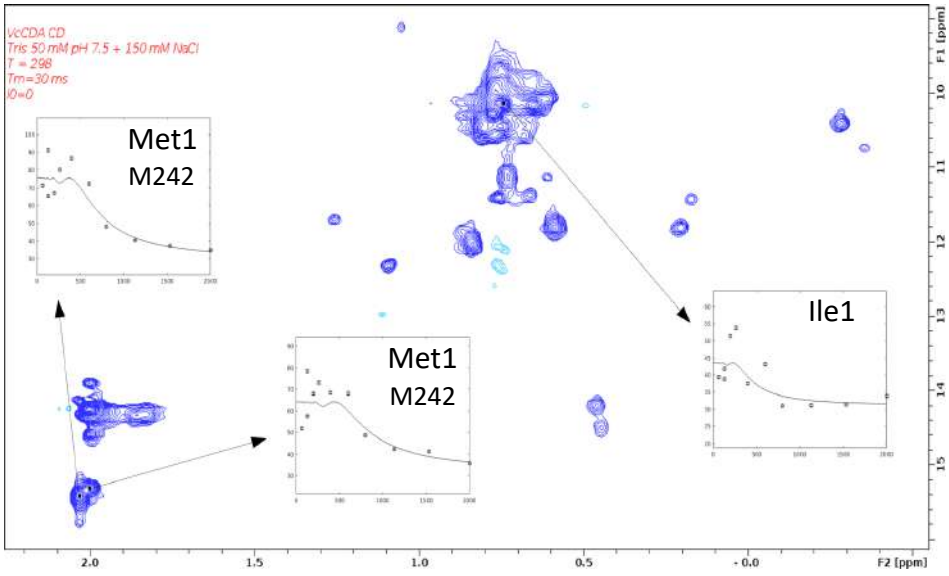


Figure 13.37. RD to investigate dynamics in a faster timescale (µs-ms). Examples of RD profile for Met242 (Met1).

The RD experiments have been collected at two fields (600 and 800 MHz) to ensure a good fitting of the data. For the CD wt protein, we could fit the data to a motion with k_{ex} in the range of 120 -180 s^{-1}, affecting seven residues from loop 4 (Met 240) and other loops, all from the major species. When this experiment was repeated for the CD PG mutant, the same residues showed exchange, but at a faster rate (k_{ex} = 850 s^{-1}).

Both for CD wt and for CD PG, the RD experiments in the presence of an excess of substrates is largely quenched and only peak 24 shows an RD profile with a very low value of kex (< 100 s^{-1}). It is educational to monitor the effect of exchange in M240 of loop 4, as shown in Table 13.10:

Table 13.10. Exchange analysis of M240.

Condition	k_{ex} (s^{-1})	pB (%)
CD VcCDA	120	3
PG VcCDA	814	8
CD VcCDA + DP2	No exchange	
CD VcCDA + DP4		
PG VcCDA + DP2		
PG VcCDA + DP4		

Taken altogether, in CD wt, the timescale of the fast motion (at least in loop 4, k_{ex} = 120 s^{-1}) it is slow enough to interfere with the release of the product. On this basis, in CD wt enzyme loops 4 and 5 slow dynamics affect the k_{off} for the substrate release after the catalysis takes place (k_{off} is the unimolecular rate constant of the EP complex dissociating to regenerate free enzyme and product). In CD PG, the timescale of the fast motion (at least in loop 4, k_{ex} = 850 s^{-1}) should be fast enough not to interfere with the release of the product and permit a fast catalysis event due to the product having a greater facility to leave the active center once the enzymatic catalysis has been carried out. This mechanism of binding would impact essentially on DP4 because it is bulkier than short substrates due to the steric hindrance. This approach distinctly explains the higher values of k_{cat} for CD PG when catalyzing DP4 compared with wt enzyme. Interestingly, the exchange process is abrogated in the presence of an excess of DP2 or DP4 for both CD wt and CD PG (Table 13.10), perhaps due to the formation of stabilizing interactions between the loops and the substrate to freeze the loops into a single conformation.

13.5. Molecular dynamics simulations of *Vc*CDA, targeting loops dynamics.

In order to elucidate, from the molecular dynamics approach, the structural determinants of the PG mutant, the dynamics of protein loops in *Vibrio cholerae* chitin deacetylase protein structure has been simulated by means of force-field based Molecular Dynamics simulations[2]. Both the wild-type and PG were simulated in the absence of substrates bound to the active site. Differences in protein loop flexibility and substrate binding accessibility have been analyzed and compared between all the simulations aimed at elaborating a dynamic model support the dynamics of the enzyme loops obtained in the NMR results.

The whole protein structure was simulated without position restraints, allowing all protein atoms to move freely at 300 K. All simulations follow a standard simulation protocol in which the initial structure of the protein, explicit solvent and/or ligands are initially equilibrated to the chosen force-field, followed by an equilibration of the simulation conditions (temperature and pressure). Once equilibrated, simulations are extended for a period of 300 ns for free MD configuration.

The displacements and fluctuations of the series of loops surrounding the active site of *Vc*CDA are evaluated after all simulations have been completed. Protein displacements are measured by means of the root-mean-squared deviation (RMSD) from the equilibrated structure. Per residue averaged fluctuations of the loops are measured by means of root-mean-squared fluctuations (RMSF) around the equilibrated structure. In the latter case, the internal fluctuations of the protein core (excluding the loops) are removed prior to the RMSF calculations.

Figure 13.38 shows the fluctuations of each amino-acid residue of the whole catalytic domain of *Vc*CDA along the free MD simulations of both the wild-type and PG. It should be stressed how the major fluctuations of the protein are localized in the regions of the loops for both enzyme variants. Interestingly, Loop 5 is the less fluctuating one, and does not show significant fluctuations for the wild-type form of the enzyme.

[2] MD-Simulations were done by Dr. Xevi Biarnés in the Laboratory of Biochemistry.

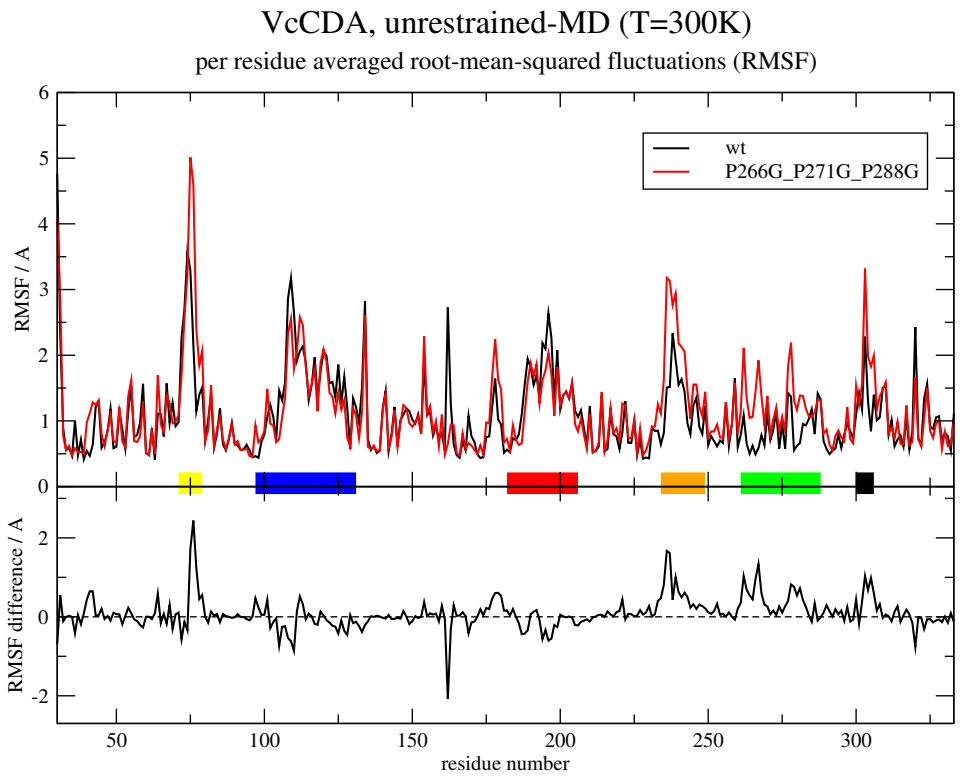


Figure 13.38. Per-residue averaged root-mean-squared fluctuations of catalytic domain of VcCDA along free MD simulations. Top panel: absolute RMSF values for both wild-type and triple mutant P266G/P271G/P280G variant. Bottom panel: differences of RMSF values between both variants. Color bars in the horizontal axes indicate the location of the six loops surrounding the active site of VcCDA: LOOP1 (yellow), LOOP2 (blue), LOOP3 (red), LOOP4 (orange), LOOP5 (green) and LOOP6 (black). The negative peak at position 162 corresponds to an important side-chain rotation of a solvent exposed Tyrosine in thisposition. Tyr162 is located in the opposite site of the catalytic center, thus this movement is not considered relevant for this discussion.

In order to better compare the differences in fluctuation between wild-type and PG, the differences in RMSF for each residue are shown in Figure 13.38 bottom panel. It is immediately apparent that the behaviour of Loop 2 and Loop 3 is equivalent for both forms of the enzyme, with equivalent fluctuations. Indeed, these two loops form the catalytic pocket platform, where part of the catalytic machinery and substrate binding spots reside. LOOP1 is the most fluctuating one, and these fluctuations are intensified for the triple mutant variant. Concertedly Loop 6 (which is in the close vicinity of Loop 1 in the structure) also shows an increase of flexibility in the triple mutant variant. Loop 1 and Loop 6 define the binding subsites for the non-reducing end of the substrate.

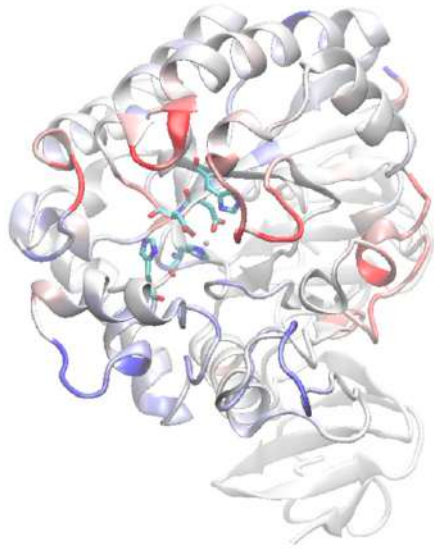


Figure 13.39. *Vc*CDA structure in cartoon representation colored according to differences of RMSF values between both variants: red, high flexibility; blue, low flexibility.

On the other side of the binding pocket, Loop 4 and Loop 5 define the binding subsites for the reducing end of the substrate. These two loops, which are close in the structure, also show a concerted increase in flexibility in the triple mutant variant of the enzyme. The flexibility increase of Loop 5 is localized in a short segment between positions 260 and 270. This results could explain the low effect of P280G mutation in loop 5 flexibilization by triple PG mutant.

Taken altogether, the analysis of RMSF differences suggests that the introduction of a PG mutant in Loop 5, induces a series of small conformational changes in the loop that propagate along the whole binding cavity, involving the loops at both sides of the substrate binding cavity (Loop 1, Loop 6, Loop 4 and Loop 5), while the catalytic platform (Loop 2 and Loop 3) remain stable. Additional binding pockets may emerge for additional substrate units at the reducing end (positive subsites).

13.6. Closing remarks of *Vc*CDA enzyme loop dynamics

From the random mutagenesis epPCR/HTS approach to modify the substrate specificity of *Vc*CDA towards longer substrates, a single mutant K275E was identified in loop 5, which increases the activity for DP4 substrate. This mutation breaks an internal loop 5 salt bridge and clearly suggests that flexibility of loop 5 could play a key role in substrate specificity of the *Vc*CDA enzyme.

In order to confirm this hypothesis, an internal disulfide bond that anchors loop 5 with loop 3 limiting the movement of this loop 5 was evaluated. For short substrates, this mutant shows the same activity as wt enzyme but for long substrates such DP4, the mutant shows almost no activity compared with wt enzyme suggesting that this substrate does need the creation of additional subsites that the movement of loop 5 would provide. This result strongly points out the role of loop 5 dynamics in the acceptance of long substrates.

We sought to investigate in more detail the 3D structure of loop 5 in order to modify the specificity of the enzyme towards longer substrates. Following the idea to increase the flexibility of loop 5, three prolines were identified and mutated to glycine (PG mutant). This mutations completely reverts the activity of the enzyme towards longer substrates increasing the k_{cat} values almost 600% compared to the wt enzyme but substrate affinity (K_M) has been affected. Individual proline to glycine mutants were analyzed to determine the role of each proline in the final construct. Results identify the P266 and P271 as a key contributors to final enzymatic activity. P280 almost has no effect in the final PG mutant.

NMR techniques were selected to elucidate the overall dynamics of PG mutant that permits us to explain the higher values of enzyme activity (k_{cat}) and the decrease of substrate affinity (K_M). Protein dynamics were evaluated using Methyl TROSY experiments. The methyl-TROSY spectrums suggest that the wt enzyme has two conformations in slow equilibrium and that minor conformation is close to the bound enzyme-substrate complex, whereas, the PG mutant has a single conformation that shifts to the equivalent minor conformation in the wt enzyme upon substrate binding. This would suggest that the minor species is an intermediate conformation in the enzyme-substrate binding reaction coordinate that explains the lower values of K_M in wt enzyme in comparation with PG mutant.

On the other hand, relaxation dispersion experiments reveal that a significant fraction of the methyl set undergoes motions in the microsecond timescale. Such residues always belong to

the major species and account for about 20% of the methyl probes of this specific conformation. The same residues show exchange in both CD- and PG *Vc*CDA, but the average exchange rate (k_{ex}) is slower for CD wt (120 ± 60 s^{-1} out of 5 residues in CD wt versus 850 ± 125 s^{-1} out of 7 residues in CD PG). Residues in loop 4 (M242 and I240) were assigned by mutation and show conformational exchange, along with residues from other regions of the protein, likely belonging to the other loops. The timescale of the fast motion for CD PG mutant (kex = 850 s^{-1}) should be fast enough NOT to interfere with the release of the product and permits a fast catalytic event due to the substrate has a greater facility to leave the active center once the enzymatic catalysis has been carried out.

Taken all previous results together, we can conclude that not only the shape of the loops surrounding the active site are key for the structural determinants in substrate specificity, but also the dynamics in different timescale motions are involved in enzyme specificity. On the other hand, NMR results confirms that *Vc*CDA wt enzyme operated under conformation selection mechanism instead of induce fit mechanism suggested by crystallographic structure. In addition, this conformational selection mechanism disappears in PG mutants affecting the substrate affinity.

CHAPTER 4. ENGINEERING VcCDA ENZYME TOWARDS NEW DEACETYLATION PATTERNS

14. CHAPTER 4. Engineering *Vc*CDA enzyme towards new deacetylation patterns.

14.1. Introduction. Subsite Capping Model and Pattern of deacetylation.

The chemical and physical interactions between the different amino acid residues within a polypeptide chain are varied and dynamic. Naturally so, the folding of an enzyme to conceal hydrophobic residues from an aqueous environment is crucial to the transformation of many kind of molecules. As the product of evolution, each enzyme forms a network of highly interconnected individual amino acids and mutation of any of these residues has the potential chance to unexpectedly alter the ability of an enzyme to function. To a first approximation, it is logical to target the active site of an enzyme with a view to alter stereo or chemo-selectivity looking for new specificities able to generate new products [131,225,226].

The combined use of rational designs and random mutagenesis has been the preferred choice in most enzyme engineering studies, and has resulted in the generation of a large genetic diversity that can easily exceed one million by the random mutagenesis of just 5–6 residues. Therefore, the practical problem at this stage is the limited throughput of the screening used to identify positives, which is usually too slow to cover the genetic diversity. These rapid profiling and screening issues have become more important with the rise of synthetic biology, which requires the optimization of multiple amino acids to generate custom-made activities or specificities in target structures.

In CE4 enzymes, as it has been previously presented in section 9.3.4.1, the Subsite Capping Model determines the substrate specificity of the different chitin de-*N*-acetylases (CDA). In *Vc*CDA, the negative subsites or the non-reducing end of the substrate, are defined by loops 1, 2 and 6, while the positive subsites, or the reducing end of the substrate, are defined by the loops 3, 4 and 5. Loops 4 and 5 are quite mobile allowing the accommodation of long substrates. As already stated in the previous chapter, reorganization of positives subsites increase the deacetylase activity towards longer substrates due to increased dynamics of loop 4 and 5 but maintaining the pattern of deacetylation.

These loops are capping both the reducing and non-reducing ends of the substrate, and this constricting topology has different consequences in terms of PA. The loops 1,2, and 6 that define the negative subsites present an extensive network of interactions that efficiently block

the negative subsites, preventing any other acetylation pattern from being obtained in the deacetylation products. Essentially, the enzyme only deacetylates the second unit from the non-reducing end because there is no space for the substrate to slide along the active site towards a new binding conformation. Thus, only one GlcNAc unit can be accommodated at the catalytic site (subsite 0), where deacetylation occurs. Zooming up the area towards the non-reducing end of the enzyme formed by loop 1 (residues 65-81), loop 2 (102-131) and loop 6 (297-311) (defined in the Subsite Capping Model) [93], a complex network of hydrogen bonds has been identified (Figure 14.3). The hypothesis is that disruption of some of these interactions may result in flexible loops that open to create new negative subsites in order to accommodate substrates in a different binding mode that may a change in the pattern of deacetylation and alters the specificity of the enzyme.

The role of loop 1 and negative subsites cavity to define the pattern of deacetylation of different enzymes is clearly observed if we compare the 3D structure of *Vc*CDA and ClCDA in complex with substrates. Different binding modes were observed for *Cl*CDA due to different loop 1 conformations since the loop 1 is less anchored to the rest of the loops permitting the substrate to shift towards negative subsites generating a different deacetylation pattern compared with *Vc*CDA wt (Figure 14.1).

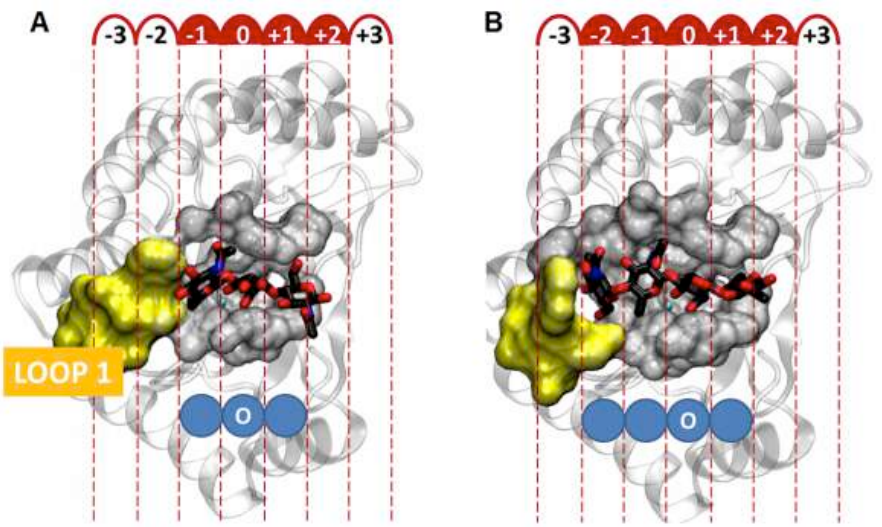


Figure 14.1. Binding subsites identification in *Colletotrichum lindemuthianum* CDA. A) model of the complex between the crystal structure of the apo form and chitotriose. B) model of the complex with chitotetraose. Surrounding loops defining the binding cavity are represented as topological surface. The rest of the protein is represented as cartoons and colored in white. The active metal cation is represented as a cyan sphere in subsite 0. The substrate is represented as thick lines.

Essentially, our strategy looks for an enzyme variant capable of rendering different patterns of acetylation to avoid the use of multiple de-*N*-acetylation and acetylation steps needed for producing specific COS products (Figure 9.5) that drastically reduce the efficiency of the reaction making the enzymatic synthesis approach not feasible.

Specifically, the main goal is to obtain an enzyme able to de-*N*-acetylate the reducing-end unit since there is no enzyme in nature able to catalyze this reaction. Single deacetylated COS at the reducing end unit are attractive molecules for specific functionalization at the reducing end site:

a) Fatty acid attachment to prepare Nod factor analogues with opposite location of the acyl group (reducing end unit instead of the natural non-reducing end unit).
b) Introduction of quaternary amines at the reducing end to enhance or modulate the antimicrobial activity of COS.
c) Combined with NodB deacetylation, the bifunctional DAAAD chitosan oligosaccharide is amenable to cross-linking by reductive amination with glutaraldehyde to access novel biomaterials.

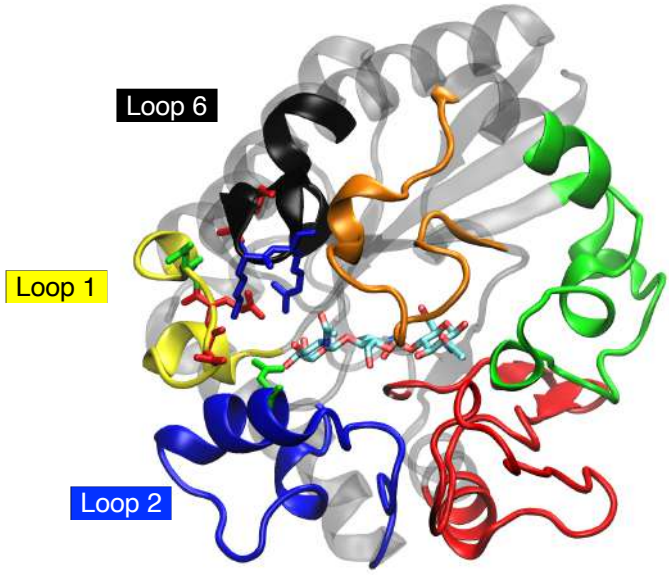


Figure 14.2. 3D structure of *Vc*CDA in complex with DP3 highlighting loop 1 (yellow), loop 2 (blue) and loop 6 (black) and residues involved in the extensive network present in negative subsites. PDB:4OUI.

In our hypothesis, Loop 1 was proposed to limit the entry of longer substrates into the negative subsites, fixing and forcing longer substrates to be placed towards the positive subsites and

consequently defining the deacetylation pattern. In order to study these phenomena, it has been proposed to eliminate the limitation exerted by loop 1, then longer substrate (e.g. DP4 or DP5) could be placed in the cavity in a new binding mode (Figure 14.2).

The engineering of the non-reducing part of the cavity can be rationally addressed mainly in two ways:

- Breaking the interactions that fix and limit the movement of loop 1.
- Directly eliminating loop 1 to open the cavity for longer substrates.

Looking at the 3D structure in detail, the complex network of interactions identified in the negative subsites (Figure 14.3) shows that two main groups of interactions are established, between loop 1 and loop 6, and between loop 1 and loop 2. The main interactions are summarized in Table 14.1. In previous work, it was proposed that if these interactions between loops 1-6 and 1-2 could be broken, provided the protein maintain the proper 3D structure, loop 1 would open allowing new binding conformations. So it was determined that the candidate residues to mutate, those whose modification maximize the loss of interactions between loops, were Lys303 and Arg304 in loop 6 and Asn119 in loop 2. In previous work, these positions were mutated to alanine to break the interactions between these loops 2 and 6 and loop 1.

On the other hand, regarding the elimination of loop 1, the main goal is to delete the sidechains that lock the loops limiting the access to potential new negative subsites, but without becoming so extreme that it destabilized the tertiary structure and consequently affect the folding of the protein. Thus, a structural alignment of chitin-deacetylases with known structure was carried out and the structure of loop 1 was compared (section 9.3.2., Figure 9.9.). The structural loop of *Vc*CDA includes residues 71-79, these aminoacids are not conserved in other enzymes, except in *Ec*PgaB, *Bc*0361 and *Cl*CDA, so a proposed deletion followed the following criteria:

- Maximize the loss of interactions.
- Do not destabilize the tertiary structure.
- Be large enough to open the cavity for new binding modes.

So it was finally decided to proceed with the deletion of residues 73 to 77 motivated by the following reasons:

- Eliminate residues Asp 74, Glu 75 and Asn 76 since they establish interaction with other residues of loops 1 and 6.
- Eliminating the entire loop can more likely lead to the loss of the tertiary structure.
- The minimum loop most similar to *Vc*CDA is *Cl*CDA. Deletion that mimics the *Cl*CDA loop but keeping the natural residues aligned from *Vc*CDA was done.

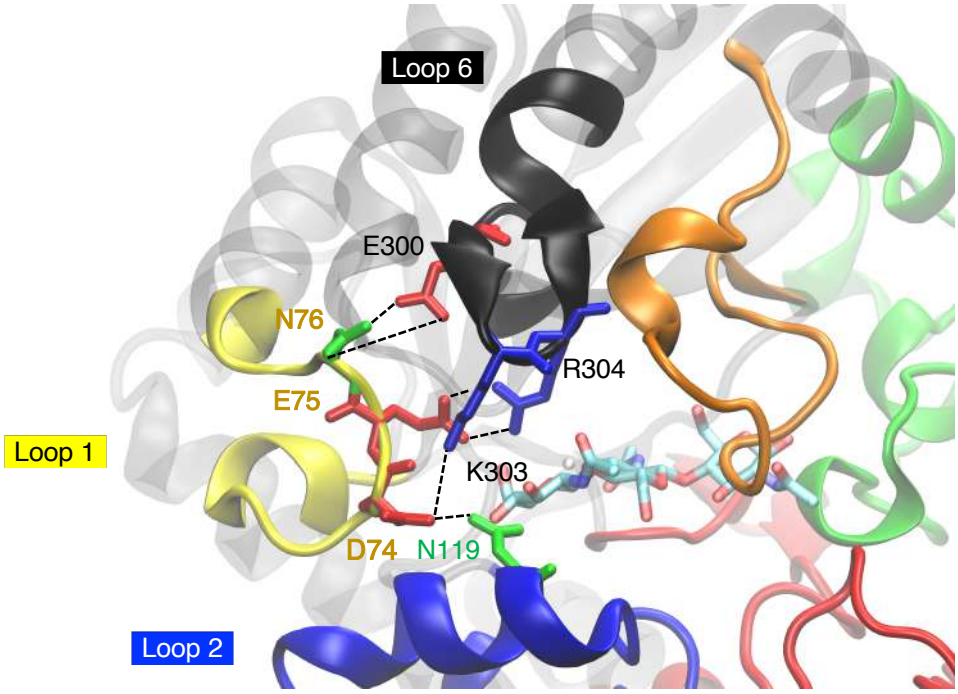


Figure 14.3. Main interactions between loop 1 (yellow), loop 2 (blue) and loop 6 (black).

Table 14.1. Distance between main interactions of loop 1, loop 2 and loop 6.

L1-L6	Distance (Å)
E75···R304	2.92
E75···R304	2.78
D74···K303	2.98
N76···E300	3.28
N76···E300	3.83
L1-L2	**Distance (Å)**
D74···N119	2.67

In previous work, mutants with deletion of loop 1 (Δ73-77), mutation N119A and mutations K303A-R304A were expressed and purified and the deacetylase activity using DP2 and DP4 was determined and compared to *Vc*CDA wt.

Table 14.2. Biochemical characterization of proposed mutants. Protein expression and deacetylase activity using DP2 and DP4 was compared to *Vc*CDA wt enzyme.

	Δloop 1 (Δ73-77)	N119A	K303-R304A
Expression	Low	Ok	Ok
Activity DP2 (%wt)	22.7	28.6	0.01
Activity DP4 (%wt)	0.26	6.10	0.18
DP2/DP4	886	47	0.8

The deacetylation activity on DP4 for all the proposed mutants falls dramatically despite the protein is still functional and well expressed. Although the activity using DP2 for the enzymes *Vc*CDA N119A and *Vc*CDA Δ73-77 does not diminish to a great extent, the activity on DP4 was practically null. More drastic is the case of *Vc*CDA K303A/R304A in which the activity on both substrates is essentially null (Table 14.2.). It was concluded that dynamics of loop 1 and its interactions are essential for the stability of the protein and the deacetylase activity of the enzyme.

After the unsuccessful results obtained by the rational design approach, we propose a directed evolution strategy focused on negative subsites of *Vc*CDA. The main goal of preparing a negative subsite libraries is finding a new variant that will be able to deacetylate longer COS substrates (DP4 or DP5) with a defined pattern of deacetylation but different from the wt enzyme.

The first strategy and the most challenging one is a series of randomized libraries that could be easily combined to perform a procedure of Iterative Saturation Mutagenesis (ISM) (Figure 14.4). Iterative saturation mutagenesis (ISM) is a well-established and efficient method for the directed evolution of functional enzymes. It reduces the necessary molecular biology work and the screening effort drastically. First, one or more amino acid positions from the chosen enzyme are assigned to multi-residue sites (i.e., groups of amino acids or "multisites", “sites”). Then, the residues in each multisite are mutated with a user-defined randomization scheme to all canonical amino acids or a reduced amino acid alphabet. Subsequently, the genes of chosen variants (usually the best but not necessarily) are used as templates for saturation mutagenesis at other multisites, and the process is repeated until the desired degree of biocatalyst improvement has been achieved. It is based on a Cartesian view of the protein structure, performing iterative cycles of saturation mutagenesis at rationally chosen sites in

an enzyme, a given site being composed of one, two or three amino acid positions [140,227]. In *Vc*CDA, an ISM approach is designed to evolve the enzyme towards new deacetylation patterns. Due to the previous results, the selected residues in loop 1, 2 and 6 are grouped in four sites to follow an ISM methodology (Figure 14.4.).

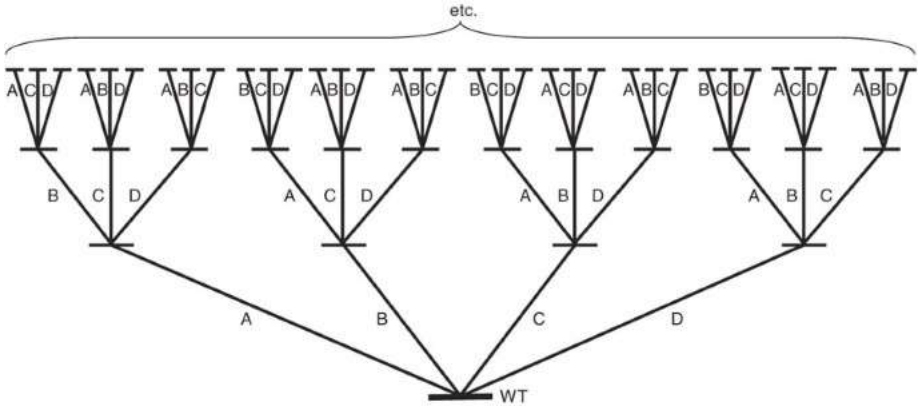


Figure 14.4. Schematic illustration of iterative saturation mutagenesis involving (as an example) four randomization sites A, B, C and D: confined protein sequence space for evolutionary enzyme optimization (redundancy in some cases is expected). Adapted from [227]**.**

The positions that are going to be evaluated are the ones indicated in the list below (Figure 14.3):

- Site I: D74X-E75X-N76X
- Site II: N119X
- Site III: E300X
- Site IV: K303X-R304X

After close detail analysis of 3D structure and taking into account the molecular insight of the key role of loop 1 in locking the access to new potential negative subsites, a decision was taken to evaluate first Site I: D74X-E75X-N76X.

Positions D74X-E75X-N76X interact with amino acids R304, K303 and E300 from loop 6 and amino acid N119 from loop 2 (Table 14.1.). These interactions define the anchor of the loop 1 to loops 2 and 6 and clearly blocks the further negative subsites that might permit the substrate to bind in a different mode that could change the substrate specificity in PA terms. Interaction N76 with N119 is the only interaction between loop 1 and 2 and is essential for protein stability as it was shown in Table 14.2.

In our directed evolution approach, genetic selection of the randomized variants is not an option, so that expensive screening must be undertaken instead. In these cases, the

screening effort increases exponentially as a function of the number of randomized positions. To alleviate this burden, it is advantageous to minimize both redundancy and the frequency of premature stop codons to reduce the library size. One step in this direction is to use NNK or NNS degenerate codons (where N= A/C/G/T, K= G/T and S= C/G) instead of NNN. For our library, NNK has been selected as a codon degeneration.

Two of these libraries (sites I and IV) were obtained using NNK randomization for each position by gene synthesis[3]. The other two libraries are single positions so they will be obtained by Quik Change Mutagenesis with NNK codon randomization. Using different sites as parallel starting libraries may provide interesting information about residue interactions in the following generations.

One of the challenges of the directed evolution approach is how to combine the randomization of the sites D74X-E75X-N76X and K303X-R304X in the fastest and cost-effective way possible. Having this in mind we came up with the design of one experiment that will allow us to combine the results of one library with the randomization of the other in a bidirectional mode.

The following strategy was proposed (Figure 14.5). The strings designed for the libraries have a series of characteristic elements: The 5' end of each strand is defined by T7 promoter sequences. This will flank the full gene including restriction enzyme target sequences in case they were needed, due to further modifications in the procedure. There is a defined sequence in the center of the gene, selected as reference for the combination of libraries. All libraries will be prepared by CPEC cloning presented in Figure 12.16.

To increase the selective pressure towards new deacetylation patterns, a disulfide bond evaluated in previous sections (see section 13.2.) was introduced in the gene library. The main objective of S-S bond is block the movement of loop 5 and forcing the substrate to slide and accommodate in the potential new negative subsites.

[3] Library of synthetic genes is equimolar for each codon.

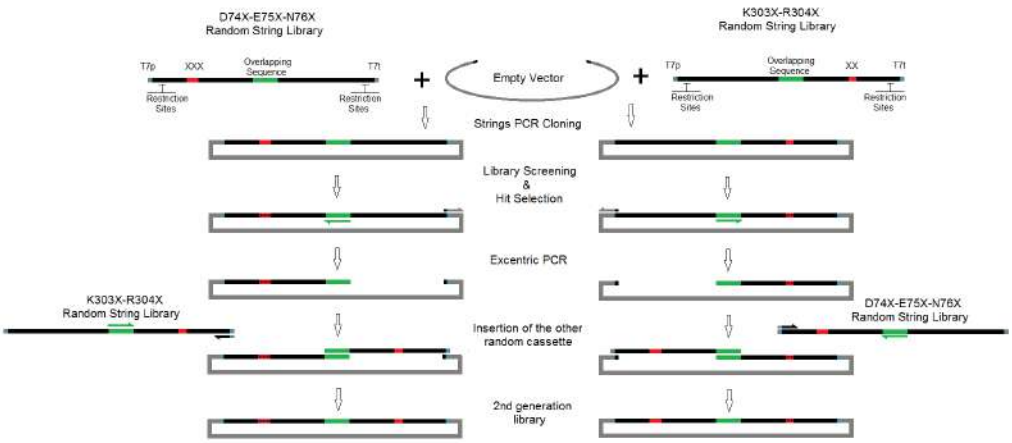


Figure 14.5. Schematic diagram of ISM strategy for directed evolution approach containing four sites. Site I: D74X-E75X-N76X, Site II: N119X, Site III: E300X, Site IV: K303X-R304X.

Different strategies were previously proposed in a way to reduce the library size of mutant libraries. Library size is an often used but ill-defined term. Generally, the term is used in a way that is synonymous with the number of transformants, unless the number of transformants greatly exceeds the maximum degeneracy of the library. Moreover, the degeneracy D of a library is the number of different individual variants among the transformants (i.e. the number of independent clones). The degeneracy of a library depends on the number of transformants T, the probability of occurrence of each specific sequence in the library and the maximum degeneracy D_{max} that could possibly be generated given the method used to create the library (i.e. the number of different members in a library of an infinite number of transformants). If all the variants are equally, the number of occurrences can be assigned by Poison distribution and degeneracy can be calculated by the following equation (Equation 14.1.):

$$D = D_{max}\left(1 - e^{-T/D_{max}}\right)$$

Equation 14.1. Degeneracy of a library. D_{max}: maximum degeneracy of a library (assuming infinite number of transformants), T: number of transformants.

Degeneracy define the number of variants in a library. Furthermore, a fundamental step of directed evolution for the creation of a library of variants of a starting protein(s) is how many variants there are for a particular protein.

$$V_H = 19^M \left[\frac{L!}{(L-M)!\,M!}\right]$$

Equation 14.2. Number of variants V_H of a protein with aminoacids length L. L: number of amino acids in a protein sequence. M: aa exchanges per molecule.

In Equation 14.2., it can be observed how quickly the potential library size grows and how the likelihood that a complete library can be created diminishes. Thus, given the limitations of the number of transformants on library size, for most proteins libraries the construction of a complete library of all variants with M = 3 is not feasible.

On the other hand, one of the key features is to define the probability of library completeness. This is found by taking the product of the probabilities that each particular library member is present in the library.

$$P_c = \prod_{i=1}^{D} P_i$$

Equation 14.3. Product of probabilities of each library is present in the library. P_i: probability that a particular sequence I is in the library.

If the probability of occurrence of each member of the library is equal, then Equation 14.3. reduces to Equation 14.4.

$$P_c = [1 - (1 - F)^T]^{D_{max}}$$

Equation 14.4. Probability of occurrence of each variant of a library. F: frequency that a particular sequence i is present in a library. D_{max}: maximum degeneracy of a library (assuming infinite number of transformants),

Finally, rearranging the Equation 14.4 to Equation 14.5 permits us to define the number of transformants needed in order to have a certain probability that the library is complete.

$$T = \frac{\ln(1 - P_c^{\frac{1}{D}})}{\ln(1 - F)} \approx \frac{\ln(1 - P_c^{\frac{1}{D}})}{-F}$$

Equation 14.5. Probability of occurrence of each variant of a library. F: frequency that a particular sequence i is present in a library. P_c: probability that the library is complete. D: degeneracy of a library.

For typical degeneracies (10^3 to 10^7), the number of transformants must exceed the degeneracy by a factor on the order of 10–25 in order to have a ≥99% probability of having a complete library.

On the basis of previous statistical analyses and using the CASTER program [228], the number of clones that should be screened for 95% coverage in the case of randomization at two and three positions, respectively (are given in Table 14.3.). Of course, full coverage is not

necessarily mandatory, and in fact in previous studies published so far it was not sought. Nevertheless, the higher the coverage, the greater the probability of finding improved variants. For instance, one position degenerated with NNK codon requires the screening of 96 colonies to be 95 percent confident that the screening contains all the possible variants, and three positions degenerated require 98000 clones.

Table 14.3. Statistical analysis of codon usage. [a]Number of clones to be screened for 95% coverage (over-sampling) when two or three amino acid positions at a given site are randomized using a specific degenerate codon. N: A/C/G/T; D: A/G/T; T ; B; not A; K:T/G; D; not G; R; A/G. Adapted from [227].

Degenerate codon	No. of codons	No. of Amino acids	No. of Stops	Amino acids encoded	95% coverage (2 positions)[a]	95% coverage (3 positions)[a]
NNK	32	20	1	All 20	3066	98163
NDT	12	12	0	RNDCGHILFSYV	430	5175
DBK	18	12	0	ARCGILMFSTWV	969	17470
NRT	8	8	0	RNDCGHSY	190	1532

In our library, two (K303X-R304X) and three (D74X-E75X-N76X) positions were randomized using NNK codon requiring a huge screening effort which demand a straightforward and reliable HTS approach for library screening.

14.2. Development and validation of a rapid screening assay for deacetylase activity.

Ideally, a directed evolution approach requires a massive clone screening (HTS) method to select candidate variants with the desired function or property. Here, we are looking for new mutants capable of generating new deacetylation patterns in chitooligosaccharides such as DP4 or DP5. As previously reported, mutants from rational design in positions Δ73-77, N119A and K303A-R304A drop substantially the activity towards longer substrates such as DP4, yielding an almost inactive enzyme (Table 14.2.) highlighting the greater importance of the interactions defined in negative subsite positions (Figure 14.3.). Given the importance of these amino acids in the catalytic activity besides the extensive libraries to be screen (Table 14.3.), a live/death primary screening has been proposed as a reliable and fast assay to discard the inactive mutants and select the active ones. This fast screening for large libraries is based on

the activity using a promiscuous substrate like methylumbelliferyl acetate (AcOMU). If hydrolyzed by esterase enzymes, the fluorogenic product 4-methylumbelliferone (MU) would be released and could be detected by measuring fluorescence increase at 340/460 nm along with reaction time (see protocol 21.2). Once functional and active mutants are selected, the secondary screening was designed to detect mutants with a potential new specificity. Previously selected mutants were repicked and expressed in a new microplate. The screening assay presented in chapter 1 with few modifications is proposed as a secondary screening assay. The adapted secondary assay is based on capture of the previously selected mutants with chitin-coated magnetic beads and evaluate the deacetylase activity of the immobilized enzymes variants on long substrates such DP4 by HPLC-MS (see section 11.2) looking for new degrees of acetylation. The paCOS product from the previous reaction are isolated and will be subjected to reaction with purified *Vc*CDA wt to identify new patterns of acetylation that will define a change in enzyme specificity (Figure 14.6.). (see protocol 21.2).

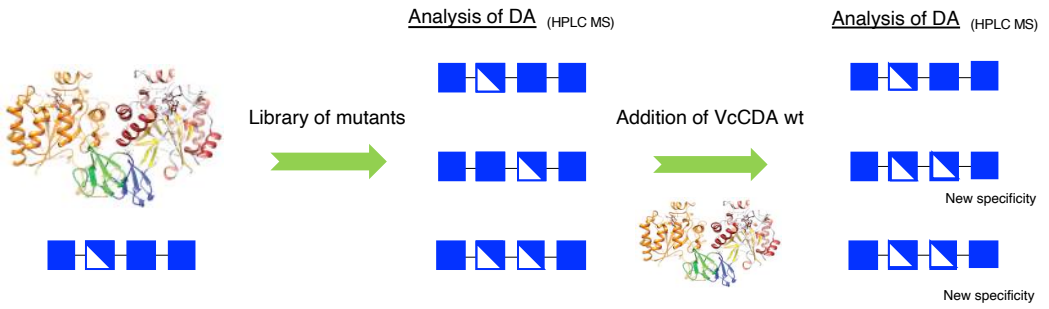


Figure 14.6. Scheme of secondary screening adapted from protocol presented in section 12.2.6.

To develop the primary screening assay, active *Vc*CDA P193C-N273C (section 13.2.) enzyme that contains the disulfide bond and inactive *Vc*CDA D39S enzyme were used as a proof-of-concept. The expression set-up in deepwell microplates (2 mL) was first optimized with the wild-type enzyme. The best condition were auto-induction for 48h at 24ºC with 300 rpm, followed by cell lysis using lysozyme (0.5 mg/mL) combined with freeze/thaw cycles to release the intracellular enzyme[4]. Clarified cell extract containing the enzyme were mixed with reaction buffer and methylumbelliferyl acetate substrate. Reaction plate was incubated at

[4] BugBuster lysis buffer was discarded due to economic reasons.

37ºC for 5 minutes and fluorescence was determined. These conditions were used in the deacetylase primary screening assay. Detailed protocol in section 21.2.

The primary screening assay was validated in a 96-microplate randomized with an inactive (*Vc*CDA D39S) and an active (*Vc*CDA P193C-N273C) enzymes as control references (Figure 14.7). The quality of the method in HTS assays is expressed by Z-factors defined as in section 12.2.6 and Equation 12.1.

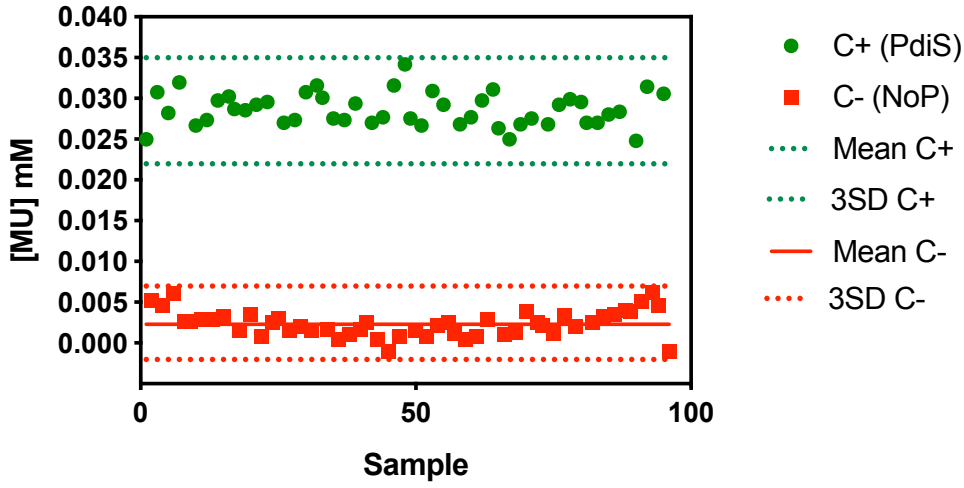


Figure 14.7. Z-factor calculation for adapted secondary screening assay. Equation 12.1. C+: Positive control (VcCDA P193C-N273C), C-: Negative control (empty plasmid, pET22b plasmid without gene cloned), 3SD: 3 times standard deviation.

The final method is able to discriminate between positive (active wt enzyme with disulfide bond) and negative (*Vc*CDA D39S) levels with a calculated Z'-factor of 0.6, which reflects a good quality for an HTS assay (Figure 14.7.). This complies with a reliable method for the screening of deacetylase activities implemented in massive HTS format.

Secondary screening assay was developed based on the HTS assay presented in Chapter 2. Optimization and validation of HTS was previously done (section 12.2.) but in our secondary screening assay some changes in the final protocol are implemented and need to be validated. Taken altogether, enzyme *Vc*CDA P193C-N273C was selected as a control reference, this double mutant in free form shows low activity in long substrates (DP4) but the same activity on short substrates (DP2) in comparation with wt enzyme due to the block of the movement of loop 5 (Figure 13.7.). Therefore, an inactive (*Vc*CDA D39S) enzyme and active enzymes (*Vc*CDA P193C-N273C and *Vc*CDA wt) were subjected to the full secondary

screening assay using DP2 and DP4 as a substrates to determine if the immobilized protein in CMB affects the overall activity of the control references.

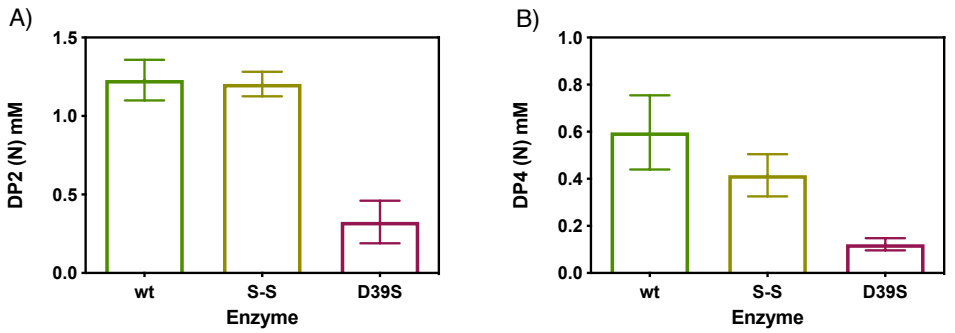


Figure 14.8. Comparation of secondary screening assay for VcCDA wt, VcCDA P193C-N273C and VcCDA D39S using DP2 (A) and DP4 (B) substrates.

As expected both the *Vc*CDA wt and *Vc*CDA P193C-N273C enzymes show the same activity using DP2 as a substrate, unlike long substrates such as DP4, where the activity is slightly lower in the case of the *Vc*CDA P193C-N273C enzyme (Figure 14.8.). The inactive mutant *Vc*CDA D39S shows almost no activity for both substrates and the activity obtained is due to the artifact of the screening itself since the values obtained are similar to the control reference of the empty plasmid (plasmid pET22b). The activity of the *Vc*CDA P193C-N273C enzyme in its immobilized state is higher than in its free form (Figure 14.8 and Figure 13.7); This unexpected increase in activity in comparation with free enzyme can be attributed to the partial reduction of the disulfide bond that allows the enzyme to behave as the wt enzyme, thus increasing its activity with long substrates.

The directed formation of S–S bonds in the diluted solutions significantly depends on structural peculiarities of a protein, in particular, on the nature and the number of amino acid residues between Cys residues, the close amino acid interactions and the protein environment. In order to evaluate the reduction or the oxidation of the disulfide bond in the secondary screening assay protocol, addition of reducing and oxidating agents were studied to determine how they affect enzyme activity. Therefore, full secondary screening assay has been performed in presence of 10 mM DTT as a reducing agent on one hand, and with 10 mM of H_2O_2 as a oxidant in the other hand. The reducing and oxidating agents were added only to reaction buffer and the enzyme was preincubated for 10 min prior enzymatic reaction to adapt to the new buffer. DTT cannot reduce buried (solvent-inaccessible) disulfide bonds.

However, disulfide bond C193-C273 is on the accessible surface of the protein. Finally, deacetylase activity was determined by fluorescence using fluorescamine (Figure 14.9.).

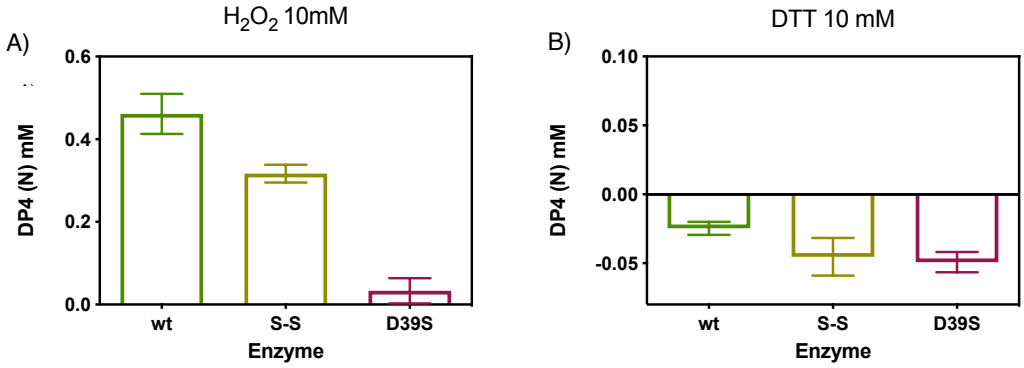


Figure 14.9. Comparation of secondary screening assay for VcCDA wt, VcCDA P193C-N273C and VcCDA D39S using DP4 as a substrate in presence of (A) oxidation agent (10 mM H_2O_2) and (B) reducing agent (10 mM DTT).

The presence of different concentrations of H_2O_2 slightly reduces the activity of all mutants compared with Figure 14.8 (Figure 14.9.A). The presence of some oxidating agent in final protocol was discarded. The reduction of disulfide bonds in the presence of DTT completely abolish the activity of all evaluated enzymes, concluding that the native disulfide bonds are essential for protein stability (Figure 14.9.B). Final protocol is presented in section 21.2.

14.3. Library screening of *Vc*CDA FL D74X-E75X-N76X.

The process of ISM was described in previous section 14.1. The first step of this method is to pinpoint several sites that are crucial for enhancing the desired property based on the 3D structure data. Considering our model in which four sites (A, B, C, D) have been identified as crucial each of these four sites is randomized by saturation mutagenesis to form four different libraries of mutant enzymes that are subsequently screened. Looking at the 3D structure in complex with DP3 in detail (PDB: 4OUI), positions D74X-E75X-N76X located in loop 1 were selected as first library to screen (Figure 14.2.).

Due to the importance of positions D74X-E75X-N76X, this library was the first to be analysed by the full screening assay. The first step to commence the directed evolution approach is to clone the string D74X-E75X-N76X with N273C-P193C mutations inside a vector, in our case is a pET22b using the CPEC technology strategy (Figure 12.16). BL21 (DE3) star cells were transformed using electroporation approach (see protocol 21.10.2). Cell picking was done using GENETIX picker (in collaboration with Dr. Sophie Bozonnet in INSA Toulouse). The

robot is based on the existing QPix and MegaPix colony picking systems (Genetix Ltd). These robots have XYZ drives that position the actuator head within the working area. The head is equipped with a camera to image colonies growing on agar. The colonies are imaged and assigned an XY coordinate, which enables them to be picked using a 'picking head' with 96 individually addressable pins. In order to develop a system for picking based on *E.coli* system, the imaging system was redesigned and equipped with a low noise, high dynamic range CCD camera. We have used 22 x 22 cm Genetix PerForma positively charged nylon membrane laid on QTrays (Genetix, Ltd.) filled with LB agar medium. Colonies are spread homogeneously on top of the membrane by applying a cell suspension after transformation and rolling 6 mm glass beads (Genetix) around the surface to spread the liquid (see protocol 21.2).

Then, 18400 clones were picked using a GENETIX picker platform following the proposed plate distribution (Figure 14.10). As already stated, using CASTER program and previous statistical analysis, picking 18400 transformants covers around 50% of all possible library combinations that is enough to evaluate the role of defined positions. Each 96-wellplate contains 2 positive controls (*Vc*CDA wt) and 2 negative controls (NoP: Empty plasmid). The rest of the plate is the library of mutants (Figure 14.10.). Different parameters were defined for cell picking: diameter: min:8 mm and max:25 mm, roundness: 0,7, Axis radio: 0,7 and proximity of 1. 2000 colonies were picked per hour.

	1	2	3	4	5	6	7	8	9	10	11	12
A												
B												
C												
D					PdiS	NoP						
E												
F												
G												
H											PdiS	NoP

Figure 14.10. 96-well microplate distribution for picking cells.

After picking, 96-well microplates were grown at 37ºC and 200 rpm for 24 h with 20%[5] of glycerol. All 96-wellplates were stored in -80ºC freezer to be ready for the screening assay. Screening of 200 plates (18400 clones) were done using methyl umbelliferyl acetate and the enzymatic activity was quantified following the primary screening assay. The average signal

[5] 20% glycerol was added directly to the LB growth media to avoid an extra growth step.

of negative controls of each plate was subtracted from the signal at each value of the plate. Representation of all 200 microplates was done in the following graph (Figure 14.11.).

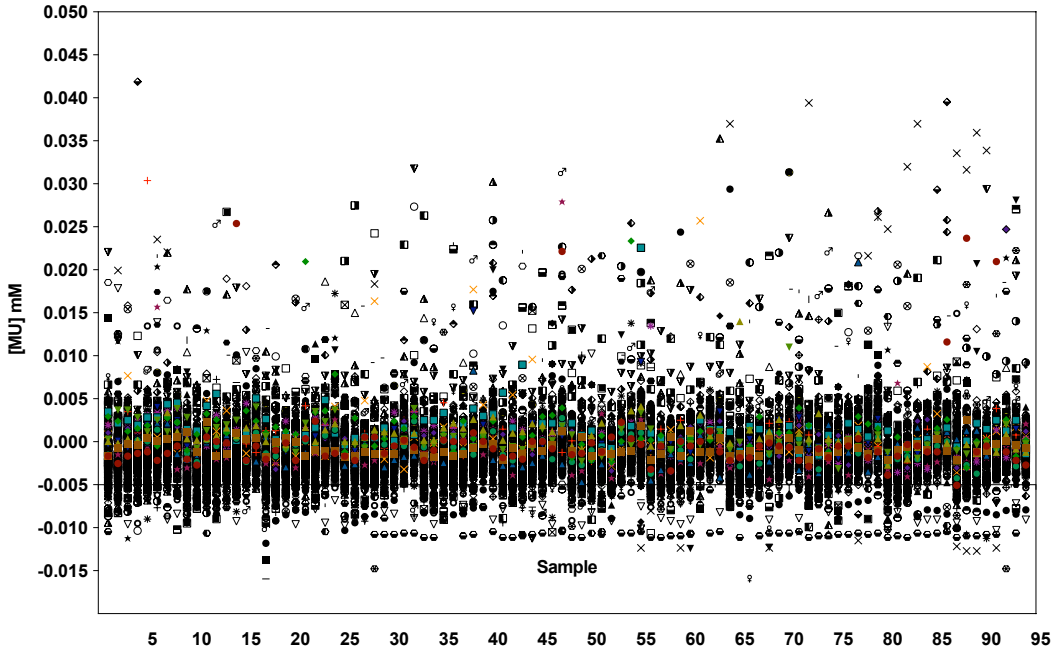


Figure 14.11. Screening of 18400 clones of the *Vc*CDA D74X-E75X-N76X library of mutants. X-axes contains all the mutants from each plate sorted by rows (A1, B1, C1, D1..). Each symbol and colour belongs to the same plate.

The distribution of screening results was homogeneous in all plates. Positive control was introduced in different positions of the plate from the sides to the center. From all the plates, we are able to obtain the mean of positive (S-S enzyme) and negative controls (empty plasmid) (Figure 14.12).

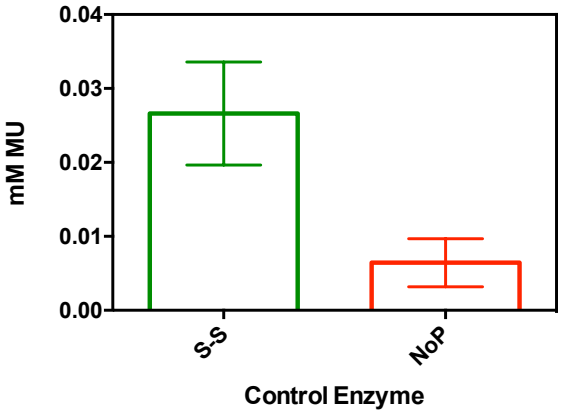


Figure 14.12. Mean of the positive controls (VcCDA P193C-N273C: S-S) and negative controls (plasmid empty: noP) of all plates screened.

From the mean of negative controls and the standard deviation (Figure 14.12.), we were able to decide the threshold to select the positive mutants from the first screening assay. 3 times the standard deviation above the negative control mean was selected as a threshold considering that almost all negative values will fall within three standard deviations. The value was 0.008 mM of MU, values lower than 0.008 mM of MU are negative mutants and values higher than 0.008 mM are considered positive hits. Following this criteria, 275 clones from 18.400 total clones were selected, this represents a 1.5% of total library (Figure 14.13).

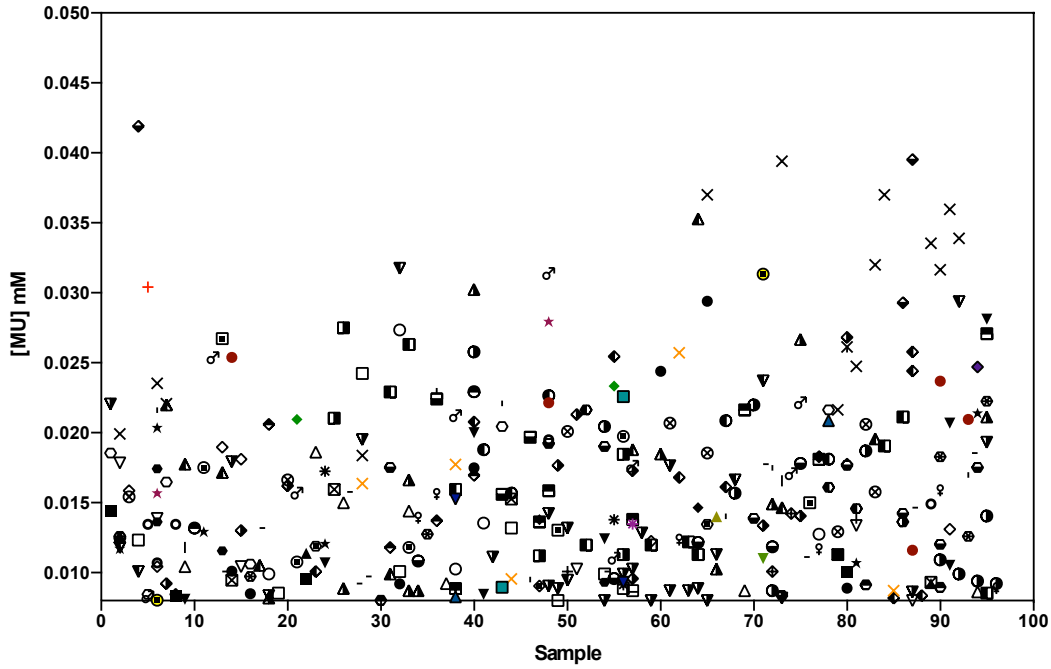


Figure 14.13. 275 clones selected as a positives hits from 18400 clones screened.

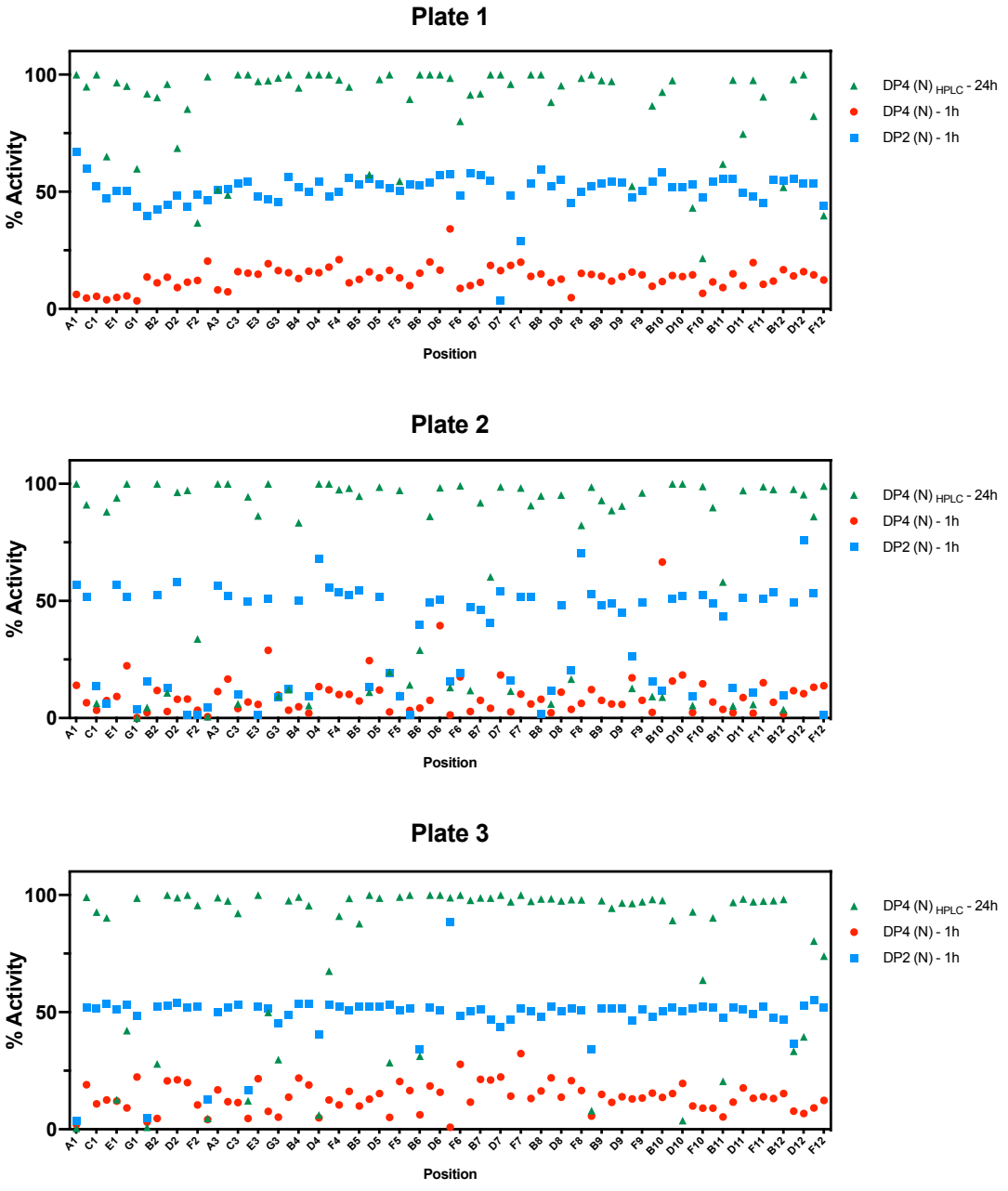
Afterwards, four 96-well microplate were filled with the positive hits. These plates were obtained repicking the cells by hand on a new 96-well microplates. Glycerinates were prepared growing cells in 1.5 mL of LB amp ON at 37ºC and 250 rpm (layout distribution shown in Figure 14.14 as example). Then, the cultures were centrifuged at 4000 rpm, 20 min at 4ºC and resuspended with 200 μl of LB Amp (20% glycerol) and stored at -80ºC freezer.

	1	2	3	4	5	6	7	8	9	10	11	12
A	P001 (F2)	P001 (H6)	P001 (G11)	P001 (B12)	P001 (E12)	P003 (B9)	P004 (G9)	P005 (E3)	P005 (G7)	P007 (H7)	P008 (F10)	P009 (F5)
B	P010 (F12)	P011 (A8)	P012 (F1)	P012 (H6)	P013 (E1)	P014 (D4)	P014 (F5)	P014 (F8)	P015 (G9)	P016 (A7)	P017 (C1)	P018 (A4)
C	P018 (D6)	P020 (C10)	P021 (B5)	P021 (D5)	P022 (H8)	P023 (A11)	P024 (A12)	P026 (B5)	P026 (H9)	P026 (C10)	P027 (H6)	P028 (D9)
D	P028 (B12)	P029 (A8)	P029 (E9)	P030 (D5)	P030 (C6)	P030 (F6)	P031 (G4)	P031 (G8)	P032 (A4)	P032 (B4)	P033 (G1)	P034 (H5)
E	P034 (C10)	P034 (D8)	P035 (H4)	P035 (D12)	P036 (A1)	P036 (D4)	P036 (E8)	P037 (D5)	P037 (H5)	P038 (F11)	P040 (A8)	P045 (F4)
F	P046 (F7)	P046 (B12)	P047 (A11)	P047 (F10)	P050 (H9)	P050 (E12)	P050 (F6)	P052 (G9)	P052 (G11)	P054 (F2)	P054 (H5)	P054 (D8)
E	P054 (A9)	P057 (F7)	P057 (G12)	C (+) (WT)			C (-) (PdiS)			C (-) (D39S)		
G	STANDARDS											

Figure 14.14. Example of a distribution for secondary screening assay. The colour of each well indicates the plate of source of each mutant (for example, mutants selected from P001 have a grey colour). Dark green (E4 to E6): positives control (*Vc*CDA wt), Pale green (E7 to E9): positives control (*Vc*CDA P193C/N273C: PdiS), red (E10 to E12): negative control (*Vc*CDA D39S).

Secondary screening was done using DP2 and DP4 (2 mM each) as substrates. The final protocol is presented in section 21.2.2. Deacetylase reaction was performed adding 70 μl of substrate to the E-CMB complex. After 1h reaction at 37ºC, 20 μl were withdrawn to determine the deacetylase activity using fluorescamine. Then, the volume reaction was kept at 37ºC for

24h. Afterwards, 10 μl were withdrawn and analyzed by HPLC-MS to check the degree of acetylation, and another aliquot of 10 μl was mixed with 40 μl of 20 μM *Vc*CDA wt enzyme and let it react for 24h at 37ºC. After 24h of reaction, 10 μl were analyzed by HPLC-MS to check the degree of acetylation. Four microplates previously prepared were analyzed using this secondary screening assay (Figure 14.15). % of deacetylated DP4 was determined taking into account the initial substrate concentration (2mM).



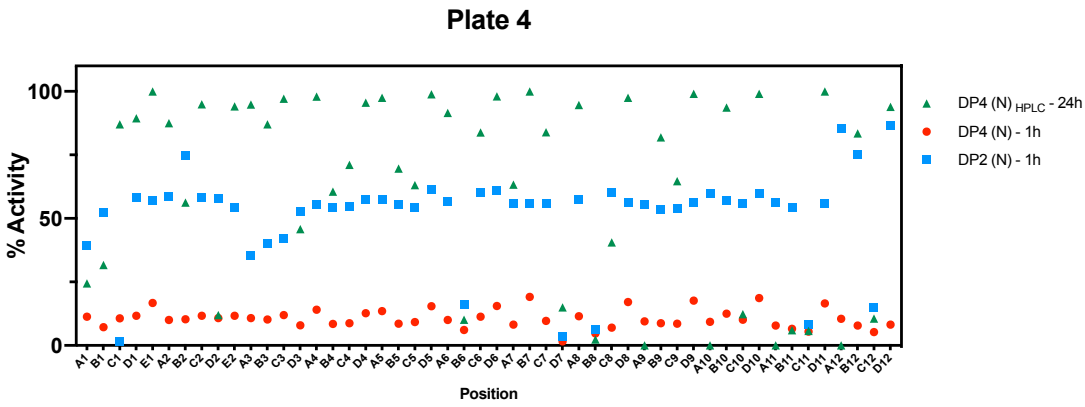


Figure 14.15. Results of secondary screening assay for 4 plates containing the positives hits from primary screening assay. % of DP4(N) regarding the total amount of DP4 (2mM). Green triangle: Activity with DP4 using HPLC-MS after 24h. Red circle: DP4(N) detected using fluorescamine at 1 hour reaction. Blue square: DP2(N) detected using fluorescamine at 1 hour reaction.

Mutants with activity lower than 80% (% of DP4 (N) regarding the total amount of DP4) were inoculated in a new microplate for DNA sequencing. Sequencing of all potential variants in 96-well microplate was done using kit NZYMiniprep 96 well plate from NZYTech. The kit is designed for the manual or automated rapid, smallscale preparation of highly pure plasmid DNA from recombinant *E.coli* strains.

We expect a change in specificity to be accompanied by low enzyme activity. This criteria permits us to select the further potential candidates since all mutants with activity greater than 80% are considered wt enzyme. On the other hand, among all the mutants analyzed, one mutant showed a change in specificity since double deacetylation appears in HPLC-MS after 24h reaction previously to the addition of *Vc*CDA wt. The full plate was sequenced using T7 promoter primer. Different mutants were identified. 16% of the sequences were wrong sequences and were unable to align with template wt plasmid. 36% of the sequences are wt enzyme. Selected sequences are shown in the following Table 14.4.

Table 14.4. Best sequences selected from sequencing promising mutants obtained from secondary screening assay. (wt seq: D74-E75-N76). % DP4(N) HPLC activity was previously shown in Figure 14.15.

Position	DNA Seq	74	75	76	% DP4(N) HPLC
E1	GGGGAGATG	G	E	M	28
A3	GATGAGACG	D	E	T	70
F3	GGTGAGGCT	G	E	A	39
A4	GATGAGTCG	D	E	S	85
E4	GATGAGGTT	D	E	V	50
A5	TGTGGGCGT	A	E	A	37
A6	TCGGTTCCG	S	V	P	51
E8	GGGGAGCGG	G	E	R	30
F12	GATGAGCGG	D	E	R	71
H7	GGTGAGCAT	G	E	H	100
F7	GGTGAGCAT	G	E	H	32
H1	GCGGAGTTG	A	E	L	12
C7	CATGCTGGG	H	A	G	14
E3	CGTGCGAGG	R	A	R	12

14 mutants were selected as the most promising ones based on the sequence and activity using DP4 as a substrate Figure 14.16.

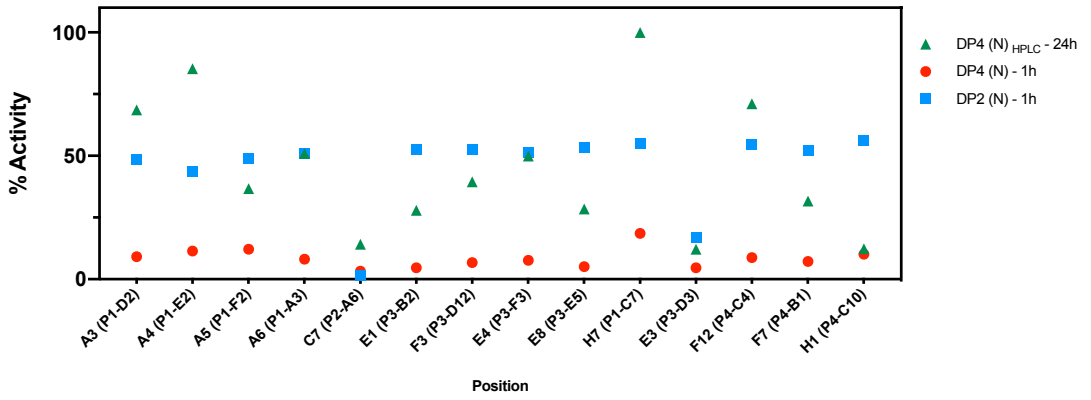


Figure 14.16. Results of secondary screening assay for 14 mutants selected as most promising ones.

A more in-depth analysis of the sequencing and the enzymatic activity results was done; mutation of E75 is almost no observed in any variant, highlighting the importance of this amino acid and its interactions. Amino acid E75 interacts with amino acid R304 with a bidentate hydrogen bond between both amino acids holding loop 1 strongly anchored to loop 6 (Figure

14.3). Presumably, the disruption of this interaction could destabilize all the loop structure. In a few variants, A6 (SVP), C7 (HAG) and E3 (RAR); amino acid 75 was mutated, but activity has strongly affected (Figure 14.16.) and the mutants have not shown any change in the PA and are thereafter discarded.

On the other hand, mutations in positions 74 and 76 have been more frequently identified. With regard to position N76, most of the mutations founded such as valine, alanine and proline completely break the interaction with amino acid E300 (Figure 14.3.). Mutation of N76 almost has no effect on enzyme activity because Loop 1 is still anchored to Loop 2 by E75-R304 bidentate interaction in most of the variants.

On the other hand, it was proposed that loop 2 acts as a platform for the substrate and does not show dynamics when substrate binding takes places. One of the most abundant mutation in 74 position is glycine, this mutation breaks the interaction of loop 1 with loop 2 at expenses of the reduction in activity using DP4 (Figure 14.16). Disruption of L1-L2 interaction by mutation of position 74 has not large impact on enzymatic activity compared to previous rational design where mutation N119A was evaluated and the enzyme did not show any activity, this could be due to new interactions were formed that stabilize the mutant.

After selection of most promising variants from the full screening assay, 11 mutants and wt enzyme were expressed in small culture (50 mL) and were purified using StrepTag spin columns (IBA life science). As example, SDS-PAGE was done to analyze protein purification using spin columns.

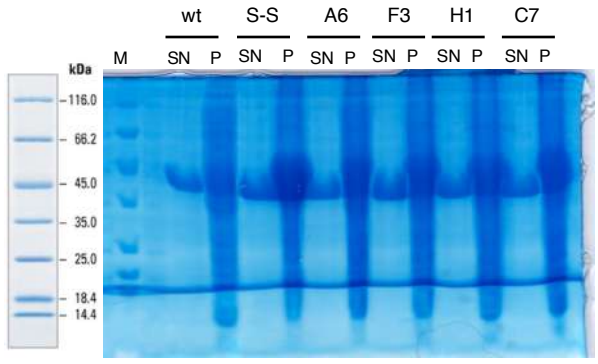


Figure 14.17. SDS-PAGE 14% of acrylamide of purification using Streptag spin column. M) molecular weight marker III. Wt , P193C-N273C (S-S) and different mutants were analyzed.

Soluble protein was recovered after elution of Streptag spin columns. Proteins purified from spin columns are pure enough for further assays. Study of the specificity of each variant has

been done. The protocol is based on the primary detection of the DA using HPLC-MS for all the variants. Afterwards, *Vc*CDA wt was added to the mixture and the deacetylation of second position from the non-reducing end was done (see section 12.2.7.2). Only the mutant D74G-E75E-N76A has shown a slightly change in deacetylation pattern. This mutant renders a small proportion of double deacetylated product (DP4(2N)) after 24 h reaction.

Table 14.5. List of DA (-wt: nefore addition VcCDA wt, +wt: after addition VcCDA wt) and PA analysis of all 11 purified mutants.

	DA		
Position	**- wt**	**+ wt**	**PA**
E1	Mono-	Mono-	No change
A3	Mono-	Mono-	No change
F3	Di-	Di-	New Specificity
A4	Mono-	Mono-	No change
E4	Mono-	Mono-	No change
A5	Mono-	Mono-	No change
E8	Mono-	Mono-	No change
F12	Mono-	Mono-	No change
H7	Mono-	Mono-	No change
F7	Mono-	Mono-	No change
H1	Mono-	Mono-	No change

Taken altogether, there is no mutant that shows a change in substrate specificity in terms of deacetylation pattern. The hydrogen network between loops 1, 2 and 6 is key for protein stability and only a slightly change was unable to generate new potential negative subsites. Results obtained suggest that randomization of amino acid pair interaction (such E75-R304) could be alternative approach to address new specificities.

14.4. Closing remarks in engineering *Vc*CDA towards new deacetylation patterns

In this chapter a new screening procedure was develop to screen large libraries of chitin deacetylases. The screening is composed of two steps. First is based on the detection of active mutants using a promiscuous substrate methyl umbelliferyl acetate for chitin deacetylases. This permits us to select the active enzymes within a large library of mutants. The second step was applied to the mutants selected in the first step, and it is formed by two different steps. First part analyzes the degree of acetylation of COS by HPLC-MS and second determines if there is a change in pattern of acetylation by addition of *Vc*CDA wt enzyme.

With this new screening methodology we sought to determine not only mutants active within a large library, we can also determine the degree of acetylation and the pattern of acetylation of mutants active on COS with a fast and straightforward screening.

This new screening methodology was applied to screen a large library of *Vc*CDA mutants focused on negative subsites towards new deacetylation patterns. More than 18000 clones where analyzed, 200 mutants where selected for secondary screening and only one mutant shows a new specificity. The hydrogen-bonding network between loops 1, 2 and 6 is key for protein stability. From a combinatorial library at loop 1 residues, only mutant D74G-E75E-N76A showed the appearance of a new product with low yield corresponding to a di-deaceetylated COS. At this stage, it is concluded that such an engineering goal for different deacetylation specificity would require a major loops restructuring.

TOPIC 2: Engineering a septanoside hydrolase

INTRODUCTION

15. TOPIC 2: Engineering a septanoside hydrolase. INTRODUCTION. Glycosidases and glycosynthases.

Glycoside hydrolases (also called glycosidases or glycosyl hydrolases) catalyze the hydrolysis of glycosidic bonds in complex sugars. Glycoside hydrolases are classified into EC 3.2.1 as enzymes catalyzing the hydrolysis of O- or S-glycosides (see section 9.3). They are extremely common enzymes with different roles in nature including degradation of biomass such as cellulose (cellulases) or starch (amylases), in anti-bacterial defense strategies (lysozymes), in pathogenesis mechanisms (viral neuraminidases) and in normal cellular function (trimming mannosidases involved in N-linked glycoprotein biosynthesis) [229,230].

There are three main basic types of these glycosidase enzymes depending on the catalytic mechanism: those in which the anomeric carbon configuration of the product inverts its initial configuration, the so-called inverting glycoside hydrolases; and those in which it is maintained, known as retaining glycoside hydrolases and retaining glycosidases by substrate-assisted catalysis [231].

Inverting glycosidases operate by a single-step mechanism in which a water molecule (with general base catalysis) effects a direct displacement at the anomeric center with protonic assistance by the general acid residue on the departing glycosidic oxygen. The catalytic residues are located approximately 10 Å apart from each other, allowing for the binding of the substrate and a water molecule in a ternary productive complex. Bond formation takes place at the same time as bond breaking (Figure 15.1.A.). Retaining glycosidases reaction mechanism operates in two steps. In the first step, the general acid protonates the glycosidic oxygen, while the deprotonated nucleophile attacks the anomeric center leading to the glycosyl-enzyme intermediate. In the second step, a water molecule activated by the conjugated base of the general acid residue attacks the anomeric carbon leading to the hydrolysis product with net retention of configuration (Figure 15.1.B). Some retaining glycosidases catalyzing glycoside bond hydrolysis in 2-acetamido sugars (i.e. GH18 chitinases, GH20 hexosaminidases, GH52 hyaluronidases, GH85 endo-β-acetylglucosaminidases and other families such 18, 20, 25, 56, 84 and 85) follow a slight variation of the retaining mechanism; they lack the enzyme's catalytic (or neighboring group participation) nucleophile and operate by substrate-assisted catalysis, where the N-acetyl group of the substrate acts as an internal nucleophile, forming an oxazoline/oxazolinium intermediate, which is then attacked by a water molecule assisted by the conjugate base of

the general acid residue to yield the product with net retention of the anomeric configuration (Figure 15.1.C).

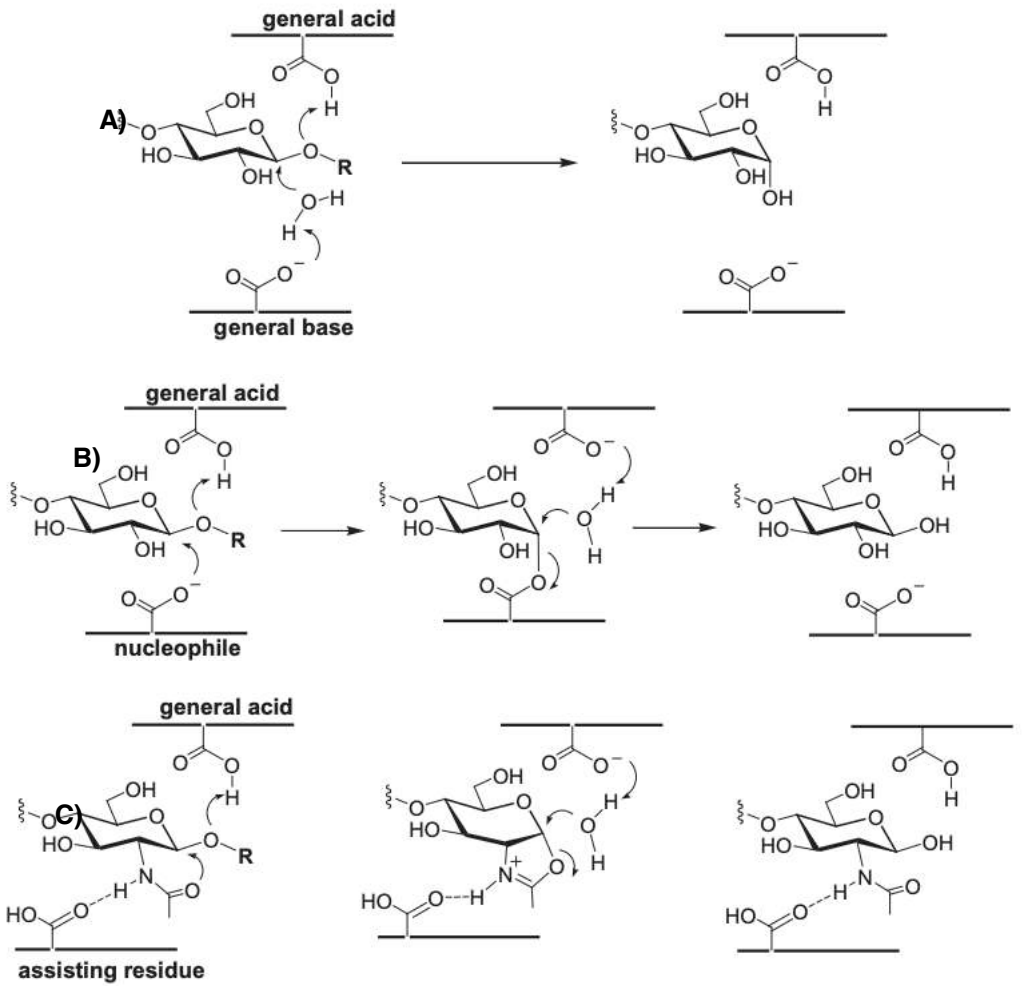


Figure 15.1. Glycosidase mechanism. A: Inverting glycosidase. B: Retaining glycosidase with enzyme nucleophile, via a glycosyl–enzyme intermediate in the two-step displacement mechanism. C: Retaining glycosidase by substrate-assisted catalysis, via an oxazolinium ion intermediate. Adapted from [232]

Furthermore, glycosynthases are glycosyl hydrolases modified by enzymatic engineering to develop an activity opposite to their natural activity. While natural retaining glycosidases activity is to hydrolase glycosidic bonds, reverse activity could be achieved by displacing the reaction equilibrium towards glycosidic bond formation (thermodynamically controlled condensation) or by using activated glycosyl donor as substrate (kinetically controlled transglycosidation). Therefore, these enzymes do not exist in nature. Its discovery occurred simultaneously in 1998 by two independent groups of researchers: Withers's group working

on an exo-glycosidase [233] and Planas's group , Laboratori de Bioquímica del Institut Químic de Sarrià, working with an endo-glycosidase [234]. Since its discovery, this technology has been applied to multiple enzymes of different nature. Glycosynthases are configuration-retained glycosyl hydrolases in which the nucleophilic residue has been replaced by an inert one. This makes the hydrolytic reaction impossible, but they are able to significantly catalyze the formation of glycosidic bonds from activated substrates. The term glycosynthase has been incorrectly applied to other types of glycosidases such as inverting the first [235] or β-*N*-acetylhexosaminidases [80]. These new enzymes with the ability to form glycosidic bonds could be called “glycosynthase like” since, generally, in these the hydrolytic activity of the original enzyme has not been completely eliminated, something that happens with the original glycosynthases.

15.1.1. Catalytic mechanism of glycosynthases

The catalytic mechanism of glycosidases has been extensively studied. Glycosidases with retention of configuration, in addition to catalyzing the hydrolysis reaction of the glycosidic bond, can catalyze their formation in their wild type version. Depending on the substrate used, there are two basic pathways through which the WT enzyme can act. One is thermodynamic control in which non-activated substrates and transglycosylation products are in equilibrium (Figure 15.2). Another more efficient route has been called kinetic control transglycosylation, this approach uses activated substrates, such as glycosyl fluorides, arylglycosides, etc. By having a better leaving group, the intermediary is formed in a faster way so that it accumulates. In this way, a faster capture of the activated glycosyl-enzyme intermediate by the glycosidic acceptor is sought than by the water molecule. However, these two methods do not provide a great yield to the reaction, since the products of this last mechanism are also capable of being hydrolyzed (Figure 15.2). By combining these strategies with the engineering of the reaction media, yields of 50% have been reached [236].

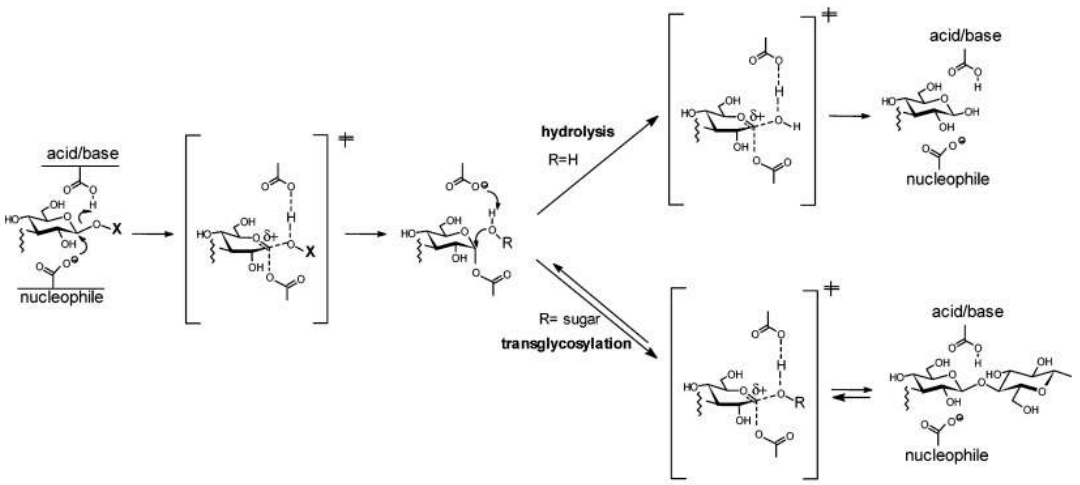


Figure 15.2. Mechanism of double displacement of a glycosidase with retention of configuration, in the lower branch of the scheme, the transglycosylation reaction with a glycosidic acceptor, is shown. Adapted from

In glycosynthases, the nucleophilic residue has been replaced by one completely inert, thus preventing from the hydrolytic reaction and avoiding the formation of glycosyl-enzyme intermediate (Figure 15.3). In this case, the substrate is a glycosyl fluoride with an anomeric carbon configuration opposite to the natural substrate (for example, a α-glycosyl fluoride for a β-glucosidase). This substrate configuration mimics the configuration that the substrate acquires when forming the glycosyl-enzyme intermediate. By eliminating the residue that acts as a nucleophile (generally glutamate) and replacing it with a less bulky one (such as serine, alanine or glycine for example) a larger cavity is created that allows the new substrate to enter the active site without hindrance [237]. The yields of glycosynthase-mediated transglycosylation are very high since the reaction product is no longer hydrolyzed by the enzyme. The IQS Biochemistry Laboratory has developed this technology in the last 20 years. Thus, it has been possible to apply various enzymes for use as biocatalysts [87].

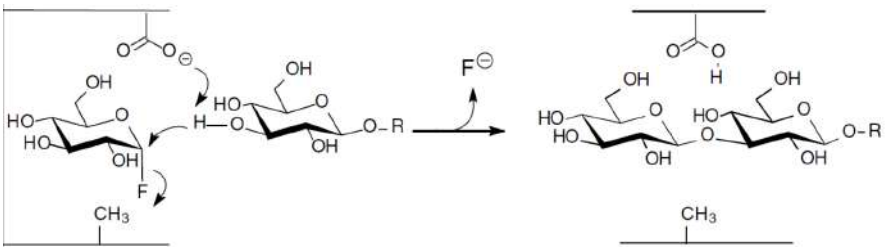


Figure 15.3. Mechanism of transglycosylation of a glycosynthase by mutation of the nucleophilic residue.

15.2. β-glycosidases

β-glycosidases are specific glycosidases that hydrolyze β-glycosidic bonds of oligosaccharides releasing nonreducing terminal glycosyl residues. Among all glycosyl hydrolases, beta-glucosidases [E.C.3.2.1.21] are a group of well-characterized, biologically important enzymes that belong to different GH families based on their metabolism and signalling functions. These enzymes have plenty of useful catalytic and sequence characteristics family GH1, which includes beta-glucosidases from bacteria, archaebacteria, plants and mammals; and family GH3, which comprises beta-glucosidases of bacterial and yeast origin.

15.2.1. *β*-glucosidase 3 from *Streptomyces sp.*

This project on protein engineering has as target the protein β-glucosidase from *Streptomyces sp.* (β-gl3). The β–gl3 β-glucosidase (β-glucoside glucohydrolase, EC3.2.1.21) from *Streptomyces sp.* has 479 amino acid sequence, 1437 bp nucleotides sequence and 52,3 kDa molecular weight. The β–glucosidase gene from *Streptomyces sp.* QM- B814 (ATCC 11238) was firstly discovered, cloned and sequenced in 1994 and classified in GH1 family. This enzyme is a retaining glycosidase showing an exo-like action pattern by releasing glucose units from the nonreducing end of different β-linked disaccharides and oligosaccharides. Like other retaining glycosidases, the βGl3 enzyme is able to perform transglycosylation reactions (i.e., transfer of glycosidic residues), catalyzing the formation of β-1,4 or β-1,3 bonds depending on the acceptor [238,239].

In recent years a vast amount of mechanistic information on the enzymatic action of retaining glycosidases has been produced through mutagenesis, enzyme kinetics, inhibition and X-ray crystallography studies (Figure 15.4. 3D structure of β-glucosidase from *Streptomyces sp.* PDB code: 1GON.). The protein was crystallized, and the 3D structure resolution of its native form and complexes with nonhydrolyzable substrate analogues was elucidated.

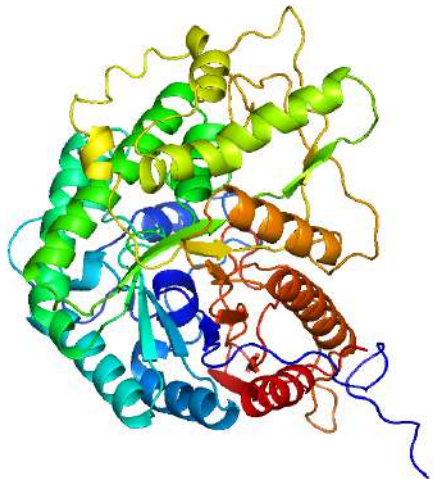


Figure 15.4. 3D structure of β-glucosidase from *Streptomyces sp.* PDB code: 1GON.

β–Glucosidase from *Streptomyces sp.* (β-gl3) has 479 amino acid sequence, 1437 bp nucleotides sequence and 52.3 kDa molecular weight. The β-glucosidase gene from *Streptomyces sp.* QM- B814 (ATCC 11238) was firstly discovered, cloned and sequenced in 1994 [239] and classified in GH1 family. β–gl3 enzyme show exo-like action pattern by releasing glucose units from the nonreducing end of different β-linked disaccharides and oligosaccharides. Among β-D-glycosides, the highest k_{cat}/K_M value is with glucoside (with substrate inhibition), fucoside, galactoside and xyloside [238]. The catalytic residues were determined, Glu178 as the general acid/base catalyst and substrates ray diffraction at 2.2 Å of resolution (PDB code: 1GON). Glu 383 as the catalytic nucleophile β-gl3 crystallographic structure was solved by X-ray diffraction at 2.2 Å of resolution (PDB code: 1GON) (Figure 15.4) [238].

Kinetics for the wild-type β-glucosidase were evaluated with different p-nitrophenyl β-glycosides as summarized in Table 15.1.

Table 15.1. Michaelis-Menten Parameters of *Streptomyces sp.* β-Glucosidases (wt and Mutants) with Cellobiose and p-Nitrophenyl β-Glycoside Substrates. Adapted from [238].

Mutant	Substrate	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M ($M^{-1} \cdot s^{-1}$)
wt	cellobiose	35.6	4.1	$8.6 \cdot 10^3$
	pNP-Glc	28.4	0.15	$1.9 \cdot 10^5$
	pNP-Fuc	37.1	0.14	$2.7 \cdot 10^5$
	pNP-Gal	118	7.3	$1.6 \cdot 10^4$
	pNP-Xyl	0.63	3.0	$2.1 \cdot 10^2$

The wt enzyme has a broad substrate specificity which is common among glycosyl hydrolases from family 1. Glucoside and fucoside substrates show the highest catalytic efficiency (in terms of k_{cat}/K_M), but a significant substrate inhibition is obtained for the p-nitrophenyl β-glucoside as opposed to the fucoside.

In previous works [240], it was explored the properties of the E383A mutant as a glycosynthase to define its activity and specificity for a wide combination of donors and acceptors. The E383A glycosynthase from *Streptomyces sp.* offers an unexpected regioselectivity compared to the other glycosynthases. For monosaccharide acceptors, the enzyme produces β-(1→3) disaccharides in excellent yields. Particularly, the disaccharides Glcβ3Man and Galβ3Man obtained have not been reported for any other glycosynthase. In contrast, β-(1→4) linkages are obtained with disaccharide acceptors. Therefore, interactions in subsite +2 of the enzyme seem to play a major role in directing regioselectivity.

15.3. Seven-member rings

Most of carbohydrates found in eukaryotic organisms such as glycolipids, structural polysaccharides, glycoproteins and other glycoconjugates, have the sugar components almost exclusive in furanose and pyranose ring forms, with a 5 or 6 carbon-rings. Adding another carbon atom to the pyranose ring homologates it to a septanose, with 7 carbon-ring.

Seven membered ring septanose sugars have the potential to be transformative glycomimetics with application in chemical biology and medicinal chemistry. A collaborative project between the Peczuh's group (University of Connecticut) and Planas' group (Laboratory of Biochemistry, IQS) addresses the development of septanoside chemistry and biochemistry and their biological applications. Our vision for the application of septanoses follows three broad scenarios. The first is analogous to beta-peptides, where septanoses compete with pyranoses as ligands of lectins, but are insensitive to degradation by hydrolytic enzymes. In a similar fashion, beta-peptides inhibit protein-protein interactions but are resistant to proteases. The previous Peczuh's project developing septanoses as ligands of the *E. coli* lectin FimH has forwarded this paradigm the farthest to date [241]. Lectin FimH is a protein which plays an essential role in the first steps of urinary tract infections. Septanose is able to bind lectin FimH protein, compromising its flexibility and enabling its activity. Furthermore, it has been shown that the protein jack bean lectin concanavalin A can selectively bind to beta-septanosides. However developing its truly potential requires

knowledge and tools to synthesize and use it as monomers for oligo- and polysaccharide formation

In the second scenario, septanoses become a “bump,” by virtue of the expanded ring, to the “hole” in an engineered hydrolase enzyme with septanoside hydrolase activity. The bioorthogonal, enzymatic unmasking of a functional aglycone will allow the highly selective and time-resolved delivery of probe compounds inside cells. Therefore our goal here is to engineer and develop a glycosidase able to accept septanoside substates that will become a new biorthogonal system with applications in chemical biology.

In the third scenario, once a successful septanoside hydrolase is achieved, its conversion to a glycosynthase will open the way to synthesis septanoside-containing oligo-saccharides and glycoconjugates as brand-new structures with broad applications in biotechnology.

In the Peczuh group from U.Conn, they have made progress toward understanding the chemistry and properties of septanoses, ring-expanded analogs of glucose. They developed syntheses of septanose-containing mono- and disaccharides, and showed that monoseptanosides take up stable, low-energy ground state conformations. In the previous work, they developed syntheses of chromogenic septanosyl glycosidase substrates (i.e., pNP Sept) (Figure 15.5) [242].

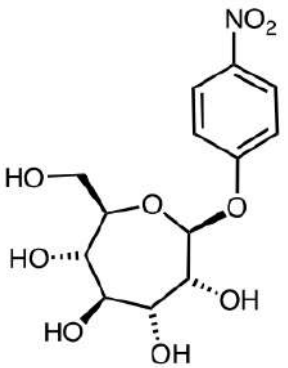


Figure 15.5. pNP-Septanoside molecule.

OBJECTIVES

16. OBJECTIVES

Glycosidases in the same family as βgl3, the enzyme to be engineered in this project, have proven amenable to modification by protein engineering. As already stated in previous sections, βgl3 enzyme was deeply characterized in previous studies. Valuable observations on the structure-based engineering of glycosidases to accept alternative substrates for hydrolysis have been reported. The objectives of the chapter are:

- The identification of a naturally occurring septanoside hydrolase or other GH sufficiently promiscuous to hydrolyze septanosides would be an important milestone. Our interest in seven-membered ring septanose sugars has inspired the development of a way to search for septanoside hydrolase activity. In first section of the results, it is described a strategy for the discovery of septanoside hydrolases that uses synthetic indolyl septanosides as chromogenic substrates. The screening strategy leverages the known dimerization of 3-hydroxy-indoles to make colored dyes, as occurs when the β-galactosidase substrate X-Gal is hydrolyzed. Among different glucosidase analyzed, only the Streptomyces sp. β-glucosidase (βgl3) resulted in faint activity against indolyl-septanoside.
- Afterwards, the second goal of this project is to engineer the βgl3 protein from *Streptomyces sp.* to obtain a mutant capable of accepting septanoside substrates with high hydrolytic activity. βgl3 protein from *Streptomyces sp*. was selected because the resulted week activity previously demonstrated and its capability to accept many different substrates.

The final goal of our group, once acquired a septanose active mutant, is to revert the hydrolase activity to glycosynthase activity. Consequently, a brand-new enzymatic tool (for transformation of seven-member ring substrates) to create oligosaccharides containing seven-member ring carbohydrates (septanose) will be achieved.

CHAPTER 5. INDOLYL SEPTANOSIDE SYNTHESIS FOR IN VIVO SCREENING OF BACTERIAL SEPTANOSIDE HYDROLASES

This chapter is presented as an article.:

Indolyl Septanoside Synthesis for In Vivo Screening of Bacterial Septanoside Hydrolases.
Aditya R. Pote, Sergi Pascual, Antoni Planas, and Mark W. Peczuh
Int. J. Mol. Sci. 2021, 22(9), 4497;
DOI: https://doi.org/10.3390/ijms22094497

17. CHAPTER 5. Indolyl Septanoside Synthesis for *In Vivo* Screening of Bacterial Septanoside Hydrolases

This chapter is presented as an article.

Int J Mol Sci. 2021 Apr 26;22(9):4497. doi: 10.3390/ijms22094497.

Indolyl Septanoside Synthesis for *In Vivo* Screening of Bacterial Septanoside Hydrolases

Aditya R. Pote[1†], Sergi Pascual[2†], Antoni Planas[2], and Mark W. Peczuh[1]

[1] **Department of Chemistry, University of Connecticut, 55 N. Eagleville Road U3060, Storrs, CT 06279, USA;**

[2] **Laboratory of Biochemistry, Institut Químic de Sarrià, University Ramon Llull, 08017 Barcelona, Spain;**

† These authors contributed equally to the work

Abstract

Building-up and breaking-down of carbohydrates are processes common to all forms of life. Glycoside hydrolases are a broad class of enzymes that play a central role in the cleavage of glycosidic bonds, which is fundamental to carbohydrate degradation. The large majority of substrates are five- and six-membered ring glycosides. Our interest in seven-membered ring septanose sugars has inspired the development of a way to search for septanoside hydrolase activity. Described here is a strategy for the discovery of septanoside hydrolases that uses synthetic indolyl septanosides as chromogenic substrates. Access to these tool compounds was enabled by a route where septanosyl halides act as glycosyl donors for the synthesis of the indolyl septanosides. The screening strategy leverages the known dimerization of 3-hydroxy-indoles to make colored dyes, as occurs when the β-galactosidase substrate X-Gal is hydrolyzed. Because screens in bacterial cells would enable searches in organisms that utilize heptoses or from metagenomics libraries, we also demonstrate that septanosides are capable of entering *E. coli* cells through the use of a BODIPY-labeled septanoside. The modularity of the indolyl septanoside synthesis should allow the screening of a variety of substrates that mimic natural structures via this general approach.

Keywords

septanoside; indolyl glycoside; glycosidase; septanoside hydrolase.

Introduction

Glycoside hydrolases (GH) are a broad superfamily of carbohydrate active enzymes that perform important functions in glycobiology—degradation of structural polysaccharides, remodeling of cell-surface proteoglycans and glycolipids, and even activation of small molecule natural products from latent precursor compounds [1–5]. Glycoside hydrolases (also referred to as glycosidases) cleave acetal linkages (glycosides) between a sugar and an aglycone moiety. Classification of GH families is based on the configuration (α- vs. β-) and location of the hydrolyzed bond (i.e., endo- vs. exo-), the identity of the sugar residue that is cleaved (e.g., D-glucose and D-galactose), and the three-dimensional shape of the hydrolase. The widely used repository of Carbohydrate Active Enzymes (CAZy database, www.cazy.org (accessed on)) [6] classifies CAZymes in families based on sequence similarity and three-dimensional (3D) folds. Among glycoside hydrolases, there are currently (March 2021) 170 GH families. Members of each family share the same reaction mechanism and 3D fold, but different substrate specificities are found in most of the families, meaning that substrate specificity is dictated by subtle structural differences at the active site. Substrates typically consist of a pyranose or a furanose ring attached to a variety of aglycone species. Seven-membered ring septanose glycosides are all but unprecedented in natural systems [7,8]. While they are not true septanosides because the seven-membered ring is not linked through a glycosidic linkage, the recently reported natural products portulasoid **1** and 20-hydroxy-ecdysone septanoside **2** from Atriplex portulacoides roots (Figure 17.1) [9] open the provocative possibility that other similar structures, in which the polyhydroxy oxepane might be a septanoside (e.g., **4** in Figure 17.1), could be discovered. In principle, it might be possible that septanoside hydrolases are present in nature or that promiscuous glycoside hydrolases might also accept septanoside substrates. Indeed, it is known that the seven-member ring iminosugar 1,6-dideoxy-1,6-imino-L-iditol was a glycosidase inhibitor [10], and it was reported that 4-nitrophenyl L-idoseptanosides, septanose analogs from hexoses, were substrates of some glycoside hydrolases, albeit weak ones [11].

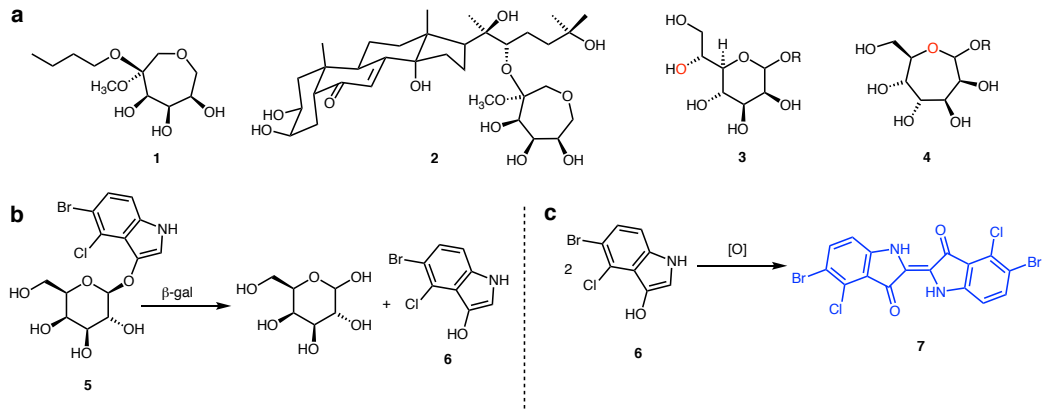


Figure 17.1. (a) Portulasoid 1 and 20-hydroxy-ecdysone septanoside 2 from Atriplex portulacoides roots; D-glycero-D-manno-heptose in its pyranose (3) and septanose (4) ring forms. (b) Hydrolysis of X-Gal 5, giving rise to D-galactose and 3-hydroxyindole 6. (c) Oxidative dimerization of 3-hydroxyindole 6 to indigo dye 7.

The identification of a naturally occurring septanoside hydrolase or other GH sufficiently promiscuous to hydrolyze septanosides would be an important milestone. Such an enzyme, or mutants thereof, would open up whole new areas of glycobiology research. For example, discovery of a septanoside hydrolase would motivate the search for septanoside-containing glycoconjugates in the producing organism. Plants are a candidate kingdom in this search because of the discovery of compounds **1** and **2**. An even more promising source, however, might be Gram-negative bacteria. There are numerous examples of heptose-containing glycoconjugates present in Gram-negative species [12,13], with the sedoheptulose/D-glycero-D-manno-heptose (3) biosynthetic pathway being a representative [14]. Cyclization of **3** through its C6 hydroxyl group (red) instead of the C5 hydroxyl could ultimately give rise to septanose glycosides **4**. Biosynthetic pathways involving heptoses are still in the early days of being thoroughly characterized, which also adds an incentive to a search for new glycosyl hydrolases. A first step in any search for septanoside hydrolases, therefore, is the development of tools—an assay strategy and matching substrate—that could be used to identify these enzymes.

One straightforward approach to search for septanoside hydrolases borrows from the β-galactosidase assay widely used in molecular biology. Upon glycoside hydrolysis of a pro-chromogenic substrate such as 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside **5** (X-Gal) [15], D-galactose is liberated along with 3-hydroxyindole **6** (Figure 17.1.b). Oxidative dimerization of **6** results in the formation of an easily detected indigo dye **7**. Development of a blue color in a bacterial colony or culture of when X-Gal is present therefore reports on the presence of the β-galactosidase enzyme. In practice, the “blue-white colony screen” is

frequently applied to the preparation of a protein expression vector [16]. The multiple cloning site where a DNA construct encoding the protein of interest will be inserted is positioned within the β-galactosidase gene on the plasmid. Plasmids lacking the insert retain a functioning β-galactosidase and hence lead to blue colonies after transformation. Successful insertion of the construct into the plasmid, on the other hand, disrupts the β-galactosidase gene and the associated hydrolase activity, giving rise to white colonies after transformation. In the context of our search, we envisioned using indolyl septanosides as candidate substrates of putative septanoside hydrolases. If hydrolytic activity toward such a substrate were present in a bacterial colony, we would anticipate the appearance of a characteristic blue color. To that end, we report here the syntheses of indolyl septanoside substrates for the discovery of septanoside hydrolases. Their structures mimic X-Gal, allowing them to be used for *in vivo* screening of bacteria. We further demonstrate that a related BODIPY–septanoside conjugate is transported into the cyto-plasm of *E. coli*, probably using any of its numerous sugar transporters. Development of these tools constitutes the first steps toward the identification of septanoside hydrolase enzymes.

Results

Synthesis of Indolyl Septanosides

Indolyl septanosides **8** and **9** (Figure 17.2.a) became the initial targets of substrate synthesis for a few reasons. First among them were the structural analogy of **8** and **9** to X-Gal **5** and reports on methods for the synthesis and utilization of indolyl glycosides in glycosidase assays [17]. The substitution pattern on the 5-bromo-3-indolyl aglycone was based on the ready availability of the 5-bromo-anthranilic acid starting material and the color of tyrian purple dye (an analog of 7 lacking the chlorine atoms) that arises from its oxidative dimerization [18,19]. Further, the stereogenic centers from C3-C6 in β-configured D-glycero-D-gulo-septanoside **8** and D-glycero-D-ido-septanoside **9** have identical relative configurations as D-glucose from C2-C5. The “phased” correspondence between these centers was important to recognition of septanosides by lectins [20–22]; it refers to the fact that C2 of the pyranose is the same as the C3 of the septanose and so forth. We also considered the possible promiscuity of glucosidases as giving us a higher probability of identifying active GHs in our search. The synthetic strategy drew from reported methods for indolyl glycoside synthesis [18,23–25] and our previous experience at preparing septanose glycosides via nucleophilic displacement on anomeric bromides [26]. It leveraged a synthesis of the per-O-acetyl septanose precursors of

the anomeric bromides that began from natural D-pyranosides such as D-glucose, suggesting that the route could be extended to other sugars.

Preparation of indolyl septanoside **8** leveraged the known conversion of per-O-acetyl septanose **10** to its corresponding α-configured anomeric bromide **11** (93%, Figure 17.2.b) [26]. Attempted glycosylation of **11** via direct displacement (SN2) conditions with 5-bromo-indoxyl (i.e., the analog of **12** lacking the C2 carboxylate group) was unsuccessful due to rapid oxidative dimerization of the indoxyl species under the basic reaction conditions (**11** + **12**, potassium tert-butoxide in acetonitrile, 0 °C to rt). We therefore resorted to a strategy that used 2-carbomethoxy indoxyl **12** as acceptor. The carbomethoxy moiety at the 2-position of the indoxyl ring is a blocking group that prevents oxidative dimerization during the glycosylation [18]. In the event, glycosylation of **11** with **12** gave protected septanoside **13** in 48% yield; the yield for the glycosylation is modest but was not extensively optimized. Recompense for successful glycosylation in this instance was the multi-step deprotection that had to be undertaken to arrive at the target. Compound **13** was therefore converted to key intermediate **14** by a sequence that included removal of the acetate protecting groups and the methyl esters in a two-step process. Re-acetylation of the hydroxyl groups in this species was concomitant with decarboxylation of the carboxy group yielding **13** in 73% over the three steps. It proved convenient to purify **13** at this stage before the final de-acetylation. Removal of the acetates under Zemplén conditions then provided 5-bromo-3-indolyl septanoside **8** in 75% yield. Overall, the yield of **8** over the six-step sequence was 24% with an average yield per step of 73%. Compound **9**, the C2′-epimer of 8, was prepared from per-O-acetyl septanose **15** using the same set of re-actions, and shown in Figure 17.2.b, in 35% yield over the six-step sequence.

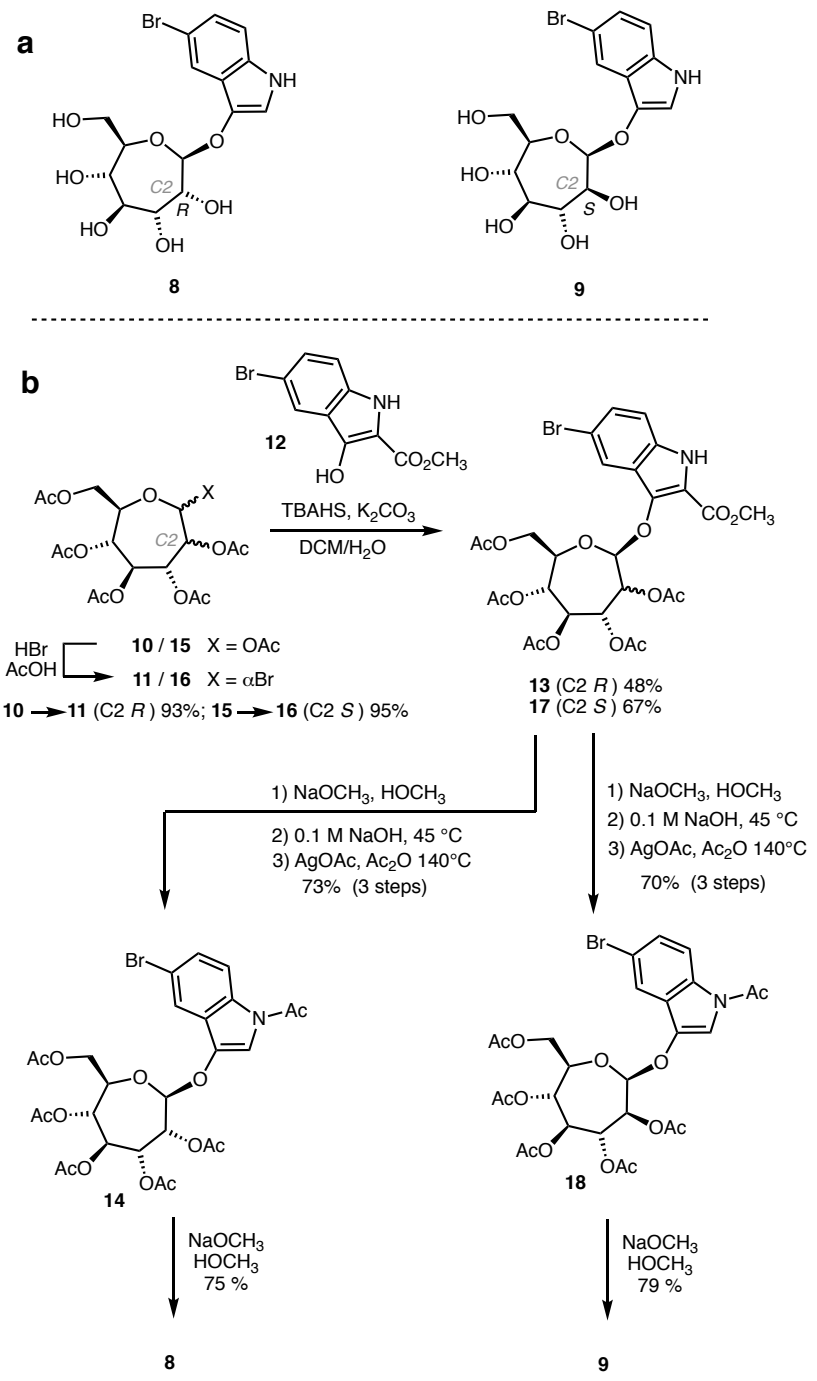


Figure 17.2. (a) Indolyl septanosides 8 and 9. (b) Synthetic routes, including yields, used to prepare indolyl septanosides 8 and 9.

Indolyl septanoside **8** was characterized both structurally and functionally. NMR spectroscopic data for **8** were consistent with a β-configured anomeric center (H1 δ = 4.93 ppm, C1 δ =

110.1 ppm) and a trans disposition of the groups at C1 and C2 ($^3J_{H1,H2}$ = 5.5 Hz). These values were comparable to other 1,2-trans β-septanosides we have synthesized, including p-nitrophenyl (pNP) septanoside **19** (vide infra). Values for the 1,2-cis indolyl septanoside **9** (H1 δ = 5.09 ppm, C1 δ = 105.5 ppm, and $^3J_{H1,H2}$ = 2.3 Hz) were also consistent with its proposed structure [8,26]. In a separate line of experimentation, septanoside **9** was exposed to buffer solutions at pH 5, 6, and 7, as well as a 1 M solution of HCl. With standing at room temperature over 8 h, color developed only in the 1 M HCl samples. We interpreted the results to be a demonstration of the stability of the indolyl glycosides at pH values near neutrality and in a range where many GH enzymes are active. The results also demonstrated the susceptibility to acid hydrolysis. In a preparative scale experiment, the indigo precipitate that arose from 1M HCl hydrolysis of **9** was collected, redissolved in N,N-dimethylformamide, and its UV–vis spectrum compared to a sample of tyrian purple synthesized independently by a different method [27–29].

Evaluation of Indolyl Septanoside 8 as a Substrate of Common Exo-Glycoside Hydrolases

Compound **8** was assayed *in vivo* with *E. coli* cells under conditions that are typical for assays that use X-Gal substrate **5**. *E. coli* BL21(DE3) Star cells harboring an inducible plasmid overexpressing the *E. coli* β-galactosidase (BL21-pGal+) and cell harboring an empty plasmid (BL21-pGal–) were grown on Petri plates containing compound **8** with isopropyl 1-thio-β-D-galactopyranoside (IPTG) as inducer. No blue colonies were observed (as opposed to the control BL21-pGal+ with **7**) (Figure 17.3). At first glance, the apparent lack of activity, as evidenced by the lack of appearance of blue colonies, may reflect that *E. coli* β-galactosidase did not recognize septanoside **8**. It may also have indicated that the substrate was not internalized into the *E. coli* cells. To test the first scenario, we assayed the purified *E. coli* β-galactosidase with compound **8** under the same experimental conditions where it shows maximum activity with its cognate substrate (Table 17.1). Once again, no hydrolytic activity was detected, confirming that septanoside **8** is not a β-galactosidase substrate.

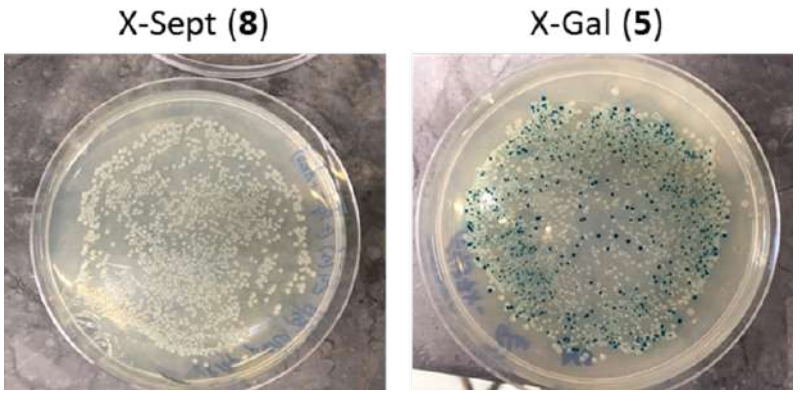


Figure 17.3. Blue-white colony screen of E. coli BL21(DE3) Star cells with indolyl septanoside 8. BL21(DE3) Star cells harboring a plasmid expressing E. coli β-galactosidase: (Left) with X-Sept (8) substrate; and (Right) with substrate X-Gal (5).

With the aim of potentially finding an enzyme that was able to recognize septanoside substrates, indolyl septanoside **8** was subjected to a panel of glycosidases to test *in vitro* for their ability to hydrolyze the substrate (Table 17.1). Each enzyme was assayed at high enzyme concentration (25–50 U/mL, activity units with its cognate substrate) with compounds **8** and **9** (1 mM) at the pH optimum of each enzyme. None of them showed measurable activity (by colorimetry of released indigo dye), including the *E. coli* β-galactosidase as indicated above. Only the *Streptomyces sp.* β-glucosidase (Bgl3) resulted in faint activity. This is a rather promiscuous enzyme reported to hydrolyze aryl glycosides of D-glucose, D-galactose, D-mannose, D-xylose and L-fucose [30]. We then analyzed the activity of Bgl3 against p-nitrophenyl septanoside **19**, prepared as previously reported [26], in greater detail. At a fixed Bgl3 concentration (9.2 μM), initial rates at increasing concentrations of **19** (Figure 17.4) were fit to give the following kinetics parameters: $k_{cat} = 5.9 \times 10^{-7}$ min^{-1}, $K_M = 3$ mM, and h = 3.8. As compared to an efficient βgl3 substrate (pNP-β-glucoside, $k_{cat} = 13.9 \times 10^{-2}$ min^{-1}, $K_M = 0.8$ mM, and $K_i = 0.6$ mM), the data show that **19** is a poor substrate for the glycosidase with a turnover of approximately five orders of magnitude slower than the glucoside.

Table 17.1. *In vitro* screen of selected glycosidases.

Enzyme	Substrate	S.A. (U/mg) [a]	Conditions
β-Galactosidase (*Escherichia coli*)	pNP-Gal	35	Phosphate buffer (100 mM) pH 6.5 40 °C
	8	n.d. [b]	
	9	n.d.	
β-Galactosidase (*Aspergillus niger*)	pNP-Gal	170	Acetate buffer (100 mM) pH 4.5 40 °C
	8	<0.001	
	9	n.d.	
β-Glucosidase (*Streptomyces* sp.)	pNP-Glc	3.3	Phosphate buffer (50 mM) pH 6.5 50 °C
	8	<0.005	
	9	n.d.	
β-Glucosidase (*Almonds*)	pNP-Glc	2	Phosphate buffer (100 mM) pH 5.0 37 °C
	8	n.d.	
	9	n.d.	
β-Glucosidase (*Thermatoga maritima*)	pNP-Glc	70	Maleate buffer (50 mM) pH 6.5 40 °C
	8	n.d.	
	9	n.d.	
β-Glucosidase (*Phanerochaete chyrosporium*)	pNP-Glc	100	Acetate buffer (100 mM) pH 5.0 40 °C
	8	n.d.	
	9	n.d.	
β-Mannosidase (*Cellulomonas fimi*)	pNP-Man	10	Maleate buffer (100 mM) pH 6.5 35 °C
	8	<0.001	
	9	<0.001	
β-Glucuronidase (*Escherichia coli*)	pNP-GlcA	110	Tris·HCl buffer (100 mM) pH 7.5 37 °C
	8	n.d.	
	9	n.d.	
β-Xylosidase (*Bacillus pumilus*)	pNP-Xyl	18	Phosphate buffer (50 mM) pH 7.5 35 °C
	8	n.d.	
	9	n.d.	

[a] Specific activity (U/mg) at 1 mM pNP-substrates. [b] n.d., not detected.

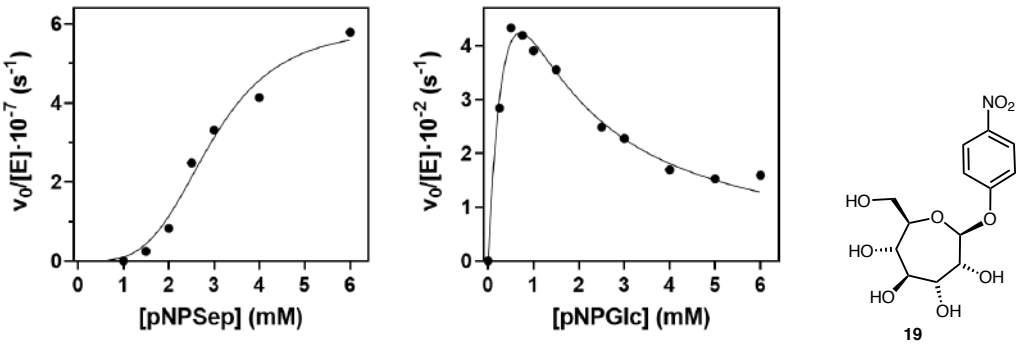


Figure 17.4. Kinetics of *Streptomyces sp.* β-galactosidase (Bgl3) with pNP-Sept (19) and pNP-Glc substrates. Conditions: 50 mM phosphate buffer, pH 6.5, and 50 °C.

Transport of a BODIPY Labeled Septanoside into *E. coli* Cells

In search of septanoside hydrolases, we intend to use compound **8** for the screening of plant, yeast, and bacterial cell extracts as well as for *in vivo* "blue-white" screening of bacterial cells harboring libraries of glycoside hydrolases (i.e., metagenomics libraries) [31]. It was necessary, therefore, to return to the question of whether or not **8** was capable of entry into *E. coli* cells. To that end, we investigated the transport of septanosides into the bacterial cytoplasm. As a proof-of-concept, a BODIPY-labeled septanoside (i.e., **23**, Figure 17.5) was synthesized as a surrogate of compound **8** to be used in confocal microscopy studies. BODIPY is a commonly used fluorophore tag that can be used to analyze uptake by cells [32,33] and for cell sorting by fluorescence-activated cell sorting (FACS) [34,35]. A Huisgen click reaction [33,36] between commercially available BODIPY alkyne **20** and the known septanosyl azide **21** [26] gave protected conjugate **22** in 48% yield. Deprotection of compound **22** under Zemplén conditions then afforded BODIPY-labeled septanoside **23**.

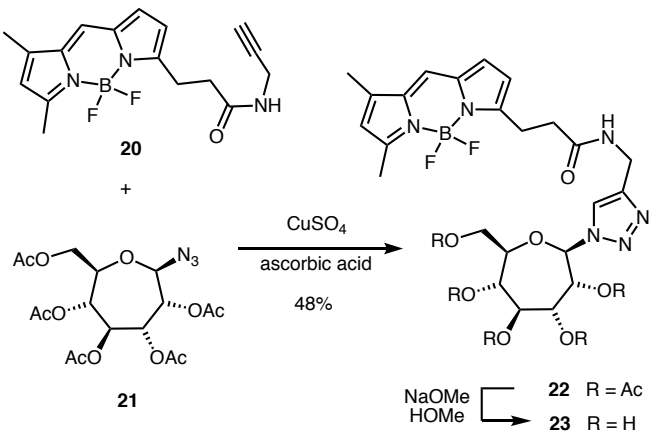


Figure 17.5. Synthesis of BODIPY-labeled septanoside 23.

Two different *E. coli* strains, MG1655 and TOP10, were used to analyze cellular uptake by confocal microscopy (Figure 17.6.). Prior to microscopy, cultures were grown to OD 1.0, washed, and then suspended in 10 mM MES buffer at pH 5.5. An aliquot of this culture was incubated with FM4-64 (localizes into bacterial membrane), BODIPY septanoside **23**, and DAPI (DNA stain), and then spun down and resuspended prior to observation. With MG1655 cells, the BODIPY fluorescence clearly localized in the cytoplasm with high intensity, confirming internalization of the septanoside. A lack of BODIPY fluorescence intensity at the cell periphery indicated that the labeled septanoside was not attached to the outer membrane.

The merged image in Figure 17.6 shows the colocalization of the blue and green images, consistent with transport of **23**, which was considered a surrogate for indolyl septanoside 8 in these experiments. Similar results were obtained in experiments with the TOP10 cells, where **23** also localized in the cytoplasm of this *E. coli* strain.

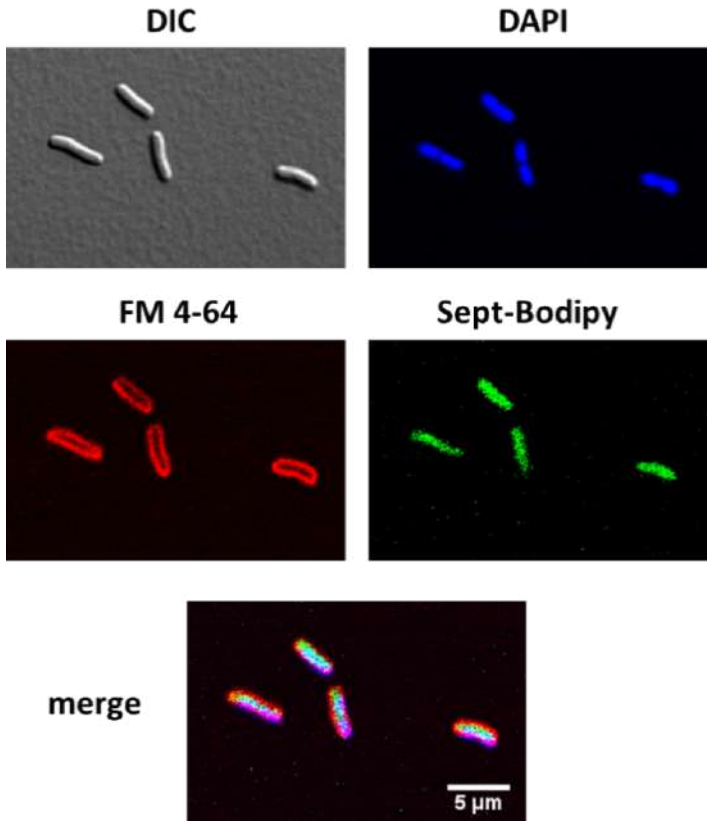


Figure 17.6. Confocal fluorescence microscopy of E. coli MG1655 cells with labeling stains. DIC, optical microscopy image; DAPI, DNA staining (bacterial chromosome); FM 4-64, membrane staining; Sept-BODIPY, BODIPY-labeled septanoside (23). Bar scale (5 µm) is the same in all images.

Discussion

The investigation reported here has established a conceptual framework behind the search for septanoside hydrolases from natural sources. Such a search will allow the identification of catabolic enzymes from organisms such as Atriplex portulacoides that produce polyhydroxy oxepanes (the source of compounds **1** and **2**) or Gram-negative organisms that utilize heptoses. Additionally, metagenomics libraries expressed in bacteria such as *E. coli* can be interrogated as well for hydrolytic activity. Our approach is fundamentally molecular—small molecule synthesis was used to prepare tools that enable searching both *in vitro* and, critically,

in vivo for enzymes. Indolyl septanosides are useful reagents because they enable the colorimetric characterization of glycoside hydrolysis in a manner analogous to X-Gal. Preliminary assays in *E. coli* to characterize the reactivity of indolyl D-glycero-D-gulo-septanoside **8** as a substrate of β-galactosidase showed no hydrolysis. The apparent lack of reactivity of this septanoside substrate was subsequently explored along two different lines. First, *in vitro* experiments with a panel of commercially available glycosidases also showed no significant hydrolysis of the unnatural substrates **8** and **9**. Significantly, Bgl3 (β-glucosidase from *Streptomyces sp.*) showed a suggestive activity with light-pale blue colonies when using substrate **8**. Further analysis with p-nitrophenyl septanoside **19** confirmed the low activity, in the range of 5–6 orders of magnitude lower than its p-nitrophenyl D-glucoside substrate. In a second line of inquiry, BODIPY-labeled septanoside **23**, prepared via click reaction with a septanosyl azide, was observed inside *E. coli* cells by confocal microscopy. It suggested that septanosides traverse the bacterial envelope. Taken together, these results illustrate an approach that opens the door to a broader search for septanoside hydrolase activity in bacterial systems.

Materials and Methods

General Experimental

Commercially available starting materials, reagents, and solvents (Sigma-Aldrich, St. Louis, MO, USA and Acros/Thermofisher, Waltham, MA, USA) were used without further purification. Septanosyl azide **21** was prepared as previously reported [26]. Reactions were performed under nitrogen atmosphere unless otherwise noted and were monitored by TLC using silica gel HL plates w/UV254, 250 µm (SiliCycle, Quebec City, PQ, Canada), visualized either under UV lamp or by charring with 2.5% *p*-anisaldehyde in H_2SO_4, AcOH and EtOH solutions. Reverse phase C18 TLC plates w/UV254, 200 µm (Sorbtech, Atlanta, GA, USA) were also used and visualized by UV. Flash chromatography was performed on silica gel (60 Å, 40–63 µm). ^{1}HNMR spectra were collected on Bruker NMR instruments at either 400 and 500 MHz with chemical shift referenced to $(CH_3)_4Si$ (δ_H 0.00 ppm) or the residual peak in $CDCl_3$ (δ_H 7.24 ppm) or CD_3OD (δ_H 3.31 ppm). ^{13}C NMR spectra were collected at 100 MHz and referenced to residual peak in $CDCl_3$ (δ_C 77.2 ppm) or CD_3OD (δ_C 49.1 ppm). High resolution mass spectrometry data were collected on a JEOL DART Electrospray-Time-of-Flight (AccuTOF-DART) instrument. Enzymes were purchased from Megazyme Ltd. (Bray, Ireland). β-Glucosidase from *Streptomyces* was expressed and purified as reported in [30]. DAPI (4′,6-

diamidine-2′-phenylindole dihydrochloride) and FM 4-64 (*N*-(3-triethylammoniumpropyl)-4-(6-(4-(tiethylamino)phenyl)hexatrienyl)pyridinium dibromide) were purchased from Sigma-Aldrich and Thermofisher, respectivelySynthesis of Indolyl Septanosides **8** and **9**

Phase-Transfer Glycosylation (General Procedure). Freshly prepared per-*O*-acetyl septanosyl bromide (i.e., **10**) [26] (0.306 g, 0.63 mmol), tetra-*N*-butylammonium hydrogen sulfate (TBAS, 0.214 g, 0.63 mmol), and 2-methoxycarbonyl-3-hydroxy-5-bromo-indole **11** (0.210 g, 0.79 mmol) were mixed in DCM (8 mL). To this solution was added 12 mL of a 1 M aqueous solution of K_2CO_3. The biphasic (organic:aqueous) reaction mixture was stirred vigorously at room temperature until consumption of the bromide was complete as determined by TLC (1:1 Hex:EtOAc). Afterwards, the organic phase was separated from the aqueous phase; the aqueous phase was extracted with additional DCM (1 × 12 mL). The combined organic layers were dried with Na_2SO_4, filtered, and the and the solvent was removed under reduced pressure. The residue was then purified by column chromatography in the solvent mixtures stated.

5-bromo-2-methoxycarbonyl-indol-3-yl 2,3,4,5,7-penta-*O*-acetyl-β-D-*glycero*-D-*gulo*-septanoside (**13**). Obtained as brownish oil from **11** and **12** to yield 0.210 g (48%) of **13**. R_f 0.4 (1:1 Hex: EtOAc); ^{1}H NMR (400 MHz, $CDCl_3$) δ 9.28 (s, 1H), 7.93 (broad s, 1H), 7.36 (dd, *J* = 8.8, 1.8 Hz, 1H), 7.23 (d, *J* = 8.8 Hz, 1H), 5.74 (dd, *J* = 4.7, 2.2 Hz, 1H), 5.64 (dd, *J* = 8.8, 2.2 Hz, 1H), 5.49–5.45 (m, 1H), 5.30 (d, *J* = 4.8 Hz, 1H), 5.22 (dd, *J* = 9.4, 6.0 Hz, 1H), 4.13 (dd, *J* = 14.3, 7.1 Hz, 1H), 4.08–4.02 (m, 1H), 3.90 (s, 3H), 3.88–3.85 (m, 1H), 2.22 (s, 3H), 2.12 (s, 3H), 2.11 (s, 3H), 2.01 (s, 3H), 1.72 (s, 3H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 170.6, 169.5, 169.4, 169.3, 169.1, 161.4, 139.0, 132.5, 129.2, 123.4, 123.2, 117.0, 113.8, 113.7, 105.4, 73.3, 72.7, 71.2, 70.9, 69.3, 63.5, 60.5, 52.1, 21.0, 20.7, 20.7(2), 20.3, 14.2; HRMS (DART-TOF) *m/z* calcd. for $C_{27}H_{30}BrNO_{12}Na$ $[M+Na]^+$ 694.0747, obs. 694.0743.

5-bromo-2-methoxycarbonyl-indol-3-yl 2,3,4,5,7-penta-*O*-acetyl-β-D-*glycero*-D-*ido*-septanoside (**17).** Reaction of **12** with septanosyl bromide **16** using the general procedure gave 0.160 g (67%) of **17** as a brownish syrup. R_f 0.3 (3:2 Hex: EtOAc) ^{1}H NMR (400 MHz, $CDCl_3$) δ 9.16 (s, 1H), 7.96 (d, J = 1.7 Hz, 1H), 7.37 (dd, J = 8.8, 1.8 Hz, 1H), 7.23 (d, J = 8.8 Hz, 1H), 5.66 (dd, J = 7.4, 2.1 Hz, 1H), 5.61 (m, 1H), 5.52 (d, J = 2.2 Hz, 1H), 5.40–5.33 (m, 2H), 4.09 (dd, J = 12.2, 5.7 Hz, 1H), 3.98 (s, 3H), 3.94 (dd, J = 12.0, 2.2 Hz, 1H), 3.92 (ddd, J = 11.0, 5.8, 2.2 Hz, 1H), 2.18 (s, 3H), 2.10 (s, 3H), 2.04 (s, 3H), 1.98 (s, 3H), 1.68 (s, 3H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 170.8, 170.0, 169.5, 169.2, 169.0, 161.4, 139.6, 132.6, 129.5, 123.4, 123.1,

116.5, 114.0, 113.7, 103.2, 76.4, 72.1, 72.0, 70.5(2), 63.4, 60.6, 52.4, 20.9, 20.7, 20.6(2), 20.3, 14.3; HRMS (DART-TOF) m/z calcd. for $C_{27}H_{31}BrNO_{12}$ $[M+H]^+$ 672.0928, obs. 672.0923.

Multistep De-acetylation, Decarboxylation, and Acetylation (General Procedure). At room temperature, a solution of approximately 0.3 mmol of the protected indolyl septanoside (i.e., **13** or **17**) in 5 mL MeOH was treated with a catalytic amount of sodium methoxide (5 mol %). After starting material disappeared from the TLC (1:1 Hex:EtOAc), indicating that deacetylation was complete, the solvent was removed, and NaOH (0.1 M aq.; 14 mL) was added to the residue. The mixture was heated to 40–45 °C and stirred until C18 reverse-phase TLC indicated that ester hydrolysis was complete, approximately 4 h (R_f of the product is 0.7 and starting material is 0.3 in 3:1 water:acetonitrile). The mixture was lyophilized, then AgOAc (3.0 equiv.), K_2CO_3. (6.0 equiv.), and Ac_2O (10.0 mL) were added. The mixture was heated to 110 °C for 1 h. Then the mixture was cooled to room temperature and diluted with 20 mL water and 20 mL CH_2Cl_2. The organic phase was washed with water (2 × 20 mL) and dilute aqueous $NaHCO_3$ (~0.1 M, 1 × 20 mL). The organic phase was dried with Na_2SO_4, filtered, and the eluent was removed under reduced pressure. The residue was then purified by column chromatography in the solvent stated.

N-Acetyl-5-bromoindol-3-yl 2,3,4,5,7-penta-*O*-acetyl-β-D-*glycero*-D-*gulo*-septanoside (**14**). Obtained as a yellowish solid from **13** using the general procedure (73% yield over 3 steps). R_f 0.3 (1:1 Hex:EtOAc) ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.25 (broad s, 1H), 7.69 (d, *J* = 1.8 Hz, 1H), 7.46 (dd, *J* = 8.8, 1.8 Hz, 1H), 7.30 (broad s, 1H), 5.57–5.54 (m, 2H), 5.44 (dd, *J* = 8.0, 4.7 Hz, 1H), 5.15 (d, *J* = 5.4 Hz, 1H), 4.33–4.31 (m, 1H), 4.17–4.09 (m, 3H), 2.62 (s, 3H), 2.20 (s, 3H), 2.14 (s, 3H), 2.13 (s, 3H), 2.06 (s, 3H), 2.03 (s, 3H), 1.93 (s, 3H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 170.6, 169.3(2), 169.2, 169.0, 168.5, 140.0, 132.3, 129.1, 126.1, 120.8, 118.1, 117.2, 112.2, 103.5, 74.0, 72.2, 71.6, 70.6, 69.2, 63.7, 60.5, 24.0, 21.2, 21.0, 20.8(2), 20.7, 14.3; HRMS (DART-TOF) *m/z* calcd. for $C_{27}H_{30}BrNO_{13}Na$ $[M+Na]^+$ 678.0798, obs. 678.0785.

N-Acetyl-5-bromoindol-3-yl 2,3,4,5,7-penta-O-acetyl-β-D-*glycero*-D-*ido*-septanoside (**18**). Obtained as yellow solid from **17** using the general procedure (70% yield over 3 steps) R_f 0.2 (3:2 Hex: EtOAc) ^{1}H NMR (400 MHz, CD_3OD) δ 8.20 (d, J = 8.8 Hz, 1H), 7.62 (d, J = 1.9 Hz, 1H), 7.41 (dd, J = 8.8, 1.9 Hz, 1H), 7.37 (broad s, 1H), 5.64 (d, J = 1.5 Hz, 1H), 5.59–5.55 (m, 2H), 5.49–5.45 (m, 1H), 5.36 (app t, J = 9.3 Hz, 1H) 4.33 (ddd, J = 9.2, 7.4, 3.6Hz, 1H), 4.23 (app d, J = 3.6 Hz, 2H), 2.60 (s, 3H), 2.15 (s, 3H), 2.06 (s, 3H), 2.02 (s, 3H), 2.00 (s, 3H), 1.76 (s, 3H); ^{13}C NMR (100 MHz, CD_3OD) δ 172.1, 171.3, 171.0, 170.9, 170.7, 170.6, 140.9, 133.4,

129.8, 127.3, 121.5, 119.0, 117.7, 114.9, 112.2, 101.6, 76.7, 73.2, 72.7, 71.3, 71.2, 64.4, 23.9, 20.6(2), 20.5; HRMS (DART-TOF) *m/z* calcd. for $C_{27}H_{30}BrNO_{13}Na$ $[M+Na]^+$ 678.0798, obs. 678.0816.

Zemplén De-acetylation (General Procedure). A solution of the starting material (1.00 mmol) in MeOH (15.0–20.0 mL) was treated with a catalytic amount of sodium methoxide (5 mol%). The solution was stirred at room temperature overnight (~16 h). If the product had precipitated during this time it was collected by filtration, otherwise the solution was neutralized with Amberlite IR-120 (H+) resin (pH 7 by pH paper) and then the solution was concentrated to near dryness. The residue was then re-dissolved in a minimum amount of water (≤5.0 mL) and subjected to lyophilization.

5-Bromoindol-3-yl β-D-*glycero*-D-*gulo*-septanoside (**8**). Obtained from **14** as tan powder after lyophilization using the general procedure (75% yield). R_f 0.4 (4:1 DCM: MeOH) ^{1}H NMR (400 MHz, CD_3OD) δ 7.75 (s, 1H), 7.23 (s, 1H), 7.19 (d, J = 8.7 Hz, 1H), 7.14 (d, J = 8.7 Hz, 1H), 4.99 (s, 6H), 4.93 (d, J = 5.5 Hz, 1H), 4.21 (app t, J = 4.4 Hz, 1H), 3.89 (dd, J = 9.1, 3.3 Hz, 1H), 3.86–3.79 (m, 2H), 3.66–3.60 (m, 2H), 3.46 (app t, J = 7.9 Hz, 1H); ^{13}C NMR (100 MHz, CD3OD) δ 137.9, 133.9, 125.6, 123.2, 121.0, 114.2(2), 112.7, 110.1, 82.3, 75.6, 75.5, 73.9, 72.2, 64.4; HRMS (DART-TOF) *m/z* calcd. for $C_{15}H_{19}BrNO_7$ $[M+H]^+$ 404.0345, obs. 404.0338.

5-Bromoindol-3-yl β-D-*glycero*-D-*ido*-septanoside (**9**). Obtained as light brown solid from **18** using the general procedure (79% yield) R_f 0.4 (4:1 DCM: MeOH) ^{1}H NMR (400 MHz, CD_3OD) δ 7.87 (s, 1H), 7.78 (d, J = 1.8 Hz, 1H), 7.20 (m, 1H), 7.16 (dd, J = 8.6, 1.8 Hz, 1H), 5.09 (d, J = 2.3 Hz, 1H), 4.88 (s, 6H), 4.00 (dd, J = 6.3, 2.2 Hz, 1H), 3.86–3.83 (m, 1H), 3.71–3.66 (m, 2H), 3.61 (dd, J = 14.1, 7.0 Hz, 1H), 3.52–3.50 (m, 2H); ^{13}C NMR (100 MHz, CD_3OD) δ 137.7, 133.9, 125.7, 123.3, 121.1, 114.2(2), 112.8, 105.5, 83.2, 76.7, 75.7, 75.5, 74.1, 64.3; HRMS (DART-TOF) *m/z* calcd. for $C_{15}H_{18}BrNO_7Na$ $[M+Na]^+$ 426.0164, obs. 426.0204.

Synthesis of BODIPY-Labeled Septanoside **23**

BODIPY:2,3,4,5,7-Penta-O-acetyl-β-D-glycero-D-gulo-septanosyl azide conjugate (**22**). BODIPY-FL alkyne **20** (0.0025 g, 0.008 mmol) and septanosyl azide **21** (0.0085 g, 0.02 mmol) and were dissolved in a mixture of THF:water (1 mL of 3:1 mixture). A solution of $CuSO_4 \cdot 5H_2O$ (0.001 g, 0.5 eq.) and sodium ascorbate (0.0012 g, 0.8 eq.) in an additional 1 mL of the THF:water was sonicated for 10 min and then added to the initial alkyne-azide solution. The mixture was allowed to react at room temperature with stirring for 12 h. After, the reaction

mixture was diluted with 5 mL EtOAc and the aqueous and organic layers separated; the aqueous layer was then extracted with additional EtOAc (2 × 5 mL). The combined organic layers were dried over Na_2SO_4, filtered, and the solvent was evaporated under reduced pressure. The crude product was then purified by column chromatography to deliver **20** as brownish oil (0.0028 g, 48% yield) Rf 0.4 (3:2 Hex: EtOAc); 1H NMR (400 MHz, CDCl3) δ 7.09 (s, 1H), 6.88 (d, *J* = 3.9 Hz, 1H), 6.28 (d, *J* = 4.0 Hz, 1H), 6.12 (s, 1H), 5.82 (broad s, 1H), 5.40 (dd, *J* = 8.0, 1.9 Hz, 2H), 5.34 (dd, J = 8.0, 3.9 Hz, 2H), 5.10 (dd, *J* = 6.5, 1.9 Hz, 2H), 5.05–5.00 (m, 2H), 4.18–4.12 (m, 3H), 3.28 (t, *J* = 7.5 Hz, 2H), 2.67–2.63 (m, 2H), 2.57 (s, 3H), 2.26 (s, 3H), 2.12 (s, 3H), 2.11 (s, 9H), 2.09 (s, 3H); 13C NMR (100 MHz, CDCl3) δ 171.5, 170.8, 169.3, 169.1 (2), 128.5, 124.0, 120.7, 117.7, 91.3, 79.7, 77.4, 76.0, 72.1, 71.8, 71.6, 70.6, 69.4, 64.3, 36.1, 29.9, 29.4, 25.0, 20.9 (3), 15.2, 11.5; HRMS (DART-TOF) m/z calcd. for C34H41BF2N6O12Na [M+Na]+ 797.2730, obs. 797.2725.

BODIPY: β-D-glycero-D-gulo-septanosyl azide conjugate (23). A solution of starting per-O-acetylated BODIPY conjugate **22** (0.0028 g, 0.004 mmol) was dissolved in dry methanol (0.75 mL) and kept under an inert atmosphere and dark conditions. To this solution was added 5 mol % of NaOMe in MeOH. The reaction was stirred at room temperature for 12 h and monitored by TLC (Rf of the product is 0.1 and starting material is 0.5 in 1:1 Hex/EtOAc). Upon complete disappearance of the SM, the reaction solvent was evaporated under the reduced pressure and the residue was re-dissolved in de-ionized water (1.00 mL) and subjected to lyophilization to obtain 23 as a brownish solid. The crude product was used in confocal microscopy experiments without further purification.

Glycosidase Activity Assay on *E. coli* Cells

Twenty microliters of 100 mM IPTG solution in water and 20 µL of a 20 mg/mL X-Gal (5) or X-Sept (**8**) solution in DMF were spread on LB agar Petri plates (20 mL medium) containing kanamycin (100 µg/mL) and allowed to dry for 20 min at 37 °C. Competent BL21 (DE3) Star cells were transformed (chemically competent cells [37]) with plasmid pRSF-β-gal encoding the *E. coli* β-galactosidase and spread onto the agar plates. Color development was analyzed after 24 h incubation at 37 °C.

Probing Glycoside Hydrolases for Septanoside Hydrolysis

Activity assays were performed at the optimal pH and temperature for each enzyme as indicated in Table 17.1. Indolyl substrates **8** and **9** (1 mM) and enzymes (concentration

ranging from 25 to 50 U/mL) in buffer were incubated in 1 mL cuvettes and the absorbance monitored at 630 nm. Specific activities of the commercial enzymes with pNP substrates are from the manufacturer, and β-glucosidase βgl3 from *Streptomyces sp*. as reported [30].

Kinetics of *Streptomyces* β-Glucosidase with p-Nitrophenyl Septanoside **19**

Kinetics were performed by monitoring pNP release by absorbance at 400 nm. Reactions were done in thermostated cuvettes at 50 °C in 50 mM phosphate buffer pH 6.5 with 0.1–10 mM substrates and 46 nM enzyme (βgl3) for pNP-Glc or 9.2 µM βgl3 for pNP-septanoside (19). Rates were obtained from the initial slopes after subtracting the blank rates (in the absence of enzyme) and kinetic constants were calculated from data fitted to a Michaelis–Menten equation with substrate inhibition (Equation 17.1) for pNP-Glc and to a sigmoidal equation (Equation 17.2) for pNP-Sept (19) using GraphPad software (Prism, San Diego, CA, USA).

$$v = k_{cat}[E]_0[S]/(K_M + [S] + [S]^2/K_i)$$

Equation 17.1. M&M equation with substrate inhibition

$$v = k_{cat}[E]_0[S]^h/({K_M}^h + [S]^h)$$

Equation 17.2. Sigmoidal equation

where [S] is the substrate concentration, $[E]_0$ the enzyme concentration, and k_{cat} (catalytic constant), K_M (Michaelis constant), K_i (substrate inhibition constant) and h (Hill index) are the adjustable parameters [38].

Cell Internalization of Compound **23** by Confocal Microscopy

E. coli MG1655 cultures were grown to an OD_{600} of ~1.0 (1×10^9 cfu/mL) in Luria Broth (LB), washed, and suspended in 10 mM MES buffer pH 5.5 to a final OD_{600} = 1. To 100 µL of cell suspension (different dilutions tested, best results with a 1:2 dilution), BODIPY-septanoside **23** was added to a final 1 µM concentration with shaking. Cells were incubated for 1 h at 37 °C. Then, 1 µL FM 4–64 (1 µg/mL stock) and 1 µL DAPI (2 µg/mL stock) were added to the cells with shaking and incubated for 30 min at 37 °C for membrane and DNA staining, respectively. Cells were spun down by centrifugation for 60 s and resuspended in 1/10 their original volume. Three microliters of cell suspension were spotted onto 1.5% agarose pads containing solid LB medium. The cells were then imaged using a Nikon A1R spectral confocal microscope with a 60× oil immersion lens.

Author Contributions: A.R.P. and S.P., experimental and data analysis; and A.P. and M.W.P., conceptualization, supervision, and writing and funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by NSF grant CHE-1506567 (to M.W.P.) and grant PID2019-104350RB-I00 from MICINN, Spain (to A.P.).

Data Availability Statement: The data that support the findings of this study are available from the corresponding authors upon reasonable request.

Acknowledgments: S.P. acknowledges a predoctoral contract from MICINN, Spain. We thank Alfredo Angeles-Boza and Samuel Juliano (UConn) for their support on confocal microscopy experiments.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] Pallister, E.; Gray, C.J.; Flitsch, S.L. Enzyme promiscuity of carbohydrate active enzymes and their applications in biocatalysis. Curr. Opin. Struct. Biol. 2020, 65, 184–192, doi:10.1016/j.sbi.2020.07.004.

[2] Kytidou, K.; Artola, M.; Overkleeft, H.S.; Aerts, J.M.F.G. Plant Glycosides and Glycosidases: A Treasure-Trove for Therapeutics. Front. Plant Sci. 2020, 11, doi:10.3389/fpls.2020.00357.

[3] Gloster, T.M. Exploitation of carbohydrate processing enzymes in biocatalysis. Curr. Opin. Chem. Biol. 2020, 55, 180–188, doi:10.1016/j.cbpa.2020.01.015.

[4] Davies, G.J.; Planas, A.; Rovira, C. Conformational Analyses of the Reaction Coordinate of Glycosidases. Acc. Chem. Res. 2012, 45, 308–316, doi:10.1021/ar2001765.

[5] Sinnott, M.L. Catalytic Mechanisms of Enzymic Glycosyl Transfer. Chem. Rev. 1990, 90, 1171–1202, doi:10.1021/cr00105a006.

[6] Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P.M.; Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014, 42, D490–D495, doi:10.1093/nar/gkt1178.

[7] Dey, S.; Samanta, G.C.; Jayaraman, N. Advancements in synthetic and structural studies of septanoside sugars. In Recent Trends in Carbohydrate Chemistry; Rauter, A.P., Christensen, B., Somsak, L., Kosma, P., Adamo, R., Eds.; Elsevier Inc.: Amsterdam, The Netherlands, 2020; pp. 217–251.

[8] Saha, J.; Peczuh, M.W. Synthesis and properties of septanose carbohydrates. In Advances in Carbohydrate Chemistry and Biochemistry; Academic Press: Cambridge, MA, USA, 2011; Volume 66, pp. 121–186.

[9] Ben Nejma, A.; Nguir, A.; Ben Jannet, H.; Hamza, M.A.; Daïch, A.; Othman, M.; Lawson, A.M. New septanoside and 20-hydroxyecdysone septanoside derivative from Atriplex portulacoides roots with preliminary biological activities. Bioorg. Med. Chem. Lett. 2015, 25, 1665–1670, doi:10.1016/J.BMCL.2015.03.028.

[10] Le Merrer, Y.; Poitout, L.; Depezay, J.C.; Dosbaa, I.; Geoffroy, S.; Foglietti, M.J. Synthesis of azasugars as potent inhibitors of glycosidases. Bioorg. Med. Chem. 1997, 5, 519–533, doi:10.1016/S0968-0896(96)00266-0.

[11] Tauss, A.; Steiner, A.J.; Stütz, A.E.; Tarling, C.A.; Withers, S.G.; Wrodnigg, T.M. l-Idoseptanosides: Substrates of D-glucosidases? Tetrahedron Asymmetry 2006, 17, 234–239, doi:10.1016/J.TETASY.2005.12.007.

[12] Elshahawi, S.I.; Shaaban, K.A.; Kharel, M.K.; Thorson, J.S. A comprehensive review of glycosylated bacterial natural products. Chem. Soc. Rev. 2015, 44, 7591–7697, doi:10.1039/C4CS00426D.

[13] Guo, Z.; Tang, Y.; Tang, W.; Chen, Y. Heptose-containing bacterial natural products: Structures, bioactivities, and biosyntheses. Nat. Prod. Rep. 2021, doi:10.1039/d0np00075b.

[14] Tang, W.; Guo, Z.; Cao, Z.; Wang, M.; Li, P.; Meng, X.; Zhao, X.; Xie, Z.; Wang, W.; Zhou, A.; et al. D-Sedoheptulose-7-phosphate is a common precursor for the heptoses of septacidin and hygromycin B. Proc. Natl. Acad. Sci. USA 2018, 115, 2818–2823, doi:10.1073/pnas.1711665115.

[15] Horwitz, J.P.; Chua, J.; Cubby, R.J.; Tomson, A.J.; Da Rooge, M.A.; Fisher, B.E.; Mauricio, J.; Klundt, I. Substrates for Cytochemical Demonstration of Enzyme Activity. I. Some Substituted 3-Indolyl-β-D-glycopyranosides. J. Med. Chem. 1964, 7, 574–575, doi:10.1021/jm00334a044.

[16] Miller, J.H. Experiments in Molecular Genetics, 1st ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 1972.

[17] Kiernan, J. Indigogenic substrates for detection and localization of enzymes. Biotech. Histochem. 2007, 82, 73–103, doi:10.1080/10520290701375278.

[18] Böttcher, S.; Thiem, J. Indoxylic Acid Esters as Convenient Intermediates Towards Indoxyl Glycosides. Eur. J. Org. Chem. 2014, 2014, 564–574, doi:10.1002/ejoc.201301198.

[19] Wolk, J.L.; Frimer, A.A. A simple, safe and efficient synthesis of Tyrian purple (6,6'-Dibromoindigo). Molecules 2010, 15, 5561–5580, doi:10.3390/molecules15085561.

[20] Castro, S.; Duff, M.; Snyder, N.L.; Morton, M.; Kumar, C.V.; Peczuh, M.W. Recognition of septanose carbohydrates by concanavalin A. Org. Biomol. Chem. 2005, 3, 3869–3872, doi:10.1039/b509243d.

[21] Duff, M.R.; Fyvie, W.S.; Markad, S.D.; Frankel, A.E.; Kumar, C.V.; Gascón, J.A.; Peczuh, M.W. Computational and experimental investigations of mono-septanoside binding by Concanavalin A: Correlation of ligand stereochemistry to enthalpies of binding. Org. Biomol. Chem. 2011, 9, 154–164, doi:10.1039/c0ob00425a.

[22] Sager, C.P.; Fiege, B.; Zihlmann, P.; Vannam, R.; Rabbani, S.; Jakob, R.P.; Preston, R.C.; Zalewski, A.; Maier, T.; Peczuh, M.W.; et al. The price of flexibility—a case study on septanoses as pyranose mimetics. Chem. Sci. 2018, 9, 646–654, doi:10.1039/C7SC04289B.

[23] Nagata, S.; Tomida, H.; Iwai-Hirose, H.; Tanaka, H.N.; Ando, H.; Imamura, A.; Ishida, H. Synthesis of a 1,2-: Cis -indoxyl galactoside as a chromogenic glycosidase substrate. RSC Adv. 2019, 9, 28241–28247, doi:10.1039/c9ra05797h.

[24] Böttcher, S.; Thiem, J. Synthesis of indoxyl-glycosides for detection of glycosidase activities. J. Vis. Exp. 2015, 2015, e52442, doi:10.3791/52442.

[25] Böttcher, S.; Hederos, M.; Champion, E.; Dékány, G.; Thiem, J. Novel efficient routes to indoxyl glycosides for monitoring glycosidase activities. Org. Lett. 2013, 15, 3766–3769, doi:10.1021/ol401710a.

[26] Pote, A.R.; Vannam, R.; Richard, A.; Gascón, J.; Peczuh, M.W. Formation of and Glycosylation with Per-O-Acetyl Septanosyl Halides: Rationalizing Complex Reactivity En Route to p-Nitrophenyl Septanosides. Eur. J. Org. Chem. 2018, 2018, 1709–1719, doi:10.1002/ejoc.201800310.

[27] de Melo, J.S.S.; Rondão, R.; Burrows, H.D.; Melo, M.J.; Navaratnam, S.; Edge, R.; Voss, G. Spectral and Photophysical Studies of Substituted Indigo Derivatives in Their Keto Forms. ChemPhysChem 2006, 7, 2303–2311, doi:10.1002/cphc.200600203.

[28] Rajesh, K.; Somasundaram, M.; Saiganesh, R.; Balasubramanian, K.K. Bromination of deactivated aromatics: A simple and efficient method. J. Org. Chem. 2007, 72, 5867–5869, doi:10.1021/jo070477u.

[29] Imming, P.; Imhof, I.; Zentgraf, M. An improved synthetic procedure for 6,6′-dibromoindigo (Tyrian Purple). Synth. Commun. 2001, 31, 3721–3727, doi:10.1081/SCC-100107023.

[30] Vallmitjana, M.; Ferrer-Navarro, M.; Planell, R.; Abel, M.; Ausín, C.; Querol, E.; Planas, A.; Pérez-Pons, J.A. Mechanism of the family 1 β-glucosidase from Streptomyces sp: Catalytic residues and kinetic studies. Biochemistry 2001, 40, 5975–5982, doi:10.1021/bi002947j.

[31] Stroobants, A.; Portetelle, D.; Vandenbol, M. New carbohydrate-active enzymes identified by screening two metagenomic libraries derived from the soil of a winter wheat field. J. Appl. Microbiol. 2014, 117, 1045–1055, doi:10.1111/jam.12597.

[32] Kowada, T.; Maeda, H.; Kikuchi, K. BODIPY-based probes for the fluorescence imaging of biomolecules in living cells. Chem. Soc. Rev. 2015, 44, 4953–4972.

[33] Uppal, T.; Bhupathiraju, N.V.S.D.K.; Vicente, M.G.H. Synthesis and cellular properties of Near-IR BODIPY-PEG and carbohydrate conjugates. Tetrahedron 2013, 69, 4687–4693, doi:10.1016/j.tet.2013.03.082.

[34] Aharoni, A.; Thieme, K.; Chiu, C.P.C.; Buchini, S.; Lairson, L.L.; Chen, H.; Strynadka, N.C.J.; Wakarchuk, W.W.; Withers, S.G. High-throughput screening methodology for the directed evolution of glycosyltransferases. Nat. Methods 2006, 3, 609–614, doi:10.1038/nmeth899.

[35] Liu, B.; Novikova, N.; Simpson, M.C.; Timmer, M.S.M.; Stocker, B.L.; Söhnel, T.; Ware, D.C.; Brothers, P.J. Lighting up sugars: Fluorescent BODIPY-: Gluco -furanose and -septanose conjugates linked by direct B-O-C bonds. Org. Biomol. Chem. 2016, 14, 5205–5209, doi:10.1039/c6ob00726k.

[36] Kolb, H.C.; Finn, M.G.; Sharpless, K.B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. Angew. Chem. Int. Ed. 2001, 40, 2004–2021, doi:10.1002/1521-3773(20010601)40:11<2004::AID-ANIE2004>3.0.CO;2-5.

[37] Green, M.R.; Hughes, H.; Sambrook, J.; MacCallum, P. Molecular Cloning; A Laboratory Manual, 4th ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 2012.

[38] Segel, I.H. Enzyme Kinetics: Behaviour and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems; John Wiley & Sons Ltd.: New York, NY, USA, 1975.

CHAPTER 6. TOWARDS ENGINEERING A β–GLUCOSIDASE TO ACCEPT SEVEN MEMBER RING SEPTANOSIDE

18. CHAPTER 6. Towards engineering a β-glucosidase to accept seven-member ring septanoside substrates

The data show in section 17 suggest that septanose molecule could be a substrate for the β-glycosidase from *Streptomyces sp*, albeit a very poor one, with kinetic parameters similar to some related idoseptanoses, septanoses lacking hydroxymethyl groups reported by Withers, that were also substrates for glycosidases [243]. With the aim to determine the molecular insights of septanose-enzyme interaction, a preliminary manual docking between βgl3 protein and a septanose substrate was performed. Manual docking of the septanoside substrate into the enzyme showed that some amino acid residues could possibly contribute the most in impeding the septanoside substrate from entering the catalytic site of the enzyme. The βgl3 from *Streptomyces sp.* structure (PDB code: 1GON) was aligned with the septanose substrate by threading with the crystallographic structure of β-glucosidase of rice *(Oryza sativa)* in complex with DNP-2FGlu (substrate) (PBD code: 3PTQ). Position of O, C1, C2 and C3 from the glucose molecule were overlapped with the same septanose atoms, leading to a proper carbon alpha (C1) orientation with the catalytic residues. Other atoms C4, C5, C6 and C7 from septanose were dealineated compared with the glucose molecule (Figure 18.1).

The structure obtained between βgl3 protein and septanose substrate was visualized and analysed (Figure 18.1). It should be stressed that there is a clear steric clash between septanose and W430 residue that apparently prevents the binding of the septanose in the active site. Other amino acids seem to be in position to interact with hydrogen bonds (Q32, H133, N177 and E437) while some may interact hydrophobically (W134 and Y309). Residues W438, F446, W356, N307, M354, C181 and N307 were also considered for proximity to substrate even doesn't interact with the septanose in this rough docking. Manual docking suggest that first shell amino acids in close contact with the substrate will play a role in accepting seven member rings in the active site.

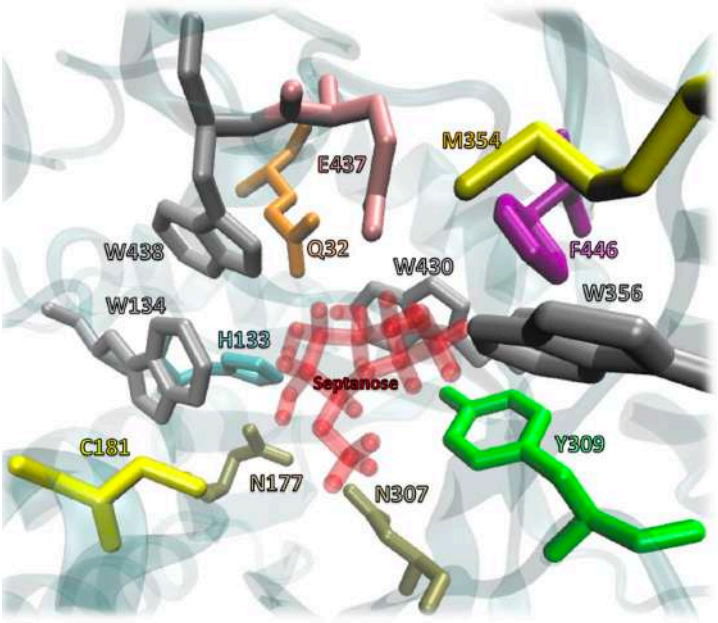


Figure 18.1. Set of mutants list highlighted in the manual docking structure between βgl3 and septanose substrate.

18.1.1. Alanine scanning mutagenesis

We sought to determine whether a small change in first shell amino acid residues allows us to create more free space in the active site to properly fit the septanose molecule.

The first experimental step of the project was an Alanine scanning mutagenesis over all amino acids nearby the catalytic center. Mutating all residues to Alanine would provide two types of information. In one hand, multiple candidates with a disturbed catalytic domain that may show septanose activity will be screened. On the other hand, checking glucosidase activity towards natural substrates will provide information about the role of every residue in enzyme activity. With the initial information, W430A seems a specially promising candidate due to the steric clash shown by the manual docking. Different positions were selected to perform the alanine scanning mutagenesis (Figure 18.1).

Multiple QuikChange® PCR were done with forward and reverse oligos containing the desired mutation (see protocol 21.3.3). Unfortunately, not all mutant list was accomplished, some mutants still depending (Figure 18.1). For future work could be interesting to obtain the remaining alanine mutants (H133A, Y309A, F446A, C181A and N307A) to completed interactions map information.

Protein expression was done using IPTG induction (see protocol 21.4.2), thereupon protein purification was done using His-tag column according protocol 21.4.5. Afterwards, kinetic assays were done with Glc-pNP and Gal-pNP, natural substrates for Bgl3 enzyme (see protocol 21.6). The objective was to determine how did mutations affect the catalytic activity compared to WT enzyme. Each mutant was also tested with Sept-pNP (Figure 15.5.B) to check if any enzyme shows hydrolytic activity towards septanose. Kinetic assays were done at 50ºC using phosphate buffer 50 mM pH 6.5 as a buffer.

Table 18.1. Specific activity and percentage of remaining activity for every mutant with Glc-pNP and Gal-pNP. Kinetics assays were done using PB 50 mM pH 7.2 NaCl 0.3M as a buffer.

	Specific Activity (s^{-1})				
	Glc-pNP	**WT ratio**	**Gal-pNP**	**WT ratio**	**Sept-pNP**
WT	4.97 ± 0.76	100%	9.31 ± 1.03	100%	None
Steric clash					
W430A	0.5 ± 0.1	10%	0.27 ± 0.05	0.02%	None
Orientation to hydroxyls					
Q32A	0.0093 ± 0.0021	0.2%	0.020 ± 0.003	0.2%	None
N177A	0.16 ± 0.02	3.2%	0.016 ± 0.001	0.2%	None
E437A	0.009 ± 0.001	0.2%	0.018 ± 0.003	0.2%	None
Hydrophobic Stacking					
W134A	0.43 ± 0.05	8.65%	0.11 ± 0.02	2.2%	None
Others					
M354A	1.29 ± 0.21	32	0.54 ± 0.13	25%	None
W356A	3.64 ± 0.12	73 %	3.64 ± 0.32	39%	None
W438A	0.04 ± 0.004	0.80%	0.004 ± 0.001	0.04%	None

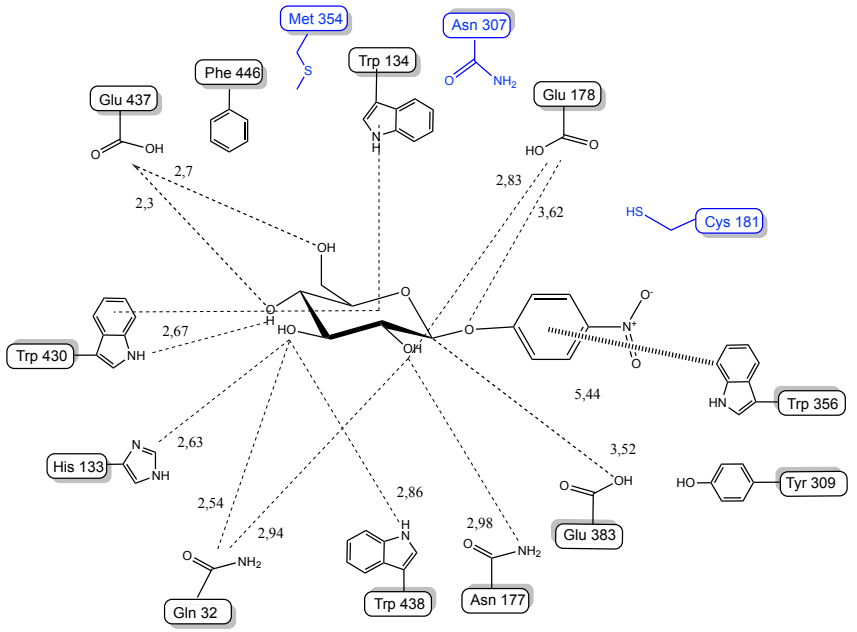


Figure 18.2. Interaction map between WT βgl3 and Glc-pNP. Catalytic residues (E178, E383). In blue, second shell residues.

Specific activity of mutant and the WT-Glu-pNP interaction map were analysed (Table 18.1 and Figure 18.2). The most active mutant was M354A with 32% activity relative to wt towards Glc-pNP. Met354 residue interacts just by Van der Waals forces with the substrate, so Met354 absence doesn't seem to highly affect the catalytic pocket interaction with the substrate. The second most active mutant was W430A with 10% remaining activity towards Glc-pNP. Trp430 didn't show any direct interactions with the substrate, it just define the pocket size. Position W430 clearly prevents the binding of the septanose molecule in the active site by steric clash. For all mutants with 3% or lower residual activity towards Glc-pNP (E437, W134A, Q32A, W438 and N177A), the respective residues interact directly with the substrate with one or two hydrogen bonds expect Trp134 which do not interact directly with the substrate but seems to have an important role. It should be stressed that there is no mutant active on Sept-pNP substrate.

Taken altogether, the results suggested that the septanose molecule occupies regions in the binding pocket already occupied by aa sidechains of the protein. That is, there are several unfavorable steric interactions with sidechains, but not the main chain of the protein. Of all the interactions, we considered W430 the best target for creating a larger volume at the active site to accommodate the septanose substrates

18.1.2. Kinetic parameters a for wt and W430A mutant

No activity towards septanose-pNP was obtained for any mutant from the Alanine scanning mutagenesis approach. According to preliminary docking studies, W430 residue seemed to have a steric clash with the septanose substrate. So, W430A mutant change the voluminous Trp for a tiny Ala, tentatively getting enough space for septanose substrate to fit in and get some catalytic activity. Michaelis Menten assays were performed for both enzymes with Glc-pNP as natural substrate to evaluate the impact of W430A mutation (see protocol 21.6).

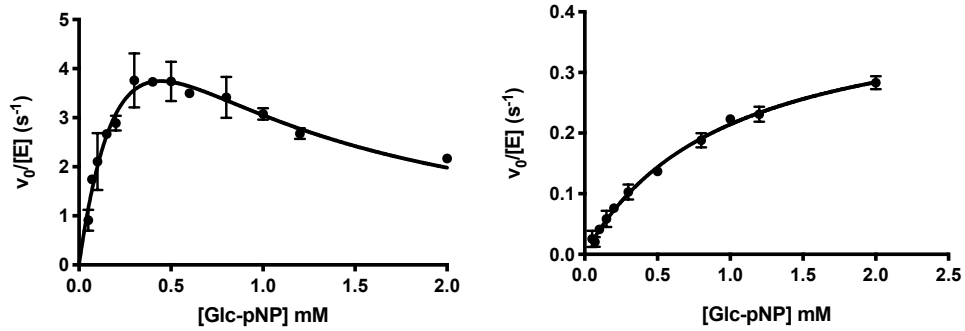


Figure 18.3. WT and W430A with Glc-pNP Michaelis Menten plots. Kinetics assays were done using PB 50 mM pH 7.2 NaCl 0.3M as a buffer.

Table 18.2. Michaelis Menten parameters for WT and W430A enzymes with Glc-pNP as substrate.

Mutant	k_{cat} (s^{-1})	K_M (mM)	K_i (mM)	k_{cat}/K_M ($M^{-1} \cdot s^{-1}$)
wt	10.8 ± 18	0.41 ± 0.2	0.47 ± 0.2	$2.6 \cdot 10^4$
W430A	0.42 ± 0.02	0.96 ± 0.08		$4.4 \cdot 10^2$

Mutant W430A shows almost 26-fold lower k_{cat} value than the wt enzyme (Figure 18.3) showing a high impact on enzyme activity towards the natural substrate pNP-Glc. On the other hand, the K_M value is only 2-fold higher than the wt enzyme. This value suggests that natural substrate pNP-Glc still bind in the active site with certain affinity but the mutant is unable to hydrolyze the substrate. In addition, mutant W430A completely losses the substrate inhibition behavior.

It should be stressed that W430 mutation does not completely disturb the catalytic center because it is still able to bind the natural substrate but the catalytic efficiently was largely affected. Results suggested that a directed evolution approach involving different mutations in the first shell residues could result in a mutant able to hydrolase septanose-pNP molecule.

Further modelling analysis using docking was done to elucidate the role of W430A mutation in the acceptance of septanose substrates.

18.1.3. In silico analysis of W430A mutant

In unpublished work, we sought to better ascertain the reasons for the low hydrolytic activity on pNP β-2 (Figure 15.5) through a computational docking experiment between methyl septanoside Me β-2 and Me β-3 and *Streptomyces sp.* Bgl3 (pdb code 1GNX).

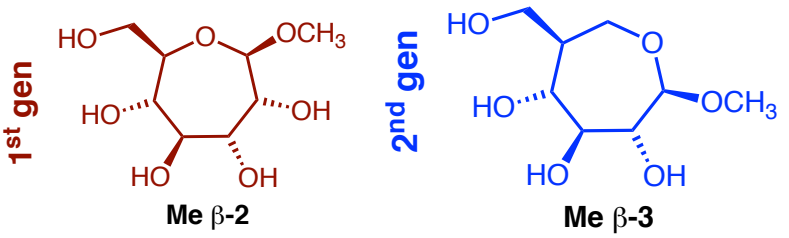


Figure 18.4. Chemical representation of 1st generation of septanoside Me β-2 and 2nd generation Me β-3.

We redirected our efforts to an in silico analysis of βgl3 association with potential septanose substrates, using 1st (Me β-2) and 2nd (Me β-3) generation septanosides to understand the molecular binding of septanoside molecules in βGl3 active site.

First, dockings with the septanose substrate Me β-2 were performed with WT protein and W430A mutant. βgl3 protein from *Streptomyces sp.* was already crystallized with PDB code: 1GON. This protein structure lacks two external high mobility loops (from 1 to 14 and from 313 to 332 residues). HHpred program was used to predict the 3D structure of both loops using BLAST and PSI-BLAST searches that find remote homologs. Even if the prediction was not accurate, these loops are far away from the catalytic pocket so probably would not affect docking results. However, a complete protein structure was needed to run adequately the docking programs. For all dockings in subsequent work, the same grid box was selected. The box contained all the catalytic pocket (including both catalytic residues) together with 3 some extra space that ensures docking reliability. Box distances were 60 x 58 x 60 Å.

A septanose Me β-2 ligand structure PDB format was available in IQS Bioinformatic Databank. First, WT docking with septanose ligand 1st (Me β-2) was performed, the result revealed what was expected, septanose molecule was too big to access the catalytic pocket (Figure 18.5).

Then, W430A mutant was obtained, just by Trp430 side-chain amino acid deletion form the previous protein structure. Considering the hypothesis of W430A mutant may leave some space in the catalytic pocket to septanose 1st (Me β-2) fit in, W430A mutant docking with septanose was performed and results shown that septanose can enter the catalytic pocket and interact like natural substrates (Figure 18.5.). Despite of the septanose fit in the catalytic pocket, the reaction didn't take place because of the septanose molecule orientation takes. Septanose is flipped compared to the expected catalytic position, that would explain why W430A doesn't have any catalytic activity towards septanose. Analysing in detail, seems that septanose C4 and C5 are strongly bounded with the residues which normally interact with Glc-pNP C3 and C4 leaving the alpha carbon out of orientation (Figure 18.6.).

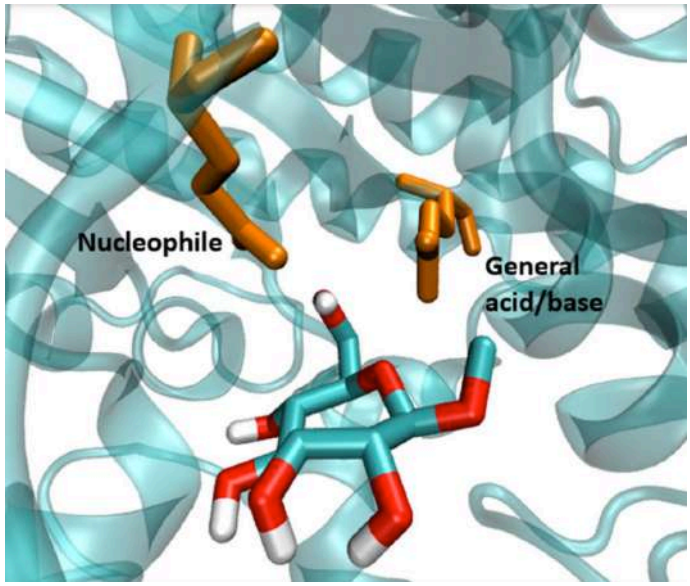


Figure 18.5. Snapshot of docking of W430A mutant using septanose 1st (Me β-2) as a substrate.

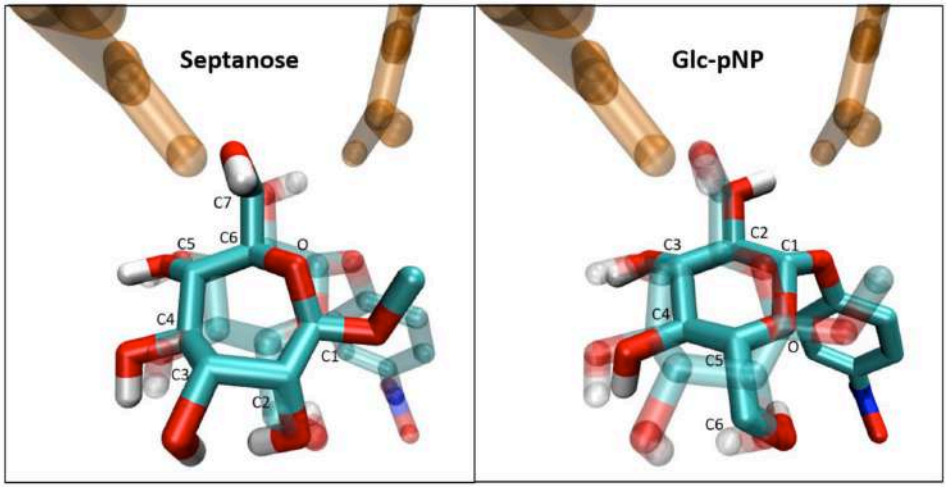


Figure 18.6. W430A Septanose docking over WT Glc-pNP. Superposition is repeated highlighting one and another, in each case, for the opaque (not transparent) molecule carbons are assigned.

We sought to rigorously determine why 1st generation septanosides were not substrates and determine how alternative designs like Me β-3 would bind into the donor site of wt βgl3. Septanose Me β-3 was designed according the results obtained in docking Figure 18.6. Me β-3 contains a septanoside ring but septanosidic oxygen was move back one position to mimic the flipped molecule obtained in Figure 18.6 to produce a structure able to bind in a catalytic orientation in the active site.

To better understand why septanosides were poor substrates of wt βgl3 and to provide preliminary support for new mutants or substrate design, we conducted docking experiments with methyl septanoside Me β-3 wt Bgl3 and two single mutants: W430A and W430L of βgl3 protein.

On the other hand, first of all docking required the generation of low energy conformers of Me β-3. Because they are highly populated, low energy conformers are the ones most likely to encounter the enzyme. An earlier conformational analysis on the parent methyl septanoside Me β-2 calculated 3J H,H coupling constants based on a Boltzmann distribution of conformers, and compared them to values collected in ^{1}H-NMR experiments. Because Me β-3 has yet to be synthesized, we relied solely on calculated conformations for the preliminary docking study here (Figure 18.7).

Conformers were generated in the following manner:

1) A configurationally accurate structure of Me β-3 was imported into Maestro (Schrödinger).
2) The structure was optimized by DFT at a B3LYP/6-31G** level of theory. 3
3) A standard, unconstrained Monte Carlo conformational search routine was then conducted on the optimized structure using the OPLS force-field and water as an implicit solvent. Conformations whose RMSD values were within 0.5 Å were considered identical and the newly generated structure was discarded.
4) After ~2500 iterations, low energy conformers were re-optimized.
5) In addition, organized into a list based on their relative free energies.
6) Structures within approximately 5-10 kcal/mol of the global minimum were inspected and manually categorized based on Stoddart's nomenclature (Figure 18.7).

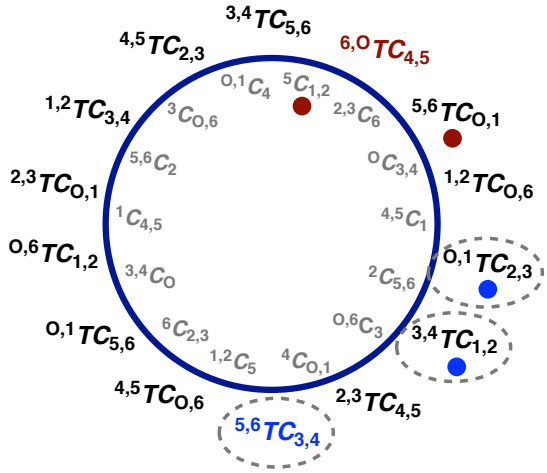


Figure 18.7. Conformational map for septanosides showing lowest energy conformers of Me β-2 (brown) and Me β-3 (blue); conformers within 3-5 kcal/mol are labelled with dots.

With low-energy conformations of Me β-3 in hand (Figure 18.7), we proceeded to docking experiments with three βgl3 sequences: wild type (wt), W430A, and W430L previously prepared form 1GON model. Manual docking identified W430 as a key residue to mutate to create a larger volume in the active site so that septanosides could be accommodated as substrates. AutoDock uses a computationally inexpensive hybrid force field that contains terms based on molecular mechanics as well as empirical terms. Its semi-empirical force field includes intramolecular terms, a "full" desolvation model, and also considers directionality in hydrogen bonds. Conformational entropy is calculated from the sum of the torsional degrees of freedom. A clustering analysis was done for every docking experiment in order to divide

docking poses into structurally homogenous groups to evaluate them. In this approach, every pose is put in a given group by using a measure of structural similarity and compared with the energy of binding obtained for every docking conformation. Root mean square deviation (RMSD) values obtained by pairwise or matrix error distances are the most widely used measure of similarity for binning structures. It permitted us to correlate the expected binding pose with the ones obtained by molecular docking.

Results from the docking experiments are summarized in Table 18.3. It shows the matrix of ligands and proteins that were explored. The most highly populated binding site on wt Bgl3 for all three compounds was the acceptor site where the aglycon would bind (++). There were instances of binding at the donor site of the active site of the wt enzyme, where the scissile glycosidic linkage is properly accommodated, but clusters were minor in comparison to the acceptor site. This observation is consistent with our earlier conclusion that, for septanoses, the W430 residue blocked the active site, precluding their ability to bind septanosides in a productive mode/pose for catalysis.

Table 18.3. Summary of Docking experiments for septanose ligands with wt βgl3 and single mutants W430A and W430L. Key: + = populated, but low counts; ++ = highly populated cluster; flip = ligand orientation is flipped relative to pyranose; ⊥ = ligand is bound perpendicular to the productive pose orientation; W-hole = ligand occupies space created by removing tryptophan residue.

	wt			W430A			W430L		
	sites			sites			sites		
substrate	active	aglyc	others	active	aglyc	others	active	aglyc	acceptor
Me β-2	+ (flip) + (⊥)	++	not obs	+ (flip) + (substr)	+	+W-hole	+ (substr)	+	+W-hole
Me β-3	+ (flip)	++	not obs	++ (substr)	+	+W-hole	++ (substr)	+	+W-hole

Docking experiments with the W430A mutant showed a different pattern of binding to Me β-2 and Me β-3. Clusters showing each ligand at the acceptor site were populated, as they were for wt (Figure 18.8).

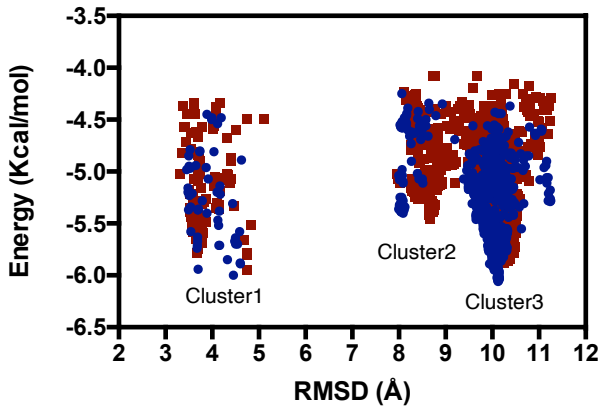


Figure 18.8.) Docking clusters from first ($^{5,6}TC_{3,4}$ - blue) and second ($^{3,4}TC_{1,2}$ – red) Me β-3 conformers with βgl3 W430A. Cluster 1: "productive" binding at donor subsite; Cluster 2: binding into the W-hole; Cluster 3: binding into the acceptor pocket.

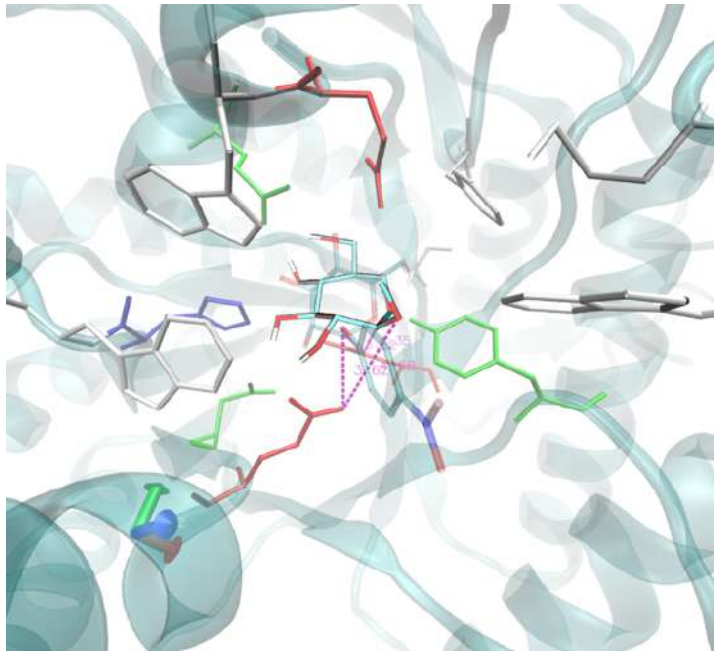


Figure 18.9. Active site of βgl W430A with Me β-3 (blue, cluster 1) and pNP β-1 (light grey).

Of greater importance were the binding modes observed at the donor site. As already stated, Me β-2 drifted into the space opened up by converting the Trp residue to an Ala (the "W-hole"). Me β-2 also binds at the active site in a flipped configuration as it did with the wt protein. Critically, Me β-2 and Me β-3 populated clusters at the active site were reminiscent of the pyranose substrate in a productive binding mode; for Me β-2 the population of this cluster was

low. For Me β-3, however, it was significantly populated. Inspection of specific poses in that cluster showed that Me β-3 is in the same plane that would be occupied by the pyranose substrate, and the acidic residue (E383A) of the enzyme is in proximity to the anomeric carbon (Figure 18.9). This close proximity is important for the progress of the hydrolysis mechanism. In further support of this result, a second conformer of Me β-3 - a $^{3,4}TC_{1,2}$ conformer (the first was a $^{5,6}TC_{3,4}$ conformer) was subjected to the same docking routine. A pattern of poses that was very similar to the original run was obtained, reinforcing the robustness of the result. Similar results were obtained with the W430L mutant: Me β-2 essentially did not bind into the donor subsites in a productive mode but Me β-3 showed a preferential productive binding. We take this as a clear indication of the potential success of 2nd generation of septanoses to be substrates of βgl3, and the enzyme amenable to optimization by mutagenesis.

18.2. Closing remarks towards engineering a β-glucosidase to accept seven-member ring septanoside substrates

Indolyl septanosides are useful reagents because they enable the colorimetric characterization of glycoside hydrolysis in a manner analogous to X-Gal. Preliminary assays in *E. coli* to characterize the reactivity of indolyl D-glycero-D-gulo-septanoside as a substrate of β-galactosidase showed no hydrolysis. The apparent lack of reactivity of this septanoside substrate was subsequently explored along two different lines. First, experiments with a panel of commercially available glycosidases also showed no significant hydrolysis of the unnatural substrates septanosides. Significantly, βgl3 (β-glucosidase from *Streptomyces sp.*) showed a suggestive activity with light-pale blue colonies when using substrate indolyl septanoside. Further analysis with p-nitrophenyl septanoside confirmed the low activity, in the range of 5–6 orders of magnitude lower than its p-nitrophenyl D-glucoside substrate. In a second line of inquiry, BODIPY-labeled septanoside was observed inside *E. coli* cells by confocal microscopy. It suggested that septanosides traverse the bacterial envelope. Taken together, these results illustrate a further directed evolution approach using βgl3 as a template.

To evaluate the faint hydrolytic activity of the β-glucosidase from *Streptomyces sp.* using septanoside substrates, septanoside Me β-2 (Figure 18.4) was manually docked into the active site to assess its interactions with the protein; the substrate was oriented so that its anomeric carbon (site of hydrolysis) was near the catalytic residues of βgl3. The results suggested that it occupies regions in the binding pocket already occupied by aa sidechains

of the protein. That is, there are several unfavorable steric interactions with sidechains, but not the main chain of the protein. Alanine scanning mutagenesis was done in the first shelf of the active site in βgl3 to evaluate the role of amino acids in close contact with the septanoside substrate. No mutant show hydrolytic activity on septanose-pNP. Of all the interactions, we considered W430 the best target for creating a larger volume at the active site to accommodate the septanose substrates. Using specific activity as a metric, β-glucosyl pNP (pNP β-1), β-galactosyl pNP (pNP β-4), and pNP β-septanoside pNP β-2 were evaluated as substrates for βgl3 W430A (Table 18.1), unfortunately, showed a slightly activity with septanose-pNP β-2 as substrate.

Furthermore, different docking approaches were done to evaluate the possible binding of the septanose in the active site in W430A. Despite of the septanose fit in the catalytic pocket, the reaction didn't take place because of the septanose molecule orientation takes. Septanose is flipped compared to the expected catalytic position, that would explain why W430A doesn't have any catalytic activity towards septanose. Alternative designs like Me β-3 (Figure 18.4) could bind into the donor site of wt βgl3. New septanoside molecule Me β-3 was analyzed by docking experiments with two single mutants: W430A and W430L of βgl3 protein. Of greater importance were the binding modes observed at the donor site. Me β-2 drifted into the space opened up by converting the Trp residue to an Ala (the "W-hole"). Me β-2 also binds at the active site in a flipped configuration as it did with the wt protein but Me β-3 is in the same plane that would be occupied by the pyranose substrate, and the acidic residue (E383A) of the enzyme is in proximity to the anomeric carbon rendering a productive binding mode. We take this as a clear indication of the potential success of 2nd generation of septanoses to be substrates of βgl3, and amenable to optimization by mutagenesis.

CONCLUSIONS

19. CONCLUSIONS

The main conclusions drawn from the work carried out in the two distinct topics of this doctoral thesis are detailed below.

19.1. Conclusions from engineering substrate specificity of Chitin deacetylases (Chapters 1 to 4)

- We have implemented a convenient and sensitive HTS assay to screen directed evolution libraries of CDAs and other CE4 enzymes for deacetylase activity on COS. It is based on a fluorescence monitoring assay of the deacetylase activity on COS substrates after capturing the expressed enzyme variants fused to a chitin binding module with chitin-coated magnetic beads. The same assay principle would be applicable to the screening of inhibitors of CE4 enzymes (including chitin and peptidoglycan deacetylases) as they are antimicrobial targets.
- The HTS assay was applied to screen mutant libraries of the model *Vibrio cholera* enzyme in the framework of a directed evolution program towards new substrate specificities. Two mutants were identified, K275E and K275E/H127R that increases the activity (k_{cat}) towards long-substrates such DP4. Mutant K275E suggests the role of loop 5 flexibility in the acceptance of long substrates.
- An internal disulfide bond P193C/N273C that anchors loop 5 with loop 3 limits the movement of this loop 5 and strongly reduces the activity towards long substrates.
- Triple proline to glycine mutant in loop 5 completely reverts the activity of the enzyme towards long substrates, increasing the k_{cat} values almost 600% compared to the wt enzyme at expenses the substrate affinity (K_M).
- Individual proline to glycine mutants were analyzed to determine the role of each proline in the final construct. Results identify P266 and P271 as key contributors to the final increase in enzymatic activity of the triple PG mutant towards long substrates, whereas P280 almost has no effect in the final PG mutant.
- Methyl TROSY experiments conclude that the wt enzyme has two conformations in slow equilibrium and that the minor conformation (about 20% population) is close to the bound enzyme·substrate complex, whereas the PG mutant has a single conformation (equivalent to the major conformation of the wt enzyme) that shifts to the equivalent minor conformation of the wt enzyme upon substrate binding. This

would suggest that the minor species in the wt enzyme is an intermediate conformation in the enzyme·substrate binding reaction coordinate that explains the lower values of K_M in wt enzyme in comparation with PG mutant.

- Relaxation dispersion experiments reveal that a significant fraction of the methyl set undergoes motions in the microsecond timescale. Such residues always belong to the major species and account for about 20% of the methyl probes of this specific conformation. Residues in loop 4 (M242 and I240) were assigned by mutation and show conformational exchange together with other Met and Ile residues located in other loops, all in the same time scale that indicates concerted dynamics.
- The timescale of the fast motion for CD PG mutant (kex = 850 s^{-1}) should be fast enough NOT to interfere with the release of the product and permits a fast catalytic turnover because the substrate has a greater facility to leave the active center once enzymatic catalysis has taken place. On the contrary, the slower fast motion (kex = 120 s-1) for the CD wt enzyme may interfere product release (and substrate binding) explaining the lower k_{cat} value for long substrates as compare to the PG mutant.
- Taken all previous results together, we can conclude that not only the shape of the loops surrounding the active site are key for the structural determinants of substrate specificity, but also their dynamics in different timescale motions are involved in enzyme specificity.
- In contrast to the initial assumption of an induced fit mechanism for ligand binding in the wt *Vc*CDA based of X-ray crystallography, NMR experiments show that *Vc*CDA wt follows, at least partially, a conformational selection mechanism. Moreover, NMR experiments confirms that PG mutant switches the binding mechanism from a conformation selection (in the wt enzyme) to an induce fit mechanism (in the PG mutant).
- A new screening procedure was developed, based on a primary screening for deacetylase activity with methylumbelliferyl acetate substrate followed of a secondary screening with COS by HPLC-MS that allows the analysis of a large library by activity, degree of acetylation and also pattern of acetylation.
- Engineering a novel pattern of acetylation has been attempted by redesigning loops at the negative subsites. The hydrogen-bonding network between loops 1, 2 and 6 is key for protein stability. From a combinatorial library at loop 1 residues, only mutant D74G-E75E-N76A showed the appearance of a new product with low yield

corresponding to a di-deaceetylated COS. At this stage, it is concluded that such an engineering goal for different deacetylation specificity would require a major loops restructuring.

19.2. Conclusions from β-glucosidase from *Streptomyces sp.*

- Indolyl septanosides are useful chromogenic substrates that enable the screening of glycosidases, either collections of wt enzymes or directed evolution libraries, for septanoside hydrolase activity. Currently, such enzyme specificity for septanoside substrates has not been identified in nature.
- After screening a panel of wild-type glycosidases, βgl3 (β-glucosidase from *Streptomyces* sp.) showed a suggestive activity with light-pale blue colonies when using indolyl septanoside and faint activity with p-nitrophenol septanoside substrates, which indicate that septanosides are poor substrates for the glycosidase. We then considered that βgl3 is a good glycosidase candidate to start an engineering program towards a septanoside hydrolase.
- W430 was selected as the best target for creating a larger volume at the active site to accommodate the septanose substrates. However it showed faint activity with indolyl septanoside substrate as with the wt enzyme.
- Alanine scanning mutagenesis was performed on first shelf residues of the active site in βgl3 and no mutant showed hydrolytic activity on septanose-pNP.
- Different docking approaches were done to evaluate the possible binding of the septanose substrate in the active site of W430A. Despite of septanose fitting into the catalytic pocket, the reaction didn't take place due to the unproductive orientation of the septanose ring, which is flipped as compared to the expected catalytic pose.
- Alternative septanose molecule designs like Me β-3 were evaluated by docking experiments and suggested that they could bind productively into the binding site of wt βgl3. Docking experiments provide a clear indication of the potential of 2nd generation septanoses to become the proper substrates for engineering βgl3 as a septanoside hydrolase by directed evolution. The project will continue under these assumptions.

MATERIALS AND METHODS

21. MATERIALS AND METHODS

In this chapter, the experimental part of the project is described. Defined protocols for different chapters were presented in detail. On the other hand, general procedures were described.

21.1. Development of HTS assay for Chitin de-*N*-Acetylases.

21.1.1. Binding efficiency of *Vc*CDA to the CMBs using different conditioning and binding buffers.

Percentage of bound protein to CMBs at increasing purified protein loads. Buffers: A. PBS (50 mM phosphate, 300 mM NaCl, pH 8.5) as conditioning and binding buffer. B. PBS+Triton X-100 (0.05%) as conditioning buffer and PBS as binding buffer. C. CBB (20 mM Tris·HCl, 500 mM NaCl, 1 mM EDTA, 0.05% Triton X-100, pH 8.0) as conditioning buffer, and PBS as binding buffer. D. CBB as both conditioning and binding buffer.

- First, the CMBs were conditioned by taking a volume of commercial CMB suspension (New England Biolabs), settling the beads by centrifugation (or applying a magnet), and washing them with water (2x) and conditioning buffer (3x), followed by resuspension in conditioning buffer (same initial volume).
- 25 μL of CMB suspension were added to 150 μL of purified protein (15-140 μg) in binding buffer, and kept with shaking at 650 rpm for 1h at 4°C.
- After binding, the beads were retained with a magnet and the supernatants transferred to a new tube.
- The protein in the supernatant was quantified by the BCA assay, and the CMB bound protein calculated by difference.

SDS-PAGE gels of bound and unbound (supernatant) protein (0-140 μg *Vc*CDA incubated with 25 μL CMB suspension with PBS pH 8.5 as conditioning and binding buffer at 4°C for 1 h, 175 μL total volume).

- After binding, beads were retained with a magnet and the supernatants transferred to a new tube (S).
- After washing the CMBs with PBS (150 μL, 3 times), both the beads (CMB gel) and the supernatant S (Supernatant gels) were analyzed by SDS-PAGE (see protocol 21.9.1). M: molecular weight marker.

21.1.2. Specific activity of CMB-immobilized *Vc*CDA

Specific activity (SA) of CMB-immobilized VcCDA was determined using tetraacetylchitotetraose (DP4) as substrate.

- After the binding step (60 μg *Vc*CDA incubated with 25 μL CMB suspension with PBS pH 8.5 as conditioning and binding butter at 4°C for 1 h), the CMBs were washed 3 times with PBS pH 8.5 and preincubated at 37°C for 5 min.
- Reactions were initiated by addition of 100 μl DP4 (2 mM).
- At different time intervals, the beads were retained with a magnet, and 10 μl aliquots were withdrawn and added to 90 μl of water/1-propanol (1:1) to stop the reaction and monitor product formation by HPLC-MS (see protocol 21.5).

21.1.3. Specificity of *Vc*CDA binding to the CMBs.

- To study possible interferences 50 μL of purified *Vc*CDA solution containing 0-140 μg of protein (as indicated in the gel lanes) were mixed with 100 μL of a cell-free extract of *E. coli* BL21(DE3) cells harboring an empty pET22b plasmid and the suspension incubated with 25 μL of CMB in PBS buffer at 4°C for 1 h with shaking (650 rpm).
- The beads were retained with a magnet and the supernatants transferred to a new tube. Bound and unbound (supernatant) proteins were analyzed by SDS-PAGE (see protocol 21.9.1). Upper gel: CBM-immobilized proteins, Lower gel: supernatants.

21.1.4. Specificity of *Vc*CDA binding to the CMBs.

- 40 μg of purified *Vc*CDA protein solution were mixed with 0-10 μg/μL of BSA in PBS buffer in total volume of 150 μL. 25 μL of CMB in PBS buffer were added and the suspension incubated at 4°C for 1 h with shaking (650 rpm).
- The beads were retained with a magnet and the supernatants discarded.
- After washing the CMBs with PBS (150 μL, 3 times), the CBM-immobilized proteins were analyzed by SDS-PAGE.

21.1.5. Kinetics of *Vc*CDA binding to the CMBs.

- 25 μL of CMB in PBS pH 8.5 buffer were added to 90 μg of purified VcCDA protein in 150 μl PBS pH 8.5, and the suspension was incubated at 4°C for different time intervals (15 to 90 min) with shaking (650 rpm).
- The beads were retained with a magnet and the supernatants transferred to a new tube (S). After washing the CMBs with PBS (150 μL, 3 times), both the beads (CMB gel) and the supernatant S (Supernatant gel) were analyzed by SDS-PAGE.
- Fig 4.7.A) SDS-PAGE of protein bound and unbound to CMB after different incubation times;
- Fig 4.7.B) quantification of bound protein by quantification of the unbound protein in the supernatant (S) by the BCA assay (see protocol 21.9.2).

21.1.6. Deacetylase activity of free and CMB-immobilized *Vc*CDA using DP2 and DP4 as substrates.

Fig 4.8.A) Activity with diacetylchitobiose (DP2) substrate monitored by HPLC-MS according to protocol 4.

- Fig 4.8. A1) Purified enzyme. Free enzyme (grey circles): 30 μg of purified *Vc*CDA were mixed with DP2 at a final concentration of 2 mM in a final volume of 100 μL, and the solution was incubated at 37°C for 5 min. At different times intervals, aliquots (10 μl) were withdrawn and added to 90 μL of water/1-propanol (1:1) to stop the reaction. Samples were analyzed by HPLC-MS. CMB-immobilized enzyme (black squares): 60 μg of purified *Vc*CDA in PBS pH 8.5 (150 μL final volume) were mixed with 25 μL CMB in PBS pH 8.5 buffer and the suspension was incubated at 4°C for 1 h with shaking (650 rpm). After the binding step (50% protein binding according to Figure S1), the beads were retained with a magnet and the supernatants transferred to a new tube (S). After washing the CMBs with PBS (150 μL, 2 times) and preincubated at 37°C for 5 min, reactions were initiated by addition of 100 μl of DP2 (2 mM). At different time intervals, the beads were retained and aliquots (10 μl) were withdrawn and added to 90 μl of water/1-propanol (1:1) to stop the reaction. Samples were analyzed by HPLC-MS. Inset: Initial rates of free and immobilized enzyme at the same protein concentration.

- Fig 4.8. A2) Cell-free extracts. 150 μl (black cicles) and 200 μl (grey squares) of a cell-free extract containing *Vc*CDA (protocol 1) were mixed with PBS pH 8.5 in a final volume of 200 μL and 25 μL of CMB in PBS pH 8.5 buffer were added and the suspension was incubated at 4ºC for 1 h with shaking (650 rpm). After binding step, the beads were retained with a magnet and the supernatants transferred to a new tube (S). Washing the CMBs with PBS (150 μL, 2 times) and preincubated at 37ºC for 5 min. Reactions were initiated by addition of 100 μl of $(GlcNAc)_2$ (2 mM) to CMB-P complex. At different time intervals, the beads were retained and aliquots (10 μl) were withdrawn and added to 90 ml of water/1-propanol (1:1) to stop the reaction. Samples were analyzed by HPLC-MS (grey squares).

Fig 4.8.B) Activity with tetraacetylchitotetraose (DP4) substrate monitored by HPLC-MS according to protocol 4.

- Fig 4.8. B1) Purified enzyme. Deacetylase activity using DP4 was performed using the same protocol described above (A1) working with 30 μg of purified protein in free form (grey circles) and 60 μg of protein bound to CMB (50% binding according to Figure S1) (black squares).
- Fig 4.8. B2) Cell-free extract. Deacetylase activity using DP4 was performed using the same protocol described above (A2) working with 100 μl of cell free extract.

21.1.7. *E. coli* expression hosts.

E. coli expression hosts, BL21(D3) and BL21(D3)ΔChb, were transformed with a pET22b plasmid containing a synthetic gene coding for VcCDA.26

- Cell cultures of both strains were grown until OD 1.2
- Fig 4.9. A) Protein expression: 0-150 μL of cell-free extract from each strain were mixed with PBS pH 8.5 to a final volume of 150 μL, then 25 μL of CMB in PBS pH 8.5 buffer were added, and the suspensions were incubated at 4°C for 1 h with shaking (650 rpm).
- After the binding step, the beads were retained with a magnet and the supernatants transferred to new tubes (S). The beads were washed with PBS (150 μL, 3 times) and analyzed by SDS-PAGE (Protocol X).

- Fig 4.9. B) Activity of immobilized enzymes: the beads (CMB-immobilized proteins) coming from 75 μL of cell-free extract from both strains as above, were kept at 37°C for 5 min.
- Reactions were initiated by addition of 100 μL DP4 (2 mM). After 60 min, the beads were retained with a magnet, 10 μL of supernatant were withdrawn and added to 90 μL of water/1-propanol (1:1) to stop the reaction, and product formation was quantified by HPLC-MS.

21.1.8. Cell lysis methods on microplates.

Cell lysis in 96-well deep well plates was studied using 3 different protocols: Bugbuster Reagent®, freeze- thaw cycles, and sonication. The lysis efficiency was determined by analyzing the free protein released to the medium.

- Cell cultures were grown in 96-well deep well plates following the autoinduction protocol (Protocol X).
- Cells were harvested by centrifugation at 3500 rpm and the supernatants were discarded.
- a) Bugbuster Reagent®: the cell pellet was resuspended with 250 μl of Bugbuster reagent with shaking at 650 rpm for 20 min at rt. The lysate was centrifuged at 3500 rpm for 1 h and the pellet discarded.
- b) Freeze-thaw cycles: the cell pellet was frozen by submerging the microplate in a dry ice/ethanol bath for 2 min. Then it was thawed by transfer to an ice/water bath for 8 minutes. The cycle was repeated 10 times. The lysate were resupended in 250 μL PBS pH 8.5, centrifuged at 3500 rpm for 1 h, and the pellet discarded.
- c) Sonication: the cell pellet was resuspended in 250 μL PBS pH 8.5 and lysed by sonication during 7 minutes (50% amplitude, 5 s ON, 15 s OFF, 2 mm diameter probe).
- The lysate was centrifuged at 3500 rpm for 1 h and the pellet discarded.
- For the three method, the protein in the clarified cell-free extracts was quantified using the BCA assay (protocol X). All the procedure was done in triplicate.

21.1.9. HTS protocol of mutant libraries

Transformant *E. coli* BL21 (DE3) cells harboring the epPCR library of the VcCDA catalytic domain, and *Vc*CDA WT used as a control, were grown overnight on LB-agar plates supplemented with 100 μg/mL ampicillin at 37°C. Colonies from the epPCR library as well as colonies from *Vc*CDA WT were picked and transferred to 96-well microtiter plates (74 library colonies and 10 wt colonies per plate), grown overnight in LB medium (100 μg/mL ampicillin) at 37°C and 250 rpm, and stored as glycerol stocks. From glycerol stocks, all the steps were performed using an Automated Liquid Handling Bravo Platform from Agilent Technologies.

1. 10 μL of glycerol stocks were used to inoculate 250 μL of LB autoinduction medium in 96 deep well plates. The plates were sealed with breathable foil and incubated at 24°C, 250 rpm for 48h.
2. After protein expression, the cells were harvested by centrifugation at 4°C, 4000 rpm for 30 minutes. The supernatants were discarded and the plates containing the cell pellets were directly used in the next step or were sealed and kept at -20ºC until use.
3. Cells were lysed by the addition of 250 μL of BugBuster® protein extraction reagent to the cell pellet with shaking at 1500 rpm on an Eppendorf MixMate agitator for 30 minutes at room temperature. Afterwards, the plates were centrifuged at 4°C, 4000 rpm for 60 minutes to remove the cell debris.
4. 25 μL of the supernatant (cell-free extract) were transferred to a new plate and mixed with 125 μL of PBS pH 8.5. Then, 10 μL of a suspension of CMB in PBS pH 8.5 were added (the CBMs had been previously conditioned by washes with water (2x), PBS pH 8.5 (3x), and resuspended in the same initial volume).
5. The 96-well microtiter plates were incubated on an shaker (Eppendorf MixMate) at 650 rpm, 4°C for 1 h to allow protein binding to the CMB. Washes of the beads (protein immobilized on the CMB) were performed using a magnet platform. Supernatants were discarded, and the beads were washed twice with 150 μL PBS buffer pH 8.5, and the final supernatant discarded.
6. The reaction plate with CMB-immobilized proteins and the containers with PBS buffer and substrate solutions were preheated 10 min to the corresponding reaction temperature. Then, 35 μL of PBS 2X pH 8.5 and 35 μL of tetraacetylchitotetraose (4 mM DP4) in MiliQ water were added to the reaction plate to give a final 2 mM substrate concentration. Each plate contained a set of glucosamine standards. The

reaction plate was placed on an Eppendorf MixMate at the reaction temperature, and shaken at 650 rpm for 1 h.

7. The deacetylase activity was quantified using fluorescamine as reagent to detect the formation of primary amines. The deacetylase reactions were stopped by transferring 40 μL of reaction supernatants from the reaction plate (applying the magnet to settle the beads) to a black 96-well microtiter plate (for fluorescence reading) containing 60 μL of PBS pH 8.5.
8. Next, 20 μl of 2 mg/mL fluorescamine in dimethyl-formamide (DMF) were added and the mixture was incubated for 10 minutes at room temperature, followed by addition of 150 μL H_2O:DMF (1:1). Fluorescence was measured on a Biotek FLx800 microplate fluorescence reader at 360 ± 20 nm excitation and 460 ± 20 nm emission wavelengths. Data were retrieved on an EXCEL file for calculations with an in-house data management application.

For assay reported in Fig 4.11.A, different times of deacetylation reaction (5, 30 and 60 minutes) were analyzed using the whole previous protocol.

For assay reported in Fig 4.11.B, different volumes of cell extract (5, 10, 25, 50 and 75 μl) were studied using the whole previous protocol.

For assay reported in Fig 4.11.C, different volumes of CMB (1, 2, 5, 10 and 15 μl) were analyzed using the previous protocol. 25 μL of the supernatant (cell-free extract) was used to study the volume of CMB in reaction.

21.1.10. Reproducibility of the HTS assay.

- The HTS protocol was applied to six glycerinates of *E.coli* BL21(DE) cells harboring the pET22b (wt *Vc*CDA) expression plasmid (see protocol 21.4.1).
- After cell lysis, the cell-free extracts from each culture were split in six aliquots of 25 μL, and the protocol was continued for each of the 6x6 samples.
- In the graph, the concentration of DP4(N) product formed is given for each of the 6 aliquots (repetitions) from each of the 6 cultures (X-axis). The mean values (of 6 repetition) and coefficient of variation (CV%) for each culture are given.

21.1.11. DNA libraries by error-prone PCR.

A library of random mutants of VcCDA was generated using error-prone PCR. First, random mutagenesis on the catalytic domain (CD) was carried out using the GeneMorph II Random Mutagenesis Kit (Agilent Technologies).

Primers were designed to flank the catalytic domain in order to randomize all catalytic domain.

Table 21.1. Flanking primers designed for error prone PCR

Primer	Cloning primer	DNA sequence
Insert vector	Fw	5'- CGCGAAATTAATACGACTCACTATAGGG-3'
	Rv	5'-CACACTCCACGCGGCGTATA-3'.

Table 21.2. PCR reaction mixture

Reagent	Volume (µL)
H_2O MiliQ	41.5
10× Mutazyme II reaction buffer	5
40 mM dNTP mix (200 µM each final)	1
primer mix (250 ng/µl of each primer)	0.5
Mutazyme II DNA polymerase (2.5 U/µl)	1
Template	1

Table 21.3. PCR reaction conditions

STEP	TEMPERATURE (ºC)	TIME	CYCLES
Initial denaturation	95	2 min	1
Denaturation	95	1	30X
Annealing	Tm – 5ºC	1	
Extension	72	1.5	
Final extension	72	10	1

The mutation frequency was controlled by the initial concentration of target DNA according to the manufacture`s protocol. The library string was subcloned into a PCR-amplified pET22b containing the two CBM12 sequences using the CPEC technology.

E. coli BL21 (DE3) cells were transformed with the plasmid library by electroporation (Eppendorf Electroporator®). Protocol

21.1.12. Temperature profile of CMB-immobilized wt *Vc*CDA.

The temperature profile of CMB-immobilized *Vc*CDA was studied between 37 and 70°C.

- 25 μL of cell-free extract were mixed with PBS pH 8.5 to a final volume of 150 μL and 10 μL of CMB were added.
- The suspension was incubated at 4°C for 1 h with shaking (650 rpm). After the binding step, the beads were retained with a magnet and the supernatants discarded.
- The beads were washed with PBS (150 μL, 3 times), resuspended 150 μL PBS buffer pH 8.5, and preincubated at the desired temperature (37-70°C) for 5 min.
- Reactions were initiated by addition of 70 μl DP4 substrate (2 mM) and the reaction mixtures kept at the corresponding temperature. After 60 min reaction time, the deacetylase reactions were stopped by transferring 40 μL of supernatants from the reaction plate (applying the magnet to settle the beads) to a black 96-well microtiter plate (for fluorescence reading) containing 60 μL of PBS pH 8.5.
- Next, 20 μL of 2 mg/mL fluorescamine in DMF were added and the mixture incubated for 10 min at room temperature, followed by addition of 150 μL H_2O:DMF (1:1). Fluorescence was measured on a Biotek FLx800 microplate fluorescence reader at 360±20 nm excitation and 460±20 nm emission wavelengths.

21.1.13. Thermal stability of wt and mutant VcCDA proteins

Melting temperatures of *Vc*CDA WT, K275E, and K275E/H127R were determined with the commercial dye Sypro Orange (Thermo Fisher), which is able to bind to the exposed hydrophobic areas of the protein during denaturation with the concomitant increase of dye fluorescence.

- A stock dilution from the Sypro Orange commercial solution (5000x) to a final concentration of 10X was prepared in the same buffer in which the protein was dissolved.
- 5 μL of Sypro Orange stock solution were added to 20 μl of protein solution (4 μM final concentration) in 50 mM phosphate buffer pH 8.5, 300 mM NaCl.
- A thermocycler Rotogene 3000 (Corbett Research) was used to quantify the fluorescence. Denaturing ramp program was: 1 min at 25°C, steps of 1°C up to 95°C maintaining the temperature for 30 s between steps.
- The excitation/emission wavelength were 483 nm and 560 nm.
- Data were fitted to equation 2 by non-linear regression using the GraphPad software (Prism) (Boltzman Sigmoidal equation, see Equation 17.2).

21.1.14. Thermotolerance of free and immobilized enzymes.

The thermotolerance of *Vc*CDA WT (green dots), K275E (blue squares), and K275-H127R (red diamonds) was determined with free enzymes at 50°C, and with CMB-immobilized enzymes at 50ºC and 60ºC.

- In Purified enzymes (2 mM) were incubated in PBS pH 8.5 buffer at 50°C. Then, aliquots were taken at different time intervals, diluted 1:10 in PBS pH 8.5, and kept at 37°C for 10 min (final enzyme concentration 0.2 μM).
- Residual activity was determined using DP4 substrate at 2 mM concentration (protocol).
- A series of tubes containing 150 μL of purified enzymes (2 μM) and 10 μL of CMB in PBS buffer pH 8.5 were incubated at 4°C for 1 h with shaking (650 rpm). After the binding step, the beads were retained with a magnet, the supernatants discarded, and the beads washed with PBS (150 μL, 2 times). The CMB-immobilized enzymes were resuspended in 150 μL PBS pH 8.5 and half of the tubes were incubated at 50° (B1) and the other half at 60°C (B2).
- At different time intervals (X-axes in the graphics), each tube was cooled to 37°C and maintained for 10 min before the substrate (2mM DP4) was added to start the reactions to measure the residual activity.
- At different time intervals, the beads were retained and aliquots (10 μl) were withdrawn and added to 90 μl of water/1-propanol (1:1) to stop the reaction. Samples were analyzed by HPLC-MS (Protocol 4).

Final curves were adjusted by one phase exponential decay, were Ki was defined the rate constant, expressed in reciprocal of the X axis time units and Half-life as the time units of the X axis computed as ln(2)/K.

Equation 21.1. One phase exponential decay equation.

$$Y=(Y0 - Plateau)*exp(-K*X) + Plateau$$

21.2. Screening assay to evaluate the deacetylation pattern

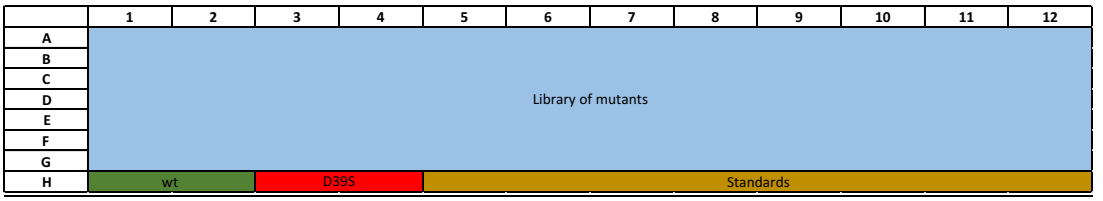
21.2.1. Rapid screening assay for large libraries of chitin deacetylases

Transformant *E. coli* BL21 (DE3) cells harboring the directed evolution library of the *Vc*CDA, and *Vc*CDA WT and *Vc*CDA D39S used as a controls, were grown overnight on LB-agar plates supplemented with 100 μg/mL ampicilin at 37°C. Colonies from the library as well as colonies from *Vc*CDA WT and *Vc*CDA D39S were picked and transferred to 96-well microtiter plates (84 library colonies and 2 wt and 2 D39S colonies per 96 well-plate, distribution was presented below), grown overnight in 1.5 ml LB medium (100 μg/mL ampicillin) at 37°C and 250 rpm, and stored as glycerol stocks. From glycerol stocks, all the steps were performed using an Automated Liquid Handling Platform Bravo from Agilent Technologies.

1) 10 μl of glycerol stocks were used to inoculate 1.5 mL of LB autoinduction medium in 96 deep well plates. The plates were sealed with breathable foil and incubated at 24°C, 250 rpm for 48h.
2) After protein expression, the cells were harvested by centrifugation at 4°C, 4000 rpm for 30 minutes. The supernatants were discarded and the plates containing the cell pellets were directly used in the next step or were sealed and kept at -20ºC until use.
3) Cells were lysed by the addition of 250 μl of Lysis buffer (PBS pH 8.5, 0.5 mg/mL lysozyme, 10U/ml of DNAse) to the cell pellet with resuspension by up and down pipetting at room temperature. After resuspension the pellet, the solution was incubated 1 hour at 30ºC with agitation. To complete the lysis, a frezze/thaw cycle was done with the freezing of the plate at -80ºC for 4 hours and thawing with 2 hours at rt. Afterwards, the plates were centrifuged at 4°C, 4000 rpm for 60 minutes to remove the cell debris.
4) 150 μl of the supernatant (cell-free extract) were transferred to new plate. The plates containing the cell-free extract were directly used in the next step or were sealed and kept at 4ºC until use.
5) 20 μl of the supernatant (cell-free extract) were transferred to a new plate (fluorescent reader plate, black plate) and mixed with 170 μl of PBS pH 8.5 (this buffer includes H_2O_2 10 mM or DTT 10 mM in Figure 14.9). The mixture was incubated at 37ºC for 10 minutes.

6) 10 μl of AcOMU (1 mM in acetonitrile) were added to cell extract solution to reach a final concentration of 0.05 mM of AcOMU. Fluorescence was measured on a FLx800 micro-plate fluorescence reader at 360±20 nm excitation and 460±20 nm emission wave-lengths at 5 min reaction.

Microplate distribution:



21.2.2. Secondary screening assay to evaluate the deacetylation pattern

For secondary screening assay to determine the deacetylation pattern, protein expression and purification using capture CBM was done according protocol 21.1.9. However, deacetylase activity measure (step 7 in protocol 21.1.9) was slightly different.

1) Deacetylase reaction was performed adding 70 ml of substrate to the E-CMB complex. After 1h reaction at 37ºC and 300 rpm, 20 ml were withdrawn to determine the deacetylase activity using fluorescamine.
2) Then, the left reaction was keep at 37ºC for 24h and 300 rpm. Afterwards, 10 μl were withdrawn and mixed with 40 μl of PBS pH 7.5, 10 μl of this mixture were analyzed by HPLC-MS (see protocol 21.5.1) in order to check the degree of acetylation looking for DP4(N), (2N) or (3N) or full deacetylated product.
3) From the same microplate, another aliquot of 10 μl were withdrawn and mixed with 40 ml of wt enzyme (final concentration of 20 mM of *Vc*CDA wt enzyme (see protocol 21.4.1)) and keep it at 37ºC for 24h in order to deacetylated the second position from non-reducing end. After 24h of reaction, 10 μl of this mixture were withdrawn and were analyzed by HPLC-MS in order to check the degree of acetylation looking for DP4(N), (2N) or (3N).

21.3. DNA manipulation

21.3.1. *Vc*CDA individual and double mutants

All mutants were obtained by site-directed mutagenesis, with a modified QuikChange® SMD protocol (Agilent). For each single mutant, a sense/antisense mutagenic primer pair was designed. Each pair contained non-overlapping sequences at their 3' end, primer-primer overlapping sequences at the 5' end and the mismatching nucleotides for the mutation located at the center of the overlapping region. The minimum number of mismatched nucleotides was introduced for each desired mutated amino acid and codon usage for E. coli was considered for the design of primers. The mutagenic primers are listed in Table 21.4.

Table 21.4. Cloning primers for *Vc*CDA individual and double mutants.

Mutant	Primer	Sequence	n	Tm (ºC)
P266G	Forward	GAATTCGGCTGTGGTACGCCGC	22	63,2
P266G	Reverse	CGTACCACAGCCGAATTCCTGGG	23	62
P271G	Forward	CGATTGAAGGCATCAATTCTAAAACCCAGG	30	59.7
P271G	Reverse	GAATTGATGCCTTCAATCGTGGTCGG	26	60.1
P280G	Forward	GAATTCGGCTGTGGTACGCC	20	58.8
P280G	Reverse	CCACAGCCGAATTCCTGGG	19	58.7
P266G-P271G	Reverse	GAATTGATGCCTTCAATCGTGGTGCC	26	60.6

Plasmid pPR-IBA *Vc*CDA CD wt was used as template for mutagenesis. PCR were carried out using IProof polymerase (0.04 U/µL, Bio-Rad) and a thermocycler Gene Pro Thermal Cycler (Bioer). Amplifications were carried out in a 40 µL final volume of the reaction mixture detailed inTable 21.5 and the PCR reaction detailed in Table 21.6.

Double mutant P266G-P271G was obtained using *Vc*CDA CD P266G as a template and using P271G forward primers and P266G-P271G reverse primer.

Table 21.5. PCR reaction mixture

Reagent	Volume (µL)
H_2O MiliQ	X (total volume)
Iproof master mix (2X)	20
Primer Forward	3
Primer Reverse	3
Template	X (3-5 ng)

Table 21.6. PCR reaction conditions

STEP	TEMPERATURE (ºC)	TIME	CYCLES
Initial denaturation	98	3 min	1
Denaturation	95	30 sec	30X
Annealing	Tm – 5ºC	30 sec	
Extension	72	2 min	
Final extension	72	10 min	1

After PCR reaction, samples were digested with Dpn1 for 2-3 hours at 37°C and, then, transformed into *E. coli* DH5α competent cells. Positive transformants were confirmed by Sanger DNA sequencing and the final constructs were used to transform *E. coli* DH5α and *E. coli* BL21(DE3) star competent cells.

21.3.2. Mutagenesis analysis for NMR experiments

*Vc*CDA CD D39S M242A and *Vc*CDA CD D39S I240L were obtained by Site directed mutagenesis. The mutagenic primers are listed in Table 21.7

Table 21.7. Cloning primers for *Vc*CDA M242A and I240L mutants for NMR studies.

Mutant	Primer	Sequence	n	Tm (ºC)
M242A	Forward	GGGCATTCCGGCGCCGGCGAATAG	24	62
M242A	Reverse	GCCGGCGCCGGAATGCCCCAGTTTTC	26	62
I240L	Forward	CTGGGGCCTGCCGATGCCGG	30	61
I240L	Reverse	GGCATCGGCAGGCCCCAGTTTTCCGG	26	61.5

Plasmid pPR-IBA *Vc*CDA CD wt was used as template for mutagenesis. PCR were carried out using IProof polymerase (0.04 U/µL, Bio-Rad) and a thermocycler Gene Pro Thermal Cycler (Bioer). Amplifications were carried out in a 40 µL final volume of the reaction mixture detailed in Table 21.5 and the PCR reaction detailed in Table 21.6.

After PCR reaction, samples were digested with Dpn1 for 2-3 hours at 37°C and, then, transformed into E. coli DH5α competent cells. Positive transformants were confirmed by Sanger DNA sequencing and the final constructs were used to transform E. coli DH5α and E. coli BL21(DE3) star competent cells.

21.3.3. Alanine scanning mutagenesis for βgl-3

Alanine scanning mutants were obtained by Site directed mutagenesis. The mutagenic primers are listed in Table 21.8.

Table 21.8. Cloning primers for alanine scanning mutants.

Mutant	Primer	Sequence	Tm (ºC)
Q32A	Forward	AGCTATGCGATTGAAGGTGCA	57.1
Q32A	Reverse	TTCAATCGCATAGCTGGCGGT	59.5
W134A	Forward	TATCATGCGGATCTGCCGCAA	59.3
W134A	Reverse	CAGATCCGCATGATACAGGGT	56.7
M354A	Forward	ACCGCAGCGGGTTGGGCAGTT	67.2
M354A	Reverse	CCAACCCGCTGCGGTGGTTTC	64
W356A	Forward	ATGGGTGCGGCAGTTGATCCG	62.7
W356A	Reverse	AACTGCCGCACCCATTGCGGT	65.2
W430A	Forward	TTCCTGGCGTCACTGCTGGAT	61.2
W430A	Reverse	CAGTGACGCCAGGAAATAACC	56
N177A	Forward	ACACTGGCGGAACCGTGGTGT	64
N177A	Reverse	CGGTTCCGCCAGTGTGGTCCA	64.5
E437A	Forward	AATTTTGCGTGGGCACATGGT	58.6
E437A	Reverse	TGCCCACGCAAAATTATCCAG	56.2
W438A	Forward	TTTGAAGCGGCACATGGTTAT	55.4
W438A	Reverse	ATGTGCCGCTTCAAAATTATC	51.8

Plasmid pET21b-βgl3 wt was used as template for mutagenesis. PCR were carried out using IProof polymerase (0.04 U/µL, Bio-Rad) and a thermocycler Gene Pro Thermal Cycler (Bioer). Amplifications were carried out in a 40 µL final volume of the reaction mixture detailed in Table 21.5 and the PCR reaction detailed in Table 21.6.

After PCR reaction, samples were digested with Dpn1 for 2-3 hours at 37°C and, then, transformed into *E. coli* DH5α competent cells. Positive transformants were confirmed by Sanger DNA sequencing and the final constructs were used to transform *E. coli* DH5α and *E. coli* BL21(DE3) star competent cells.

21.4. Expression and purification of proteins

21.4.1. Protein expression with autoinduction media

With this method (contrary to the IPTG induction strategy) is not necessary to add an inducer to the culture at a certain point since the growth of the expression strain is performed in a medium that promotes the induction of the T7 promoter, leading to the expression of the protein of interest.

- *E. coli* BL21 (DE3) cells were transformed with a pET22b or pPR-IBA2 plasmid containing a synthetic gene coding for *Vc*CDA or a construct previously designed (without signal peptide and with a C-terminus Step-tag sequence WSHPQFEK).
- Cells were grown following an autoinduction protocol in LB medium supplemented with 100 μg/ml ampicillin, 25 mM Na_2HPO_4, 25 mM KH_2PO_4, 50 mM NH_4Cl, 5 mM Na_2SO_4, 2 mM $MgSO_4$, 0.5% glycerol, 0.05% glucose and 0.2% lactose, in an orbital shaker at 180 rpm and 24°C for 60 h.
- Cells were harvested by centrifugation at 3500 rpm and resuspended in PBS buffer (50 mM phosphate pH 8.5, 0.3 M NaCl).
- The cell suspensions was lysed by sonication during 7 minutes (50% amplitude, 5 s ON, 15 s OFF, 6 mm diameter probe).
- The lysate was centrifuged at 12000 rpm for 1h, and the pellet discarded. Supernatant was used for purification of the recombinant protein.

21.4.2. Protein expression with IPTG induction

Protein expression cultures:

- 3 mL LB with 3 µL ampicillin (100 mg/mL) if needed. Incubated overnight at 37ºC and 200 rpm.
- 75 mL LB with 32,5 µL ampicillin (100 mg/mL) in 250 mL shake flasks baffled were inoculated with its respective preinoculums at a final 0.1 OD.
- After 2 hours at 37ºC and 250 rpm, when the OD reaches approximately 0.5 OD, cultures were induced with 0,5 mM IPTG.

21.4.3. Protein expression in minimal media

M9 media composition, stocks and final concentration

- Freshly chemical competent cells BL21 (DE3) pLyss star were transformed with a plasmid pPRIBA (CD PG D39S or CD wt D39S).
- Pre-inoculums of 5 mL of LB (Ampicilline and Chloramphenicol) medium were inoculated from 2 colonies and let it grow at 37ºC and 250 rpm from 9 am to 6 pm (final OD around 4.5).
- 100 mL of M9 minimal media were inoculated with 1.5 mL of preinoculum to get an OD=0,1 (at 6pm same day).
- OD next day was 3.78 (CD PG D39S) and 3.89 (PG D39S). 600 mL of M9 media were inoculated with 15 mL of previous culture to get an OD=0.1.
- OD600 were monitored. When the OD600= 1 (around 3pm), add 600 μl of IPTG 1M to get 1 mM final concentration. Let it grow the culture ON at 24ºC and 170 rpm.
- Harvest the cells at 5500 rpm/20'/4ºC.
- Resuspend the pellet with 60 mL of PBS pH 8,5 and PMSF.
- Lysis the cells with sonication during 8 min (10'' ON/20'' OFF, 50% amplitude) and centrifuge the lysate 1 hour at 12000 rpm's and 4ºC.
- Cell extract were purified using StrepTag column 5 mL.

Table 21.9. M9 media composition, stock and final concentration. All solutions except salts (autoclaved) are prepared by filter sterilization.

Solutions		Mass concentration Stock (g/L)	Final concentration
Glucose		40	4 g/L
NH_4Cl (^{15}N labeling)		20	2 g/L
Mg Salts ($MgS0_4$)		24.3	0.002 M
Fe Salts ($FeCl_3$)		0.16	0.000001 M
Thiamine		10 mg/mL (H_2O)	10 mg/L
Biotin		10 mg/mL (H_2O-pH basic)	10 mg/L
Ampicillin		100 mg/mL	0.1mg/L
Chloramphenicol		25 mg/mL	0.025 mg/L
Salts (10X)	Na_2HPO_4	70.98	1X
	KH_2PO_4	68.04	
	Na_2SO_4	2.67	
Traces (pH:1-5/2)	$CaCl_2$	1.239	3 mL/L culture
	$AlCl_3 \cdot 6H_2O$	0.04	
	$ZnSO_4 \cdot 7H_2O$	0.87	
	$CoCl_2 \cdot 6H_2O$	0.16	
	$CuSO_4 \cdot 5H_2O$	1.967	
	H_3BO_3	0.01	
	$MnCl_2$ $4H_2O$	1.42	
	$NiCl_2$ $6H_2O$	0.01	
	Na_2MoO_4	0.02	

Table 21.10. M9 media composition, 100 mL and 600 mL final volumes.

Solutions: M9	Volume (100 mL culture)	Volume (600 mL culture)j
Water	67	400
Salts	10	60
Glycerol	10	60
NH_4Cl	10	60
Traces	0.3	1,8
Mg salts	1	6
$FeCl_3$	1	6
Thiamine	100 μL	600 μL
Biotin	100 μL	600 μL
Amp	100 μL	600 μL
Cm	100 μL	600 μL

21.4.4. Strep-trap affinity chromatography

A strep-tag II can be found in all plasmids containing V*c*CDA constructs or mutants, either on the C-terminus (pET22b(+) or the N-terminus (pPR-IBA2) end of proteins. This tag is used for the initial purification by affinity chromatography.

Supernatant from cell extract centrifugation was loaded to a pre-conditioned Strep-Trap 1 mL or 5 mL column in several cycles. The conditioning procedure for a previously used column included several steps, and it was performed at a 2-3 mL/min flow rate:

- 5 column volumes of H_2O.
- 5 column volumes of 0.5M NaOH.
- 5 column volumes of H_2O (to completely remove NaOH, it is recommended to check pH).
- 5 column volumes of the buffer in which the protein to purify is resuspended (usually PBS pH 8.5).
- 2 column volumes of HABA 1 mM in PBS (or until column color has completely changed to red).
- PBS to remove HABA, until column color has completely changed back to white.

When using a new column, it can be directly washed with H_2O to remove the conservation solution and, then, conditioned with the corresponding buffer. All solutions that were loaded to the column were previously filtrated with a 0.45 µm filter.

The procedure for each loading-eluting cycle was the following, the total lyase volume was divided to load 10-12 ml each time. First, the strep-trap column was connected to the Äkta fraction collector (ÄKTA FPLC, Abersham Biosciences). 10 to 12 mL of clarified lysate were loaded at 2-3 mL/min and monitored with ÄKTA's UV absorbance detector. The column was washed with PBS until absorbance values reached the baseline and elution was performed by adding d-desthibiotin 2.5 mM in PBS pH 8.5. The fraction corresponding to the appearance of a peak in the absorbance measure was kept. The column was then washed with 2 volumes of PBS, regenerated with 5 volumes of HABA 1 mM in PBS and, finally, washed again with PBS to eliminate HABA.

A standard purification could include 3 to 5 cycles depending on the initial volume and on the quantity of protein observed in the elution steps. After purification, the column was washed with H_2O and conserved in a 20% EtOH solution.

21.4.5. His-tag column affinity

A his-tag can be found in the pET21b(+) plasmid containing β-glucosidase genes, on the C-terminus end of proteins. This tag is used for their purification by affinity chromatography.

- Supernatant from cell extract centrifugation was loaded to a 5 mL Ni^{2+} charged His-Trap chelating column connected to the Äkta fraction collector (Äkta FPLC, Abersham Biosciences).
- From this loading, all purification steps were monitored with Äkta's UV absorbance detector. The column was washed with PBS buffer (20 mM phosphate, 150 mM NaCl, pH 7) until absorbance values reached the baseline and elution was performed with an elution buffer (20 mM phosphate, 150 mM NaCl, 500 mM imidazole pH 7) gradient (from 0 to 100% in 60 minutes).
- The fractions corresponding to the appearance of a peak in the absorbance (corresponding to elution buffer concentration between 10 and 40%) was collected and kept. The column was then washed with PBS and H2O and conserved in a 20% EtOH solution.

21.4.6. Size Exclusion Chromatography

21.4.6.1. Preparative chromatography

For preparative separation of different oligomeric fractions of the studied proteins in this work, a Superdex 200 16/600 column was used (GE Healthcare). Before using it for purification and connecting it to the monitored fraction collector ÄKTA, the Superdex column was conditioned with a flow rate of 1 mL/min, adding 100 mL of H_2O and 100 mL of the desired buffer (usually, PBS pH 7). All solutions that were loaded to the column were previously filtrated with a 0.45 µm filter.

All elution fractions from the previous affinity chromatography were pooled together and concentrated to a maximum volume of 2 ml with a 10 kDa MWCO Amicon Ultra-centrifugal filter (Merck Millipore). The concentrated protein was loaded to the column through a previously conditioned 2 ml loop (washed with H_2O and PBS) and elution was performed using the same buffer at 1 mL/min flow rate. The elution of the oligomeric fractions was monitored by online measure of 280 nm absorbance and 1 mL fractions were collected in deep-well microplates.

Once the last fraction was eluted, the column was washed with 100 mL of H_2O and 100 mL of 20% EtOH solution.

All fractions were conserved for their analysis by SDS-PAGE electrophoresis and, once purity and identity (by molecular mass determination) of the interest protein were determined, all aliquots of the desired oligomeric fraction were pooled together. In the case of all *Vc*CDA variants, this fraction was the monomer, used for further protein characterization.

21.4.6.2. Column calibration

Calibration of the gel filtration column allows to precisely identify the apparent molecular weight of a protein (or any other compound loaded to the column) depending on the needed volume to move it through the column (elution volume).

In order to calibrate the preparative column, the dead volume (V_0) of the column was estimated and a calibration curve was obtained using standard proteins of known weight. The procedure, in which a set of standards from Sigma-Aldrich was used (Gel Filtration Markers Kit for Protein Molecular Weights 6,500-66,000 Da), include conditioning the column as previously described.

Then, a series of solution with the standard proteins suspended in PBS were prepared (Table 7.7) and filtered through a 0.45 μm filters.

The Superdex 200 column was connected to Äkta FPLC and 1 ml of each of the three solutions was loaded to the column through the injection loop in the order that is indicated in Table 7.7. The elution of proteins was performed at the same flow rate used for preparative separation, 1 mL/min, and it was monitored by online measuring of 280 nm absorbance. Chromatograms data were extracted, and the maximum signal of each protein/compound was considered its elution volume (Ve). A linear regression fit was performed, using Log(Mw) and V_e/V_0 values (Figure 21.1).

Table 21.11. Elution volumes for calibration of Superdex 200 column. Elution times of the proteins and compounds used for the calibration of the gel filtration column and data transformation to obtain the standard curve.

Compound	MW (kDa)	Log (kDa)	V_e (mL)	V_e/V_0
Aprotinin	**6.5**	**0.81**	**101.2**	**2.4**
Cytochrome C	**12.4**	**1.09**	**95.5**	**2,1**
Carbonic Anhydrase	**29**	**1.46**	**84.4**	**1.9**
Albumin	**66**	**1.82**	**72.4**	**1.7**
Dextran Blue (V_0)	**2000**		**45.9**	**1**

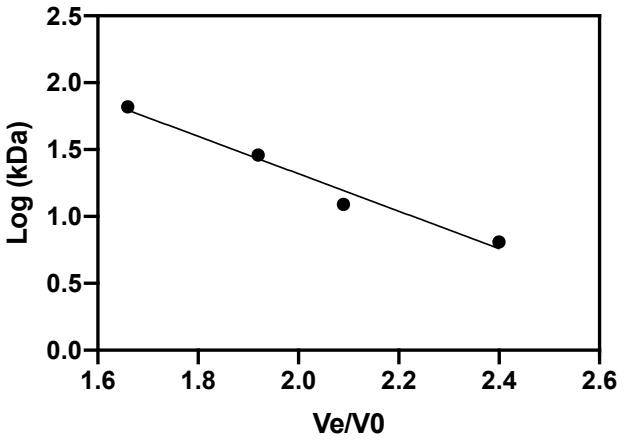


Figure 21.1. Standard curve of the Superdex 200 column calibration. "Y = -1.399*X + 4.118"

The obtained equation was used to calculate the apparent Mw of the proteins or fractions that were separated using this methodology (Equation 8.2).

Equation 21.2. Calculate the apparent Mw of the proteins

$$Mw = 10^{-1{,}399\left(\frac{V_e}{V_0}\right)+4.118}$$

21.5. Kinetics characterization of *Vc*CDA enzymes

21.5.1. Deacetylase activity determination by HPLC-MS

The formation of products with different degrees of acetylation from chitooligosaccharides substrates, $(GlcNAc)_{2-5}$, was monitored by HPLC-MS (Agilent 1260 HPLC-MS, electrospray ionization (ESI +), single quadrupole detector) using a XBridge BEH Amide 2.5 μM, 3.0 x 100 mm XP column (Waters) in combination with an XBridge BEH Amide Guard Cartridge (2PK) precolumn (2.5 μM x 20 mm) (Waters).

21.5.2. Specific activity of enzymes

Specific activities for the initial monodeacetylation reactions with the same substrates were determined with the same procedure for all enzymes presented in this work. Reactions were performed with different enzyme concentration (between 0.25 and 5 μM) and using the corresponding monodeacetylated product standard for quantification (section 21.5.3). The standards were prepared by reaction of the corresponding $(GlcNAc)_n$ substrate with *Vc*CDA, as previously reported, and several dilutions were prepared for each set of standards.

21.5.3. HPLC-MS methodology for the estimation of monodeacetylated products: Preparation of standards.

To obtain the different chitosan oligosaccharide standards 10mg of each of the fully acetylated chitooligosaccharides (DP2-DP4-DP5) were dissolved in 0.5ml of PBS. The de-*N*-acetylation reaction started by the addition of the previously purified *VcCDA* to a final concentration of 3μM and a total reaction volume of 1,5ml. The reaction cocktail was incubated a 37ºC, the conversion rate was frequently checked by the HPLC-MS. Once the reaction reached total substrate consumption, enzyme was removed by affinity chromatography. The pure mono de-*N*-acetylated products were extracted from the deproteinized reaction mixture by cation exchange chromatography using a Dowex 50wx8 resin (Sigma-Aldrich). Products were eluted using an NH_3 gradient. Ammonia was evaporated and the purity of the fractions was analyzed by mass spectrometry. All fractions containing pure mono de-*N*-acetylated chitosan oligosaccharides were pooled and freeze-dried.

Purified deacetylated di-, tri-, and tetrasaccharide compounds were dissolved in 1-propanol:water 1:1 to a concentration of 1mM and kept at -20ºC as stocks. Starting from these solutions, standard curves were diluted to a series of determined concentrations ranging from 0.05μM to 225μM.

21.5.4. Michaelis-Menten reaction

Initial rates of deacetylation by VcCDA were determined with $(GlcNAc)_n$, n=2, 4, 5. Substrate solutions (100 μM to 4 mM final concentrations) in PBS buffer pH 8.5 were preincubated at 37°C for 5 min. Reactions were initiated by addition of the enzyme (0.05-1 μM final concentrations) in a reaction volume of 100 μL. At different time intervals, aliquots (10 μl) were withdrawn and added to 90 μl of water/1-propanol (1:1) to stop the reactions. Samples were analyzed by HPLC-MS (Agilent 1260 HPLC-MS, electrospray ionization (ESI+), single quadrupole MS detector) using a X-Bridge BEH Amide 2.5 μm 3.0x100 mm XP Column (Waters), 5 μL injection, and isocratic elution at 60°C with acetonitrile/water 65:35 v/v, 0.1% formic acid, at a flow rate of 0.4 ml/min. Standards for quantification of monodeacetylated di-,tetra , and pentasaccharide products were prepared by enzymatic deacetylation of $(GlcNAc)_n$, n=2, 4, 5, with VcCDA at 100% conversion.

Enzyme kinetics were done in sealed 96-well microtiter plates, each microplate containing two replicas of product standards to be analyzed at the beginning and at the end of each reaction set (section 21.5.3.). Product formation in the enzymatic reactions was quantified by HPLC-MS (SIM for [M+H]+, m/z 383, 789, and 993 for monodeacetylated di-, tetra-, and pentasaccharide products, respectively). Initial rates were calculated as the slope of the linear time-course [product] vs. time (typically 4 to 6 time points).

Initial velocities are calculated by linear regression on a standard line, and the parameters kcat and KM are obtained by non-linear regression of v0 / [E] against the [substrate] using the Michaelis & Menten equation using Graphpad software.

21.6. Kinetic characterization of β-glucosidase enzymes

21.6.1. Specific velocity assay

Specific velocity assays were performed in 96-well PVC plates and read in 96-well plates with flat transparent bottom (Thermo Scientific).

- Using Robot Bravo 100 µL of different enzyme concentration were mixed with 50 µL substrate per well at 37ºC.
- At different times (2', 5', 7', 10', 12' and 15') 20 µL of the reaction mix were dropped into 120 µL NaOH 2M in the reading plate wells, NaOH solution pH ensure reaction stops and allows pNP visibility.
- Absorbance reading was performed with ELx808 Absorbance Reader (BioTek®) at 405 nm wavelength. A pNP standard curve of calibration ranging from 0mM to the maximum substrate concentration was used.
- For septanose hydrolytic activity using Sept-pNP, same previous protocol was done.

21.6.2. Michaelis-Menten assay

Michaelis Menten assays were performed for wild type enzyme and W430A enzyme with Glc-pNP as substrate. The procedure used was the same as specific velocity assays but maintaining enzyme concentration constant and changing substrate concentration.

Equation 17.1 was used to determine the enzyme inhibition model.

21.7. Modelling of βgl3 enzyme in complex with substrates

A pair of programs were used to perform and visualize all substrate-protein dockings. The program used to calculate docking positions was Autodock tools-1.5.6. (Molecular Graphics Laboratory, The Scripps Research Institute) with autogrid4 and autodock4 as extension programs. The program used to visualize PDB files was VMD 1.9.3.

Initial PDB data was obtained from RCSB PDB (https://www.rcsb.org/). Every PDB structure have its related PDB code (Example: 1GON, β-glucosidase from *Streptomyces sp.*).

Starting from PDB molecule data Autodock tools allow the user to add polar hydrogens. For a proper docking AD4 atom type assignation and Gasteiger charge computation is needed. Molecule data was stored as PDBQT. For the ligand is necessary to detect the torsion tree and choose bond torsions. The grid box centre and dimensions were selected including all catalytic pocket. Autodock4 calculation were performed with 1000 docking per single run using the default genetic algorithm provided by the program.

21.8. NMR methodology

Methyl-TROSY and NOESY spectrum (500 ms) were collected on an IdM-labelled sample at 298K and on a 800 MHz Bruker Avance III spectrometer equipped with a cryoprobe. Transversal relaxation dispersion experiments were acquired at two fields on the abovementioned 800 MHz spectrometer and on a 600 MHz Avance III spectrometer, at 298 K and using a relaxation compensated pulse CPMG sequence** (60 ms total CPMG time) and variable effective fields: 33, 66 (x2), 133, 200, 266, 400, 600, 800, 1133, 1533 and 2000 Hz. Different datasets were collected for the following samples: CD Apo, CD:DP2 (excess), CD:DP4 (excess), PG Apo, PG:DP2 (excess) and PG:DP4 (excess), consisting of 150 mM of the receptor, 20 mM of Tris (pH 7.4), 150 mM NaCl and 10% D_2O. Dispersion data were fit to the Carver and Richards equation using in-house Matlab© scripts, either to one field alone or simultaneously using data from two fields. Duplicate data were used to obtain an estimation of the error.

Table 21.12. NMR Experiments conducted in the present project.

Experiment	Protein	Magnetic Fields (MHz)	Points measured (Hz)	Sample conditions
CD-TROSY	CD	800	n.a.	150 μM CD, 20 Tris pH 7.4, 150 mM NaCl. T = 298K
CD-RD	CD	800(x2)*, 600(x2)*	33, 66(x2)**, 133, 200, 266, 400, 600, 800, 1133, 1533, 2000	150 μM CD, 20 Tris pH 7.4, 150 mM NaCl. T = 298K
CD/DP2-TROSY	CD + DP2	800	n.a.	150 μM CD, 20 Tris pH 7.4, 150 mM NaCl+ excess of DP2 (solubility limit). T = 298K
CD/DP2-RD	CD + DP2	800, 600	33, 66(x2)**, 133, 200, 266, 400, 600, 800, 1133, 1533, 2000	150 μM CD, 20 Tris pH 7.4, 150 mM NaCl+ excess of DP2 (solubility limit). T = 298K
CD/DP4-TROSY	CD + DP4	800	n.a.	150 μM CD, 20 Tris pH 7.4, 150 mM NaCl + excess of DP4 (solubility limit). T = 298K
CD/DP4-RD	CD + DP4	800, 600	33, 66(x2)**, 133, 200, 266, 400, 600, 800, 1133, 1533, 2000	150 μM CD, 20 Tris pH 7.4, 150 mM NaCl+ excess of

				DP4 (solubility limit). T = 298K
PG-TROSY	PG	800	n.a.	150 μM PG, 20 Tris pH 7.4, 150 mM NaCl. T = 298K
PG-RD	PG	800, 600	33, 66(x2)**, 133, 200, 266, 400, 600, 800, 1133, 1533, 2000	150 μM PG, 20 Tris pH 7.4, 150 mM NaCl. T = 298K
PG/DP2-TROSY	PG + DP2	800	n.a.	150 μM PG, 20 Tris pH 7.4, 150 mM NaCl+ excess of DP2 (solubility limit). T = 298K
PG/DP2-RD	PG + DP2	800, 600	33, 66(x2)**, 133, 200, 266, 400, 600, 800, 1133, 1533, 2000	150 μM PG, 20 Tris pH 7.4, 150 mM NaCl+ excess of DP2 (solubility limit). T = 298K
PG/DP4-TROSY	PG + DP4	800	n.a.	150 μM PG, 20 Tris pH 7.4, 150 mM NaCl+ excess of DP4 (solubility limit). T = 298K
PG/DP4-RD	PG + DP4	800, 600	33, 66(x2)**, 133, 200, 266, 400, 600, 800, 1133, 1533, 2000	150 μM PG, 20 Tris pH 7.4, 150 mM NaCl+ excess of DP4 (solubility limit). T = 298K

21.9. Basic Biochemistry protocols

21.9.1. SDS-PAGE electrophoresis

Different solution should be prepared:

- Solution A: 30% acrylamide, 2.7% bisacrylamide (Thermo-Fisher Scientific). ·
- Solution B: 1.5M Tris-HCl pH 5.8 (Sigma-Aldrich®).
- Solution C: 0.5M Tris-HCl pH 5.8 (Sigma-Aldrich®).
- Solution D: 10% SDS (w / V) (Sigma-Aldrich®). · TEMED (Sigma-Aldrich®). ·
- Ammonium persulfate (Sigma-Aldrich®).
- 10x electrophoresis buffer: 30g / L Tris-Base, 144g / L glycine 10g / L SDS pH 8.3. · 5x loading buffer: 62.5 mM Tris-HCl, 25% glycerol, 2% SDS, 5% β-mercaptoethanol, a small amount of bromophenol blue.
- Molecular weight marker: 2: 4: 14 Stock marker: 5X loading buffer: H_2O MilliQ (Low Range Standard marker, Bio-Rad).

In this work, the SDS-PAGE gels made are always 14% acrylamide. The volumes of each solution are indicated below.

- Stacking gel: 2.7 mL of miliQ water, 4.7 mL of solution A, 2.5 mL of solution B, 100 μL of solution D and 5 μL of TEMED
- Running gel: 6.1 mL of miliQ water, 1.3 mL of solution A, 2.5 mL of solution C, 100 μL of solution D and 10 μL of TEMED

The gel support (Mini-Protean, Bio-Rad) is assembled. The separator gel solution is made and just before pouring it into the mold, a spatula tip of ammonium persulfate is added and it is poured into the mold leaving space for the stacking gel. A small volume of water is added on top of the gel to level and it is allowed to gel. Once gel was solidified, the water is eliminated and the stacking gel is added on the separator, preparing it in an identical way to the separator with the proportions indicated in the table. The well mold is introduced and allowed to gel. Once gelled, the mold is removed, and the assembly is introduced into the tray. The gels are covered with 1x electrophoresis buffer.

21.9.2. Protein quantification by BCA

Protein concentration was determined by the BCA method (Thermo Scientific) and Bovine Serum Albumin (BSA) was used as protein reference. Several dilutions of BSA (from 0 to 2 mg/mL in the same buffer in which the interest protein was dissolved) were prepared as standard curve. The protocol indicated by the manufacturer was followed for different dilutions of protein samples and for the standard curve.

20 μL of each sample (different dilutions of the interest protein and standard curve samples) was mixed with 160 μL of the working reagent (mixture described by the manufacturer's protocol) and incubated for 30 minutes at 37°C in a microtiter plate. Absorbance at 595 nm was measured with a microplate reader.

21.10. Basic molecular biology protocols

21.10.1. Chemical competent cells preparation and transformation

This protocol was used to obtain chemical competent cells for both DNA subcloning (*E. coli* DH5α, Invitrogen) and protein expression (*E. coli* BL21(DE3) star, Invitrogen).

- 50 mL of LB were inoculated with target cells and incubated at 37°C and 250 rpm until $0D_{600}$ reached a 0.6 value.
- Then, cells were kept on ice for 20 minutes and centrifuged for 5 minutes at 5000xg and 4°C. Supernatant was removed by decanting and cells were resuspended in 25 mL of chilled sterile 50 mM NaCl.
- A second centrifugation step in the same conditions was performed and cells were resuspended in 5 mL of chilled sterile 50 mM NaCl2 after removing supernatant.
- 750 μL of sterile glycerol were added to cell suspension (final glycerol concentration of 15%) and, finally, aliquots of 200 μL of these cells were transferred to sterile Eppendorf tubes and stored at -80°C until needed.
- For transformation, between 2 and 10 μl of DNA were added to competent cells after thawing them on ice. Several controls were always set up together with the targeted DNA:
    - As a negative control of contamination, sterile water was added to competent cells, which at the final step were plated in medium with antibiotic.
    - As a positive control of viability of cells, competent cells were plated in medium without antibiotic.
    - As positive control of transformation, a plasmid with known transformation efficacy was transformed.
- Cells with DNA/water were kept for 30 minutes on ice, then incubated for 2 minutes at 42°C and transferred to ice for 5 minutes.
- 500 μL of LB medium were added and cells were incubated for 45 minutes on a shaker at 250 rpm and 37°C. After that time, cells were plated to plates with the desired medium and antibiotic.

21.10.2. Electrochemical competent cells preparation and transformation

- Inoculate 1 colony from a fresh plate of the strain to be made electrocompetent into 10 ml of LB in a 125 ml flask and incubate for 16-18 hours at 37°C and 250 rpm.
- 250 mL flasks containing 50 mL each of LB pre-warmed to 37°C. Add 1 mL of the overnight culture to each of the flasks.
- Shake at 37°C and 250 rpm until the cultures reach an OD_{600} of 0.5-0.7. Be sure to turn on centrifuge and cool rotor to 4°C well in advance of harvesting cells.
- Place cultures on ice for 15 minutes. From this point on the cultures must be kept ice cold.
- Centrifuge at 5000 rpm for 10 min. Pour off the supernatant and aspirate any residual broth.
- Centrifuge at 5000 rpm for 10 min. Pour off the supernatant, it is not necessary to aspirate. Completely suspend the cells in 25 mL H_2O miliQ with 10% glycerol and re-centrifuge.
- Centrifuge at 5000 rpm for 10 min. Pour off the supernatant, it is not necessary to aspirate. Completely suspend the cells in 10 mL H_2O miliQ with 10% glycerol and re-centrifuge.
- Pour off the supernatant and suspend the cells in the residual in 250 μL H_2O miliQ with 10% glycerol by pipetting up and down. Aliquoted in 50 μL in Eppendorf tubes previously autoclaved.
- At this point you can electroporate or freeze the cells away. Once you have used all of the culture, transfer the tubes to dry ice for 10 minutes. Once the cultures are frozen, transfer them to a -80°C freezer. The cultures should be good for >6 months.
- Turn on electroporator and set to 1.7-2.5 kv (optimize for strain). Place recovery SOC in 37ºC water bath.
- Pre-warm LB-antibiotic plates at 37ºC.
- Thaw cells on ice for 10 min or use freshly made cells.
- Place appropriate number of microcentrifuge tubes and 1 mm-electroporation cuvettes on ice.
- Flick the tube containing cells a few times to mix and add 25 μL to the microcentrifuge tubes.

- Add 1 μl of a 20 ng/μL DNA solution (in DI water) to the cells in the microcentrifuge tube.
- Transfer the DNA-cell mixture to the cold cuvette, tap on countertop 2X, wipe water from exterior of cuvette and place in the electroporation module and press pulse (don't hold the button down).
- Immediately add 600 μL of 37ºC SOC, mix by pipetting up and down once and transfer to a 15 ml-falcon tube.
- Spread the cells in LB agar plates.

21.10.3. DNA obtention and purification

For plasmidic DNA extraction, different commercial miniprep kits were used (Sigma Aldrich; Clinisciences). Protocol provided by each manufacturer was followed after growing an overnight culture of the transformed strain with the interest plasmid in 5 mL of LB medium with the required antibiotic.

For purification of DNA fragments or specific bands from an agarose gel, the GenElute™ Gel Extraction Kit (Sigma Aldrich) was used. The protocol supplied by the manufacturer was followed after excising the desired band or bands from 1% agarose gel.

For purification of PCR products, the QIAquick spin – PCR purification Kit (Qiagen) was used, following the protocol provided by the manufacturer.

21.10.4. DNA quantification

The methodology used in this work to quantify DNA was the Qubit® system (Thermo Fisher), based on target-selective dyes that emit fluorescence when bound to DNA. Two different kits were used, depending on the expected concentration of DNA of the sample (high sensitivity, to quantify samples between 10 and 100 ng/μL; and broad range, to quantify samples between 100 and 1000 ng/μL).

21.10.5. DNA electrophoresis gel

DNA electrophoresis agarose gels is a current biomolecular separation method based on their size. The electrophoresis gel composition varies depending on the DNA size, in this case 1% agarose (ranging from 500 to 10.000 bp) in Tris-acetate-EDTA (TAE) buffer.

- Samples were loaded using x6 Dye Blue loading buffer (4g sucrose and bromophenol blue in 10 mL H2O) in 1:5 ratio (Dye Blue loading buffer: Sample).
- DNA marker used was Marker III (Sigma-Aldrich, ranging from 564 - 21.226 bp) diluted 4/2/6 (marker/ x6 Dye Blue loading buffer/water).
- Gel composition was 30 mL TAE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) with 1% agarose and 1 µL ethidium bromide (revealing agent).
- Gel is immersed in TAE buffer and a 90-120 V voltage applied until the dye reached the end (approximately 45 minutes).
- DNA bands were revealed with ultraviolet light because of ethidium bromide fluoresce.
- If needed, DNA can be covered using GenEluteTM Gel Extraction Kit (Sigma-Aldrich). The kit makes possible the partial DNA recovery from an agarose gel band.

BIBLIOGRAPHY

22. BIBLIOGRAPHY

1. C, R. Des substances amylacées dans les tissus des animaux, spécialement des articulés (chitine) Available online: https://scholar.google.com/scholar_lookup?journal=Comp.+Rend&title=Des+substances+amylacées+dans+les+tissus+des+animaux,+spécialement+des+articulés+(chitine)&author=C.+Rouget&volume=48&publication_year=1859&pages=792-795& (accessed on May 4, 2021).

2. Hoppe-Seyler, F. Ueber Chitin und Cellulose. *Berichte der Dtsch. Chem. Gesellschaft* **1894**, *27*, 3329–3331, doi:10.1002/cber.189402703135.

3. Ghormade, V.; Kulkarni, S.; Doiphode, N.; Rajamohanan, P. R.; Deshpande, M. V Chitin deacetylase: A comprehensive account on its role in nature and its biotechnological applications. *Curr. Res. Technol. Educ. Top. Appl. Microbiol. Microb. Biotechnol.* **2010**, 1054–1066.

4. Rinaudo, M.; Ã, M. R. Chitin and chitosan: Properties and applications. *Prog. Polym. Sci.* **2006**, *31*, 603–632, doi:10.1016/j.progpolymsci.2006.06.001.

5. Pradip Kumar Dutta*, J. D. and V. S. T. Chitin and chitosan: Chemistry, properties and applications. *Chitin Chitosan Prop. Appl.* **2019**, 1–510, doi:10.1002/9781119450467.

6. Kurita, K. Controlled functionalization of the polysaccharide chitin. *Prog. Polym. Sci.* 2001, *26*, 1921–1971.

7. Varki, A. Essentials of Glycobiology. *Cold Spring Harb.* **2015**, 823.

8. John P. Zikakis *Chitin, Chitosan, and Related Enzymes*; Elsevier, 1984;

9. Di Mario, F.; Rapanà, P.; Tomati, U.; Galli, E. Chitin and chitosan from Basidiomycetes. *Int. J. Biol. Macromol.* **2008**, *43*, 8–12, doi:10.1016/j.ijbiomac.2007.10.005.

10. Grifoll-Romero, L.; Pascual, S.; Aragunde, H.; Biarnés, X.; Planas, A. Chitin deacetylases: Structures, specificities, and biotech applications. *Polymers (Basel).* **2018**, *10*, doi:10.3390/polym10040352.

11. Basa, S.; Nampally, M.; Honorato, T.; Das, S. N.; Podile, A. R.; Gueddari, N. E. El; Moerschbacher, B. M. The Pattern of Acetylation Defines the Priming Activity of Chitosan Tetramers. *Cite This J. Am. Chem. Soc* **2020**, *142*, doi:10.1021/jacs.9b11466.

12. Kuroki, M.; Okauchi, K.; Yoshida, S.; Ohno, Y.; Murata, S.; Nakajima, Y.; Nozaka, A.; Tanaka, N.; Nakajima, M.; Taguchi, H.; Saitoh, K. I.; Teraoka, T.; Narukawa, M.; Kamakura, T. Chitin-deacetylase activity induces appressorium differentiation in the rice blast fungus Magnaporthe oryzae. *Sci. Rep.* **2017**, *7*, 3–5, doi:10.1038/s41598-017-10322-0.

13. Zhou, J.; Wen, B.; Xie, H.; Zhang, C.; Bai, Y.; Cao, H.; Che, Q.; Guo, J.; Su, Z. Advances in the preparation and assessment of the biological activities of chitosan oligosaccharides with different structural characteristics. *Food Funct.* **2021**, *12*, 926–951, doi:10.1039/d0fo02768e.

14. Tsigos, I.; Martinou, A.; Kafetzopoulos, D.; Bouriotis, V. Chitin deacetylases: New, versatile tools in biotechnology. *Trends Biotechnol.* 2000, *18*, 305–312.

15. Krajewska, B. Application of chitin- and chitosan-based materials for enzyme immobilizations: A review. *Enzyme Microb. Technol.* **2004**, *35*, 126–139, doi:10.1016/j.enzmictec.2003.12.013.

16. Kurita, K. Chitin and chitosan: Functional biopolymers from marine crustaceans. *Mar. Biotechnol.* 2006, *8*, 203–226.

17. Guan, G.; Abul Kalam Azad, M.; Lin, Y.; Kim, S. W.; Tian, Y.; Liu, G.; Wang, H. Biological effects and applications of chitosan and chito-oligosaccharides. *Front. Physiol.* 2019, *10*, 516.

18. Muanprasat, C.; Chatsudthipong, V. Chitosan oligosaccharide: Biological activities and potential therapeutic applications. *Pharmacol. Ther.* **2017**, *170*, 80–97, doi:10.1016/j.pharmthera.2016.10.013.

19. Mendis, E.; Kim, M. M.; Rajapakse, N.; Kim, S. K. An in vitro cellular analysis of the radical scavenging efficacy of chitooligosaccharides. *Life Sci.* **2007**, *80*, 2118–2127, doi:10.1016/j.lfs.2007.03.016.

20. Huang, R.; Mendis, E.; Kim, S. K. Factors affecting the free radical scavenging behavior of chitosan sulfate. *Int. J. Biol. Macromol.* **2005**, *36*, 120–127, doi:10.1016/j.ijbiomac.2005.05.001.

21. Forrester, S. J.; Kikuchi, D. S.; Hernandes, M. S.; Xu, Q.; Griendling, K. K. Reactive oxygen species in metabolic and inflammatory signaling. *Circ. Res.* 2018, *122*, 877–902.

22. Wang, J.; Wang, J.; He, W.; He, W.; Tsai, P. J.; Chen, P. H.; Ye, M.; Guo, J.; Su, Z. Mutual interaction between endoplasmic reticulum and mitochondria in nonalcoholic fatty liver disease. *Lipids Health Dis.* 2020, *19*.

23. Zhao, Q.; Yin, L.; Zhang, L.; Jiang, D.; Liu, L.; Ji, H. Chitoheptaose Promotes Heart Rehabilitation in a Rat Myocarditis Model by Improving Antioxidant, Anti-Inflammatory, and Antiapoptotic Properties. *Oxid. Med. Cell. Longev.* **2020**, *2020*, doi:10.1155/2020/2394704.

24. Yousef, M.; Pichyangkura, R.; Soodvilai, S.; Chatsudthipong, V.; Muanprasat, C. Chitosan oligosaccharide as potential therapy of inflammatory bowel disease: Therapeutic efficacy and possible mechanisms of action. *Pharmacol. Res.* **2012**, *66*, 66–79, doi:10.1016/j.phrs.2012.03.013.

25. Chung, M. J.; Park, J. K.; Park, Y. Il Anti-inflammatory effects of low-molecular weight chitosan oligosaccharides in IgE-antigen complex-stimulated RBL-2H3 cells and asthma model mice. *Int. Immunopharmacol.* **2012**, *12*, 453–459, doi:10.1016/j.intimp.2011.12.027.

26. Riaz Rajoka, M. S.; Zhao, L.; Mehwish, H. M.; Wu, Y.; Mahmood, S. Chitosan and its derivatives: synthesis, biotechnological applications, and future challenges. *Appl. Microbiol. Biotechnol.* 2019, *103*, 1557–1571.

27. Zhang, P.; Liu, W.; Peng, Y.; Han, B.; Yang, Y. Toll like receptor 4 (TLR4) mediates the stimulating activities of chitosan oligosaccharide on macrophages. *Int. Immunopharmacol.* **2014**, *23*, 254–261, doi:10.1016/j.intimp.2014.09.007.

28. Feng, J.; Zhao, L.; Yu, Q. Receptor-mediated stimulatory effect of oligochitosan in macrophages. *Biochem. Biophys. Res. Commun.* **2004**, *317*, 414–420, doi:10.1016/j.bbrc.2004.03.048.

29. Pokhis, K.; Bitterlich, N.; Cornelli, U.; Cassano, G. Efficacy of polyglucosamine for weight loss-confirmed in a randomized double-blind, placebo-controlled clinical investigation. *BMC Obes.* **2015**, *2*, doi:10.1186/s40608-015-0053-5.

30. Trivedi, V.; Satia, M.; Deschamps, A.; Maquet, V.; Shah, R.; Zinzuwadia, P.; Trivedi, J. Single-blind, placebo controlled randomised clinical study of chitosan for body weight reduction. *Nutr. J.* **2016**, *15*, doi:10.1186/s12937-016-0122-8.

31. Ylitalo, R.; Lehtinen, S.; Wuolijoki, E.; Ylitalo, P.; Lehtimäki, T. Cholesterol-lowering properties and safety of chitosan. *Arzneimittel-Forschung/Drug Res.* 2002, *52*, 1–7.

32. Kao, C. H.; Hsiang, C. Y.; Ho, T. Y. Assessment of chitosan-affected metabolic response by peroxisome proliferator-activated receptor bioluminescent imaging-guided transcriptomic analysis. *PLoS One* **2012**, *7*, 34969, doi:10.1371/journal.pone.0034969.

33. Chang, Y.; Huang, Z.; Jiao, Y.; Xu, J. F.; Zhang, X. PH-Induced Charge-Reversal Amphiphile with Cancer Cell-Selective Membrane-Disrupting Activity. *ACS Appl.*

Mater. Interfaces **2018**, *10*, 21191–21197, doi:10.1021/acsami.8b06660.

34. Huang, R.; Mendis, E.; Rajapakse, N.; Kim, S. K. Strong electronic charge as an important factor for anticancer activity of chitooligosaccharides (COS). *Life Sci.* **2006**, *78*, 2399–2408, doi:10.1016/j.lfs.2005.09.039.

35. Zhai, X.; Yuan, S.; Yang, X.; Zou, P.; Li, L.; Li, G.; Shao, Y.; Abd El-Aty, A. M.; Haclmüftüoğlu, A.; Wang, J. Chitosan Oligosaccharides Induce Apoptosis in Human Renal Carcinoma via Reactive-Oxygen-Species-Dependent Endoplasmic Reticulum Stress. *J. Agric. Food Chem.* **2019**, doi:10.1021/acs.jafc.8b06941.

36. Zou, P.; Yuan, S.; Yang, X.; Zhai, X.; Wang, J. Chitosan oligosaccharides with degree of polymerization 2–6 induces apoptosis in human colon carcinoma HCT116 cells. *Chem. Biol. Interact.* **2018**, *279*, 129–135, doi:10.1016/j.cbi.2017.11.010.

37. Mei, Y. xia; Chen, H. xia; Zhang, J.; Zhang, X. dan; Liang, Y. xiang Protective effect of chitooligosaccharides against cyclophosphamide-induced immunosuppression in mice. *Int. J. Biol. Macromol.* **2013**, *62*, 330–335, doi:10.1016/j.ijbiomac.2013.09.038.

38. Garcia, L. G. S.; Guedes, G. M. de M.; da Silva, M. L. Q.; Castelo-Branco, D. S. C. M.; Sidrim, J. J. C.; Cordeiro, R. de A.; Rocha, M. F. G.; Vieira, R. S.; Brilhante, R. S. N. Effect of the molecular weight of chitosan on its antifungal activity against Candida spp. in planktonic cells and biofilm. *Carbohydr. Polym.* **2018**, *195*, 662–669, doi:10.1016/j.carbpol.2018.04.091.

39. Verlee, A.; Mincke, S.; Stevens, C. V. Recent developments in antibacterial and antifungal chitosan and its derivatives. *Carbohydr. Polym.* 2017, *164*, 268–283.

40. Laokuldilok, T.; Potivas, T.; Kanha, N.; Surawang, S.; Seesuriyachan, P.; Wangtueai, S.; Phimolsiripol, Y.; Regenstein, J. M. Physicochemical, antioxidant, and antimicrobial properties of chitooligosaccharides produced using three different enzyme treatments. *Food Biosci.* **2017**, *18*, 28–33, doi:10.1016/j.fbio.2017.03.004.

41. Tsai, G. J.; Su, W. H.; Chen, H. C.; Pan, C. L. Antimicrobial activity of shrimp chitin and chitosan from different treatments and applications of fish preservation. *Fish. Sci.* **2002**, *68*, 170–177, doi:10.1046/j.1444-2906.2002.00404.x.

42. Patrulea, V.; Ostafe, V.; Borchard, G.; Jordan, O. Chitosan as a starting material for wound healing applications. *Eur. J. Pharm. Biopharm.* 2015, *97*, 417–426.

43. Ahmed, T. A.; Aljaeid, B. M. Drug Design, Development and Therapy Dovepress Preparation, characterization, and potential application of chitosan, chitosan derivatives, and chitosan metal nanoparticles in pharmaceutical drug delivery. *Drug Des. Devel. Ther.* **2016**, 10–483, doi:10.2147/DDDT.S99651.

44. Li, J.; Cai, C.; Li, J.; Li, J.; Li, J.; Sun, T.; Wang, L.; Wu, H.; Yu, G. Chitosan-based nanomaterials for drug delivery. *Molecules* 2018, *23*.

45. Hayashi, K.; Ito, M. Antidiabetic action of low molecular weight chitosan in genetically obese diabetic KK-Ay mice. *Biol. Pharm. Bull.* **2002**, *25*, 188–192, doi:10.1248/bpb.25.188.

46. Niaz, T.; Shabbir, S.; Manzoor, S.; Rehman, A.; Rahman, A.; Nasir, H.; Imran, M. Antihypertensive nano-ceuticales based on chitosan biopolymer: Physico-chemical evaluation and release kinetics. *Carbohydr. Polym.* **2016**, *142*, 268–274, doi:10.1016/j.carbpol.2016.01.047.

47. Jiang, T.; Ji, H.; Zhang, L.; Wang, Y.; Zhou, H. Chitosan Oligosaccharide Exerts Anti-Allergic Effect against Shrimp Tropomyosin-Induced Food Allergy by Affecting Th1 and Th2 Cytokines. *Int. Arch. Allergy Immunol.* **2019**, *180*, 10–16, doi:10.1159/000500720.

48. Sarvaiya, J.; Agrawal, Y. K. Chitosan as a suitable nanocarrier material for anti-Alzheimer drug delivery. *Int. J. Biol. Macromol.* 2015, *72*, 454–465.

49. Kulikov, S. N.; Chirkov, S. N.; Il'ina, A. V.; Lopatin, S. A.; Varlamov, V. P. Effect of the molecular weight of chitosan on its antiviral activity in plants. *Prikl. biokhimiia i Mikrobiol.* **2006**, *42*, 224–228.

50. Das, S. N.; Madhuprakash, J.; Sarma, P. V. S. R. N.; Purushotham, P.; Suma, K.; Manjeet, K.; Rambabu, S.; Gueddari, N. E. El; Moerschbacher, B. M.; Podile, A. R. Biotechnological approaches for field applications of chitooligosaccharides (COS) to induce innate immunity in plants. *Crit. Rev. Biotechnol.* **2015**, *35*, 29–43, doi:10.3109/07388551.2013.798255.

51. Basa, S.; Nampally, M.; Honorato, T.; Das, S. N.; Bovdilova, A.; Podile, A. R.; El Gueddari, N. E.; Moerschbacher, B. M. The pattern of acetylation defines priming activity of chitosan tetramers in rice. 1–28.

52. Basa, S.; Nampally, M.; Honorato, T.; Das, S. N.; Podile, A. R.; Gueddari, N. E. El; Moerschbacher, B. M. The Pattern of Acetylation Defines the Priming Activity of Chitosan Tetramers. *Cite This J. Am. Chem. Soc* **2020**, *142*, doi:10.1021/jacs.9b11466.

53. Akter Mukta, J.; Rahman, M.; As Sabir, A.; Gupta, D. R.; Surovy, M. Z.; Rahman, M.; Islam, M. T. Chitosan and plant probiotics application enhance growth and yield of strawberry. *Biocatal. Agric. Biotechnol.* **2017**, *11*, 9–18, doi:10.1016/j.bcab.2017.05.005.

54. Choay, J.; Petitou, M.; Lormeau, J. C.; Sinaÿ, P.; Casu, B.; Gatti, G. Structure-activity

relationship in heparin: A synthetic pentasaccharide with high affinity for antithrombin III and eliciting high anti-factor Xa activity. *Biochem. Biophys. Res. Commun.* **1983**, *116*, 492–499, doi:10.1016/0006-291X(83)90550-8.

55. McDougall, G. J.; Fry, S. C. Structure-Activity Relationships for Xyloglucan Oligosaccharides with Antiauxin Activity. *Plant Physiol.* **1989**, *89*, 883–887, doi:10.1104/pp.89.3.883.

56. Zhang, X.; Li, K.; Liu, S.; Xing, R.; Yu, H.; Chen, X.; Li, P. Size effects of chitooligomers on the growth and photosynthetic characteristics of wheat seedlings. *Carbohydr. Polym.* **2016**, *138*, 27–33, doi:10.1016/j.carbpol.2015.11.050.

57. Li, K.; Xing, R.; Liu, S.; Qin, Y.; Yu, H.; Li s, P. Size and pH effects of chitooligomers on antibacterial activity against Staphylococcus aureus. *Int. J. Biol. Macromol.* **2014**, *64*, 302–305, doi:10.1016/J.IJBIOMAC.2013.11.037.

58. Zhang, X.; Li, K.; Liu, S.; Zou, P.; Xing, R.; Yu, H.; Chen, X.; Qin, Y.; Li, P. Relationship between the degree of polymerization of chitooligomers and their activity affecting the growth of wheat seedlings under salt stress. *J. Agric. Food Chem.* **2017**, *65*, 501–509, doi:10.1021/acs.jafc.6b03665.

59. Xiong, C.; Wu, H.; Wei, P.; Pan, M.; Tuo, Y.; Kusakabe, I.; Du, Y. Potent angiogenic inhibition effects of deacetylated chitohexaose separated from chitooligosaccharides and its mechanism of action in vitro. *Carbohydr. Res.* **2009**, *344*, 1975–1983, doi:10.1016/j.carres.2009.06.036.

60. Zou, P.; Tian, X.; Dong, B.; Zhang, C. Size effects of chitooligomers with certain degrees of polymerization on the chilling tolerance of wheat seedlings. *Carbohydr. Polym.* **2017**, *160*, 194–202, doi:10.1016/j.carbpol.2016.12.058.

61. Djordjevic, M. A.; Bezos, A.; Susanti; Marmuse, L.; Driguez, H.; Samain, E.; Vauzeilles, B.; Beau, J. M.; Kordbacheh, F.; Rolfe, B. G.; Schwörer, R.; Daines, A. M.; Gresshoff, P. M.; Parish, C. R. Lipo-chitin oligosaccharides, plant symbiosis signalling molecules that modulate mammalian angiogenesis in vitro. *PLoS One* **2014**, *9*, doi:10.1371/journal.pone.0112635.

62. Li, K.; Xing, R.; Liu, S.; Li, R.; Qin, Y.; Meng, X.; Li, P. Separation of chito-oligomers with several degrees of polymerization and study of their antioxidant activity. *Carbohydr. Polym.* **2012**, *88*, 896–903, doi:10.1016/j.carbpol.2012.01.033.

63. Madhavan, P.; Ramachandran Nair, K. G. *Utilization of prawn waste: isolation of chitin and its conversion to chitosan*;

64. Muxika, A.; Etxabide, A.; Uranga, J.; Guerrero, P.; de la Caba, K. Chitosan as a bioactive polymer: Processing, properties and applications. *Int. J. Biol. Macromol.*

2017, *105*, 1358–1368.

65. Elieh-Ali-Komi, D.; Hamblin, M. R. Chitin and Chitosan: Production and Application of Versatile Biomedical Nanomaterials. *Int. J. Adv. Res.* **2016**, *4*, 411–427.

66. Bonfante-Alvarez, H.; De Avila-Montiel, G.; Herrera-Barros, A.; Torrenegra-Alarcón, M.; Darío González-Delgado, Á. Evaluation of Five Chitosan Production Routes with Astaxanthin Recovery from Shrimp Exoskeletons. In *CHEMICAL ENGINEERING TRANSACTIONS*; 2018; Vol. 70.

67. Dhillon, G. S.; Kaur, S.; Brar, S. K.; Verma, M. Green synthesis approach: Extraction of chitosan from fungus mycelia. *Crit. Rev. Biotechnol.* 2013, *33*, 379–403.

68. Kaur, S.; Dhillon, G. S. Recent trends in biological extraction of chitin from marine shell wastes: A review. *Crit. Rev. Biotechnol.* 2015, *35*, 44–61.

69. Percot, A.; Viton, C.; Domard, A. Characterization of shrimp shell deproteinization. *Biomacromolecules* **2003**, *4*, 1380–1385, doi:10.1021/bm034115h.

70. No, H. K.; Lee, K. S.; Meyers, S. P. Correlation between physicochemical characteristics and binding capacities of chitosan products. *J. Food Sci.* **2000**, *65*, 1134–1137, doi:10.1111/j.1365-2621.2000.tb10252.x.

71. Antonino, R. S. C. M. D. Q.; Fook, B. R. P. L.; Lima, V. A. D. O.; Rached, R. Í. D. F.; Lima, E. P. N.; Lima, R. J. D. S.; Covas, C. A. P.; Fook, M. V. L. Preparation and characterization of chitosan obtained from shells of shrimp (Litopenaeus vannamei Boone). *Mar. Drugs* **2017**, *15*, doi:10.3390/md15050141.

72. No, H. K.; Meyers, S. P. Preparation and characterization of chitin and chitosan- a review. *J. Aquat. Food Prod. Technol.* **1995**, *4*, 27–52, doi:10.1300/J030v04n02_03.

73. Hao, W.; Li, K.; Li, P. Review: Advances in preparation of chitooligosaccharides with heterogeneous sequences and their bioactivity. *Carbohydr. Polym.* 2021, *252*.

74. Rupley, J. A. The hydrolysis of chitin by concentrated hydrochloric acid, and the preparation of low-molecular-weight substrate for lysozyme. *BBA - Spec. Sect. Mucoproteins Mucopolysaccharides* **1964**, *83*, 245–255, doi:10.1016/0926-6526(64)90001-1.

75. Jia, Z.; Shen, D. Effect of reaction temperature and reaction time on the preparation of low-molecular-weight chitosan using phosphoric acid. *Carbohydr. Polym.* **2002**, *49*, 393–396, doi:10.1016/S0144-8617(02)00026-7.

76. Chang, K. L. B.; Tai, M. C.; Cheng, F. H. Kinetics and products of the degradation of

chitosan by hydrogen peroxide. *J. Agric. Food Chem.* **2001**, *49*, 4845–4851, doi:10.1021/jf001469g.

77. Yang, Y.; Yu, B. Recent advances in the synthesis of chitooligosaccharides and congeners. *Tetrahedron* 2014, *70*, 1023–1046.

78. Aly, M. R. E.; Ibrahim, E. S. I.; El Ashry, E. S. H.; Schmidt, R. R. Synthesis of chitotetraose and chitohexaose based on dimethylmaleoyl protection. *Carbohydr. Res.* **2001**, *331*, 129–142, doi:10.1016/S0008-6215(01)00024-6.

79. Kröck, L.; Esposito, D.; Castagner, B.; Wang, C. C.; Bindschädler, P.; Seeberger, P. H. Streamlined access to conjugation-ready glycans by automated synthesis. *Chem. Sci.* **2012**, *3*, 1617–1622, doi:10.1039/c2sc00940d.

80. Ohnuma, T.; Fukuda, T.; Dozen, S.; Honda, Y.; Kitaoka, M.; Fukamizo, T. A glycosynthase derived from an inverting GH19 chitinase from the moss Bryum coronatum. *Biochem. J.* **2012**, *444*, 437–443, doi:10.1042/BJ20120036.

81. Alsina Verdú, C. Enginyeria de glicosintases derivades de quitinases GH18 per a la polimerització de quitooligosacàrids. *TDX (Tesis Dr. en Xarxa)* **2019**.

82. G. Skjak-Braek, T. Anthonsen, P. A. S. Chitin and Chitosan: Sources, Chemistry, Biochemistry, Physical Properties. Available online: https://books.google.es/books/about/Chitin_and_Chitosan.html?id=mUIRAAAAMAAJ&redir_esc=y (accessed on May 19, 2021).

83. Muthu, M.; Gopal, J.; Chun, S.; Devadoss, A. J. P.; Hasan, N.; Sivanesan, I. Crustacean waste-derived chitosan: Antioxidant properties and future perspective. *Antioxidants* **2021**, *10*, 1–27, doi:10.3390/antiox10020228.

84. Cheung, R. C. F.; Ng, T. B.; Wong, J. H.; Chan, W. Y. *Chitosan: An Update on Potential Biomedical and Pharmaceutical Applications.*; 2015; Vol. 13; ISBN 8523943803.

85. Hembach, L.; Cord-Landwehr, S.; Moerschbacher, B. M. Enzymatic production of all fourteen partially acetylated chitosan tetramers using different chitin deacetylases acting in forward or reverse mode. *Sci. Rep.* **2017**, *7*, 17692, doi:10.1038/s41598-017-17950-6.

86. Cantarel, B. I.; Coutinho, P. M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B. The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Res.* **2009**, *37*, 233–238, doi:10.1093/nar/gkn663.

87. Planas, A.; Faijes, M.; Codera, V. When Enzymes Do It Better: Enzymatic

Glycosylation Methods. In *Carbohydrate Chemistry: State of the Art and Challenges for Drug Development*; IMPERIAL COLLEGE PRESS, 2015; pp. 215–245.

88. Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P. M.; Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **2014**, *42*, doi:10.1093/nar/gkt1178.

89. Bürger, M.; Chory, J. Structural and chemical biology of deacetylases for carbohydrates, proteins, small molecules and histones. *Commun. Biol.* 2018, *1*, 1–11.

90. Nakamura, A. M.; Nascimento, A. S.; Polikarpov, I. Structural diversity of carbohydrate esterases. *Biotechnol. Res. Innov.* **2017**, *1*, 35–51, doi:10.1016/j.biori.2017.02.001.

91. Pascual, S.; Planas, A. Carbohydrate de-N-acetylases acting on structural polysaccharides and glycoconjugates. *Curr. Opin. Chem. Biol.* 2021, *61*, 9–18.

92. Bamford, N. C.; Le Mauff, F.; Van Loon, J. C.; Ostapska, H.; Snarr, B. D.; Zhang, Y.; Kitova, E. N.; Klassen, J. S.; Codée, J. D. C.; Sheppard, D. C.; Howell, P. L. Structural and biochemical characterization of the exopolysaccharide deacetylase Agd3 required for Aspergillus fumigatus biofilm formation. *Nat. Commun.* **2020**, *11*, 2450, doi:10.1038/s41467-020-16144-5.

93. Blair, D. E.; Schüttelkopf, A. W.; MacRae, J. I.; Van Aalten, D. M. F. Structure and metal-dependent mechanism of peptidoglycan deacetylase, a streptococcal virulence factor. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 15429–15434, doi:10.1073/pnas.0504339102.

94. Andres, E.; Albesa-Jove, D.; Biarnes, X.; Moerschbacher, B. M.; Guerin, M. E.; Planas, A. Structural basis of chitin oligosaccharide deacetylation. *Angew. Chemie - Int. Ed.* **2014**, *53*, 6882–6887, doi:10.1002/anie.201400220.

95. Ahangar, M. S.; Furze, C. M.; Guy, C. S.; Cooper, C.; Maskew, K. S.; Graham, B.; Cameron, A. D.; Fullam, X. E. Structural and functional determination of homologs of the Mycobacterium tuberculosis N-acetylglucosamine-6-phosphate deacetylase (NagA). *J. Biol. Chem.* **2018**, *293*, 9770–9783, doi:10.1074/jbc.RA118.002597.

96. Pascual, S.; Planas, A. Carbohydrate de-N-acetylases acting on structural polysaccharides and glycoconjugates. *Curr. Opin. Chem. Biol.* **2021**, *61*, doi:10.1016/j.cbpa.2020.09.003.

97. John, M.; Rohrig, H.; Schmidt, J.; Wieneke, U.; Schell, J. Rhizobium NodB protein involved in nodulation signal synthesis is a chitooligosaccharide deacetylase. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 625–629, doi:10.1073/pnas.90.2.625.

98. Blair, D. E.; Schüttelkopf, A. W.; MacRae, J. I.; van Aalten, D. M. F. Structure and metal-dependent mechanism of peptidoglycan deacetylase, a streptococcal virulence factor. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 15429–34, doi:10.1073/pnas.0504339102.

99. Blair, D. E.; Hekmat, O.; Schüttelkopf, A. W.; Shrestha, B.; Tokuyasu, K.; Withers, S. G.; Van Aalten, D. M. F. Structure and Mechanism of Chitin Deacetylase from the Fungal Pathogen Colletotrichum lindemuthianum †, ‡. **2006**, doi:10.1021/bi0606694.

100. Aragunde, H.; Biarnés, X.; Planas, A. Substrate recognition and specificity of chitin deacetylases and related family 4 carbohydrate esterases. *Int. J. Mol. Sci.* **2018**, *19*, 1–30, doi:10.3390/ijms19020412.

101. Schmitz, C.; Auza, L. G.; Koberidze, D.; Rasche, S.; Fischer, R.; Bortesi, L. Conversion of chitin to defined chitosan oligomers: Current status and future prospects. *Mar. Drugs* 2019, *17*.

102. Naqvi, S.; Moerschbacher, B. M. The cell factory approach toward biotechnological production of high-value chitosan oligomers and their derivatives: an update. *Crit. Rev. Biotechnol.* 2017, *37*, 11–25.

103. Cord-Landwehr, S.; Melcher, R. L. J.; Kolkenbrock, S.; Moerschbacher, B. M. A chitin deacetylase from the endophytic fungus Pestalotiopsis sp. efficiently inactivates the elicitor activity of chitin oligomers in rice cells OPEN. *Nat. Publ. Gr.* **2016**, doi:10.1038/srep38018.

104. Aragunde Pazos, H. Ingeniería enzimática de quitina desacetilasas y glicosintasas como biocatalizadores: diseño racional de la especificidad y evolución dirigida. *TDX (Tesis Dr. en Xarxa)* **2017**.

105. Wang, Y.; Niu, X.; Guo, X.; Yu, H.; Liu, Z.; Zhang, Z.; Yuan, S. Heterologous expression, characterization and possible functions of the chitin deacetylases, Cda1 and Cda2, from mushroom Coprinopsis cinerea. *Glycobiology* **2018**, *28*, 318–332, doi:10.1093/glycob/cwy007.

106. Bai, Y.; Wang, Y.; Liu, X.; Zhao, J.; Kang, L. Q.; Liu, Z.; Yuan, S. Heterologous expression and characterization of a novel chitin deacetylase, CDA3, from the mushroom Coprinopsis cinerea. *Int. J. Biol. Macromol.* **2020**, *150*, 536–545, doi:10.1016/j.ijbiomac.2020.02.083.

107. Xie, M.; Zhao, X.; Yang, L.; Jin, C. Chitin deacetylases Cod4 and Cod7 are involved in polar growth of Aspergillus fumigatus. *Microbiologyopen* **2020**, *9*, doi:10.1002/mbo3.943.

108. Hembach, L.; Bonin, M.; Gorzelanny, C.; Moerschbacher, B. M. Unique subsite

specificity and potential natural function of a chitosan deacetylase from the human pathogen Cryptococcus neoformans. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 3551–3559, doi:10.1073/pnas.1915798117.

109. Lam, W. C.; Upadhya, R.; Specht, C. A.; Ragsdale, A. E.; Hole, C. R.; Levitz, S. M.; Lodge, J. K. Chitosan Biosynthesis and Virulence in the Human Fungal Pathogen Cryptococcus gattii . *mSphere* **2019**, *4*, doi:10.1128/msphere.00644-19.

110. Shao, Z.; Thomas, Y.; Hembach, L.; Xing, X.; Duan, D.; Moerschbacher, B. M.; Bulone, V.; Tirichine, L.; Bowler, C. Comparative characterization of putative chitin deacetylases from Phaeodactylum tricornutum and Thalassiosira pseudonana highlights the potential for distinct chitin-based metabolic processes in diatoms. *New Phytol.* **2019**, *221*, 1890–1905, doi:10.1111/nph.15510.

111. Ye, W.; Ma, H.; Liu, L.; Yu, J.; Lai, J.; Fang, Y.; Fan, Y. Biocatalyzed route for the preparation of surface-deacetylated chitin nanofibers. *Green Chem.* **2019**, *21*, 3143–3151, doi:10.1039/c9gc00857h.

112. Bhat, P.; Pawaskar, G. M.; Raval, R.; Cord-Landwehr, S.; Moerschbacher, B.; Raval, K. Expression of Bacillus licheniformis chitin deacetylase in E. coli pLysS: Sustainable production, purification and characterisation. *Int. J. Biol. Macromol.* **2019**, *131*, 1008–1013, doi:10.1016/j.ijbiomac.2019.03.144.

113. Ma, Q.; Gao, X.; Bi, X.; Tu, L.; Xia, M.; Shen, Y.; Wang, M. Isolation, characterisation, and genome sequencing of Rhodococcus equi: a novel strain producing chitin deacetylase. *Sci. Rep.* **2020**, *10*, 1–9, doi:10.1038/s41598-020-61349-9.

114. Chai, J.; Hang, J.; Zhang, C.; Yang, J.; Wang, S.; Liu, S.; Fang, Y. Purification and characterization of chitin deacetylase active on insoluble chitin from Nitratireductor aquimarinus MCDA3-3. *Int. J. Biol. Macromol.* **2020**, *152*, 922–929, doi:10.1016/j.ijbiomac.2020.02.308.

115. Liu, X.; Cooper, A. M. W.; Zhang, J.; Zhu, K. Y. Biosynthesis, modifications and degradation of chitin in the formation and turnover of peritrophic matrix in insects. *J. Insect Physiol.* 2019, *114*, 109–115.

116. Yu, R. R.; Liu, W. M.; Zhao, X. M.; Zhang, M.; Li, D. Q.; Zuber, R.; Ma, E. B.; Zhu, K. Y.; Moussian, B.; Zhang, J. Z. LmCDA1 organizes the cuticle by chitin deacetylation in Locusta migratoria. *Insect Mol. Biol.* **2019**, *28*, 301–312, doi:10.1111/imb.12554.

117. Zhang, M.; Ji, Y.; Zhang, X.; Ma, P.; Wang, Y.; Moussian, B.; Zhang, J. The putative chitin deacetylases Serpentine and Vermiform have non-redundant functions during Drosophila wing development. *Insect Biochem. Mol. Biol.* **2019**, *110*, 128–135, doi:10.1016/j.ibmb.2019.05.008.

118. Liu, Y.; Yang, J.; Yao, L.; Li, S.; Chen, Y.; Yang, H.; Fan, D. Chitin deacetylase: A potential target for Mythimna separata (Walker) control. *Arch. Insect Biochem. Physiol.* **2020**, *104*, e21666, doi:10.1002/arch.21666.

119. Liu, L.; Zhou, Y.; Qu, M.; Qiu, Y.; Guo, X.; Zhang, Y.; Liu, T.; Yang, J.; Yang, Q. Structural and biochemical insights into the catalytic mechanisms of two insect chitin deacetylases of the carbohydrate esterase 4 family. *J. Biol. Chem.* **2019**, *294*, 5774–5783, doi:10.1074/jbc.RA119.007597.

120. Liu, L.; Qu, M.; Liu, T.; Chen, Q.; Guo, X.; Yang, J.; Yang, Q. Biochemical characterization of three midgut chitin deacetylases of the Lepidopteran insect Bombyx mori. *J. Insect Physiol.* **2019**, *113*, 42–48, doi:10.1016/j.jinsphys.2019.01.005.

121. Colwell, R. R. Global climate and infectious disease: The cholera paradigm. In *Science*; American Association for the Advancement of Science, 1996; Vol. 274, pp. 2025–2031.

122. Ohishi, K.; Yamagishi, M.; Ohta, T.; Motosugi, M.; Izumida, H.; Sano, H.; Adachi, K.; Miwa, T. Purification and Properties of Two Deacetylases Produced by Vibrio alginolyticus H-8. *Biosci. Biotechnol. Biochem.* **1997**, *61*, 1113–1117, doi:10.1271/bbb.61.1113.

123. Kadokura, K.; Rokutani, A.; Yamamoto, M.; Ikegami, T.; Sugita, H.; Itoi, S.; Hakamata, W.; Oku, T.; Nishio, T. Purification and characterization of Vibrio parahaemolyticus extracellular chitinase and chitin oligosaccharide deacetylase involved in the production of heterodisaccharide from chitin. *Appl. Microbiol. Biotechnol.* **2007**, *75*, 357–365, doi:10.1007/s00253-006-0831-6.

124. Li, X.; Wang, L. X.; Wang, X.; Roseman, S. The chitin catabolic cascade in the marine bacterium Vibrio cholerae: Characterization of a unique chitin oligosaccharide deacetylase. *Glycobiology* **2007**, *17*, 1377–1387, doi:10.1093/glycob/cwm096.

125. Hirano, T.; Kadokura, K.; Ikegami, T.; Shigeta, Y.; Kumaki, Y.; Hakamata, W.; Oku, T.; Nishio, T. Heterodisaccharide 4-O-(N-acetyl-??-D-glucosaminyl)-D-glucosamine is a specific inducer of chitinolytic enzyme production in Vibrios harboring chitin oligosaccharide deacetylase genes. *Glycobiology* **2009**, *19*, 1046–1053, doi:10.1093/glycob/cwp088.

126. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*; Humana Press, 2005; pp. 571–607.

127. Li, X.; Wang, L. X.; Wang, X.; Roseman, S. The chitin catabolic cascade in the marine bacterium Vibrio cholerae: Characterization of a unique chitin oligosaccharide

deacetylase. *Glycobiology* **2007**, *17*, 1377–1387, doi:10.1093/glycob/cwm096.

128. Davies, G. J.; Wilson, K. S.; Henrissat, B. Nomenclature for sugar-binding subsites in glycosyl hydrolases [1]. *Biochem. J.* 1997, *321*, 557–559.

129. Tokuyasu, K.; Mitsutomi, M.; Yamaguchi, I.; Hayashi, K.; Mori, Y. Recognition of chitooligosaccharides and their N-acetyl groups by putative subsites of chitin deacetylase from a Deuteromycete, Colletotrichum lindemuthianum. *Biochemistry* **2000**, *39*, 8837–8843, doi:10.1021/bi0005355.

130. Tuveng, T. R.; Rothweiler, U.; Udatha, G.; Vaaje-Kolstad, G.; Smalås, A.; Eijsink, V. G. H. Structure and function of a CE4 deacetylase isolated from a marine environment. *PLoS One* **2017**, *12*, doi:10.1371/journal.pone.0187544.

131. Schmidt-Dannert, C.; Arnold, F. H. Directed evolution of industrial enzymes. In *Trends in Biotechnology*; Trends Biotechnol, 1999; Vol. 17, pp. 135–136.

132. Cherry, J. R.; Fidantsef, A. L. Directed evolution of industrial enzymes: An update. *Curr. Opin. Biotechnol.* 2003, *14*, 438–443.

133. Romero, P. A.; Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* 2009, *10*, 866–876.

134. Rubingh, D. N. Protein engineering from a bioindustrial point of view. *Curr. Opin. Biotechnol.* **1997**, *8*, 417–422, doi:10.1016/S0958-1669(97)80062-6.

135. Craik, C. S.; Largman, C.; Fletcher, T.; Roczniak, S.; Barr, P. J.; Fletterick, R.; Rutter, W. J. Redesigning trypsin: alteration of substrate specificity. *Science (80-.).* **1985**, *228*, 291–297, doi:10.1126/science.3838593.

136. Yang, K. K.; Wu, Z.; Arnold, F. H. Machine-learning-guided directed evolution for protein engineering. *Nat. Methods* 2019, *16*, 687–694.

137. Pokala, N.; Handel, T. M. Review: Protein design - Where we were, where we are, where we're going. *J. Struct. Biol.* 2001, *134*, 269–281.

138. Marshall, S. A.; Lazar, G. A.; Chirino, A. J.; Desjarlais, J. R. Rational design and engineering of therapeutic proteins. *Drug Discov. Today* 2003, *8*, 212–221.

139. Packer, M. S.; Liu, D. R. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* 2015, *16*, 379–394.

140. Sayous, V.; Lubrano, P.; Li, Y.; Acevedo-Rocha, C. G. Unbiased libraries in protein

directed evolution. *Biochim. Biophys. Acta - Proteins Proteomics* 2020, *1868*.

141. Gargiulo, S.; Soumillion, P. Directed evolution for enzyme development in biocatalysis. *Curr. Opin. Chem. Biol.* **2021**, *61*, 107–113, doi:10.1016/j.cbpa.2020.11.006.

142. Bergeron, L. M.; Gomez, L.; Whitehead, T. A.; Clark, D. S. Self-renaturing enzymes: Design of an enzyme-chaperone chimera as a new approach to enzyme stabilization. *Biotechnol. Bioeng.* **2009**, *102*, 1316–1322, doi:10.1002/bit.22254.

143. Munier-Lehmann, H.; Chaffotte, A.; Pochet, S.; Labesse, G. Thymidylate kinase of Mycobacterium tuberculosis: A chimera sharing properties common to eukaryotic and bacterial enzymes. *Protein Sci.* **2001**, *10*, 1195–1205, doi:10.1110/ps.45701.

144. Rattray, A. J.; Strathern, J. N. Error-prone DNA polymerases: when making a mistake is the only way to get ahead. *Annu. Rev. Genet.* **2003**, *37*, 31–66, doi:10.1146/annurev.genet.37.042203.132748.

145. Reetz, M. T.; Kahakeaw, D.; Lohmer, R. Addressing the numbers problem in directed evolution. *ChemBioChem* **2008**, *9*, 1797–1804, doi:10.1002/cbic.200800298.

146. Arnold, F. H.; Wintrode, P. L.; Miyazaki, K.; Gershenson, A. How enzymes adapt: Lessons from directed evolution. *Trends Biochem. Sci.* 2001, *26*, 100–106.

147. Huimin Zhao and Frances H Arnold Combinatorial protein design: strategies for screening protein libraries. *Curr. Opin. Struct. Biol.* 480–485.

148. Renata, H.; Wang, Z. J.; Arnold, F. H. Expanding the enzyme universe: Accessing non-natural reactions by mechanism-guided directed evolution. *Angew. Chemie - Int. Ed.* 2015, *54*, 3351–3367.

149. Wilson, D. S.; Keefe, A. D. Random mutagenesis by PCR. *Curr. Protoc. Mol. Biol.* **2001**, *Chapter 8*, doi:10.1002/0471142727.MB0803S51.

150. Ness, J. E.; Welch, M.; Giver, L.; Bueno, M.; Cherry, J. R.; Borchert, T. V.; Stemmer, W. P. C.; Minshull, J. DNA shuffling of subgenomic sequences of subtilisin. *Nat. Biotechnol.* **1999**, doi:10.1038/12884.

151. Stemmer, W. P. C. Rapid evolution of a protein in vitro by DNA shuffling. *Nat. 1994 3706488* **1994**, *370*, 389–391, doi:10.1038/370389a0.

152. Neylon, C. Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution. *Nucleic Acids Res.* **2004**, *32*, 1448, doi:10.1093/NAR/GKH315.

153. Chronopoulou, E. G.; Labrou, N. E. Site-saturation mutagenesis: A powerful tool for structure-based design of combinatorial mutation libraries. *Curr. Protoc. Protein Sci.* **2011**, 1–10, doi:10.1002/0471140864.ps2606s63.

154. Cunningham, B. C.; Wells, J. A. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science* **1989**, *244*, 1081–1085, doi:10.1126/SCIENCE.2471267.

155. Marsischky, G.; LaBaer, J. Many paths to many clones: A comparative look at high-throughput cloning methods. *Genome Res.* 2004, *14*, 2020–2028.

156. Celie, P. H. N.; Parret, A. H. A.; Perrakis, A. Recombinant cloning strategies for protein expression. *Curr. Opin. Struct. Biol.* 2016, *38*, 145–154.

157. Macauley-Patrick, S.; Fazenda, M. L.; McNeil, B.; Harvey, L. M. Heterologous protein production using the Pichia pastoris expression system. *Yeast* 2005, *22*, 249–270.

158. Lindenburg, L.; Huovinen, T.; van de Wiel, K.; Herger, M.; Snaith, M. R.; Hollfelder, F. Split & mix assembly of DNA libraries for ultrahigh throughput on-bead screening of functional proteins. *Nucleic Acids Res.* **2020**, *48*, E63, doi:10.1093/NAR/GKAA270.

159. Muranaka, N.; Sharma, V.; Nomura, Y.; Yokobayashi, Y. An efficient platform for genetic selection and screening of gene switches in Escherichia coli. *Nucleic Acids Res.* **2009**, *37*, 39, doi:10.1093/nar/gkp039.

160. Dixit, M.; Panchal, K.; Pandey, D.; Labrou, N. E.; Shukla, P. Robotics for enzyme technology: innovations and technological perspectives. *Appl. Microbiol. Biotechnol.* **2021**, 1–9, doi:10.1007/s00253-021-11302-1.

161. Victorino da Silva Amatto, I.; Gonsales da Rosa-Garzon, N.; Antônio de Oliveira Simões, F.; Santiago, F.; Pereira da Silva Leite, N.; Raspante Martins, J.; Cabral, H. Enzyme engineering and its industrial applications. *Biotechnol. Appl. Biochem.* 2021.

162. Yang, G.; Withers, S. G. Ultrahigh-throughput FACS-based screening for directed enzyme evolution. *ChemBioChem* 2009, *10*, 2704–2715.

163. Garrabou, X.; Macdonald, D. S.; Wicky, B. I. M.; Hilvert, D. Stereodivergent Evolution of Artificial Enzymes for the Michael Reaction. *Angew. Chemie - Int. Ed.* **2018**, *57*, 5288–5291, doi:10.1002/anie.201712554.

164. FW, S. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **2005**, *41*, 207–234, doi:10.1016/J.PEP.2005.01.016.

165. Fatima, B. Quantitative Analysis by IR: Determination of Chitin/Chitosan DD. *Mod. Spectrosc. Tech. Appl.* **2020**, doi:10.5772/intechopen.89708.

166. Tan, S. C.; Khor, E.; Tan, T. K.; Wong, S. M. The degree of deacetylation of chitosan: Advocating the first derivative UV-spectrophotometry method of determination. *Talanta* **1998**, *45*, 713–719, doi:10.1016/S0039-9140(97)00288-9.

167. Hirai, A.; Odani, H.; Nakajima, A. Determination of degree of deacetylation of chitosan by 1H NMR spectroscopy. *Polym. Bull.* **1991**, *26*, 87–94, doi:10.1007/BF00299352.

168. Jiang, X.; Chen, L.; Zhong, W. A new linear potentiometric titration method for the determination of deacetylation degree of chitosan. *Carbohydr. Polym.* **2003**, *54*, 457–463, doi:10.1016/j.carbpol.2003.05.004.

169. Souza, J. M.; Plumbridge, J. A.; Calcagno, M. L. N-acetylglucosamine-6-phosphate deacetylase from Escherichia coli: Purification and molecular and kinetic characterization. *Arch. Biochem. Biophys.* **1997**, *340*, 338–346, doi:10.1006/abbi.1997.9780.

170. Kauss, H.; Bauch, B. Chitin deacetylase from Colletotrichum lindemuthianum. *Methods Enzymol.* **1988**, *161*, 518–523, doi:10.1016/0076-6879(88)61070-6.

171. Tang, M. C.; Nisole, A.; Dupont, C.; Pelletier, J. N.; Waldron, K. C. Chemical profiling of the deacetylase activity of acetyl xylan esterase A (AxeA) variants on chitooligosaccharides using hydrophilic interaction chromatography-mass spectrometry. *J. Biotechnol.* **2011**, *155*, 257–265, doi:10.1016/j.jbiotec.2011.06.041.

172. Muraki, E.; Yaku, F.; Kojima, H. Preparation and crystallization of d-glucosamine oligosaccharides with dp 6-8. *Carbohydr. Res.* **1993**, *239*, 227–237, doi:10.1016/0008-6215(93)84217-T.

173. Cord-Landwehr, S.; Ihmor, P.; Niehues, A.; Luftmann, H.; Moerschbacher, B. M.; Mormann, M. Quantitative Mass-Spectrometric Sequencing of Chitosan Oligomers Revealing Cleavage Sites of Chitosan Hydrolases. **2017**, doi:10.1021/acs.analchem.6b04183.

174. Pacheco, N.; Trombotto, S.; David, L.; Shirai, K. Activity of chitin deacetylase from Colletotrichum gloeosporioides on chitinous substrates. *Carbohydr. Polym.* **2013**, *96*, 227–232, doi:10.1016/j.carbpol.2013.03.051.

175. Prochazkova, S.; Vårum, K. M.; Ostgaard, K. Quantitative determination of chitosans by ninhydrin. *Carbohydr. Polym.* **1999**, *38*, 115–122, doi:10.1016/S0144-8617(98)00108-8.

176. Pascual, S.; Planas, A. Screening Assay for Directed Evolution of Chitin Deacetylases: Application to Vibrio cholerae Deacetylase Mutant Libraries for Engineered Specificity. *Anal. Chem.* **2018**, *90*, doi:10.1021/acs.analchem.8b02729.

177. Cord-Landwehr, S.; Melcher, R. L. J.; Kolkenbrock, S.; Moerschbacher, B. M. A chitin deacetylase from the endophytic fungus Pestalotiopsis sp. efficiently inactivates the elicitor activity of chitin oligomers in rice cells. *Sci. Rep.* **2016**, *6*, 38018, doi:10.1038/srep38018.

178. Tsigos, I.; Zydowicz, N.; Martinou, A.; Domard, A.; Bouriotis, V. Mode of action of chitin deacetylase from Mucor rouxii on N-acetylchitooligosaccharides. *Eur. J. Biochem.* **1999**, *261*, 698–705.

179. Hattori, T.; Anraku, N.; Kato, R. Capillary electrophoresis of chitooligosaccharides in acidic solution: Simple determination using a quaternary-ammonium-modified column and indirect photometric detection with Crystal Violet. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **2010**, *878*, 477–480, doi:10.1016/j.jchromb.2009.11.042.

180. Ae, M. B. H.; Ekborg, N. A.; Weiner, R. M.; Steven, A. E.; Hutcheson, W. Detection and characterization of chitinases and other chitin-modifying enzymes., doi:10.1007/s10295-003-0096-3.

181. Liu, J.; Jia, Z.; Li, S.; Li, Y.; You, Q.; Zhang, C.; Zheng, X.; Xiong, G.; Zhao, J.; Qi, C.; Yang, J. Identification and characterization of a chitin deacetylase from a metagenomic library of deep-sea sediments of the Arctic Ocean. *Gene* **2016**, *590*, 79–84, doi:10.1016/j.gene.2016.06.007.

182. Bai, N.; Roder, H.; Dickson, A.; Karanicolas, J. Isothermal Analysis of ThermoFluor Data can readily provide Quantitative Binding Affinities. *Sci. Rep.* **2019**, *9*, 1–15, doi:10.1038/s41598-018-37072-x.

183. Wu, T.; Yu, J.; Gale-Day, Z.; Woo, A.; Suresh, A.; Hornsby, M.; Gestwicki, J. Three Essential Resources to Improve Differential Scanning Fluorimetry (DSF) Experiments. *bioRxiv* **2020**, 2020.03.22.002543, doi:10.1101/2020.03.22.002543.

184. Fontenete, S.; Guimarães, N.; Wengel, J.; Azevedo, N. F. Prediction of melting temperatures in fluorescence in situ hybridization (FISH) procedures using thermodynamic models. *Crit. Rev. Biotechnol.* **2016**, *36*, 566–577, doi:10.3109/07388551.2014.993589.

185. Andr??s, E.; Albesa-Jov??, D.; Biarn??s, X.; Moerschbacher, B. M.; Guerin, M. E.; Planas, A. Structural basis of chitin oligosaccharide deacetylation. *Angew. Chemie - Int. Ed.* **2014**, *53*, 6882–6887, doi:10.1002/anie.201400220.

186. Tuveng, T. R.; Rothweiler, U.; Udatha, G.; Vaaje-Kolstad, G.; Smalås, A.; Eijsink, V.

G. H.; 1 Structure and function of a CE4 deacetylase isolated from a marine environment. *PLoS One* **2017**, *12*, e0187544, doi:10.1002/cbin.10887.

187. Shao, W.; Wiegel, J. Purification and characterization of two thermostable acetyl xylan esterases from Thermoanaerobacterium sp. strain JW/SL-YS485. *Appl. Environ. Microbiol.* **1995**, *61*, 729–733.

188. Hekmat, O.; Tokuyasu, K.; Withers, S. G. Subsite structure of the endo-type chitin deacetylase from a deuteromycete, Colletotrichum lindemuthianum: an investigation using steady-state kinetic analysis and MS. *Biochem. J.* **2003**, *374*, 369–380, doi:10.1042/BJ20030204.

189. Mishra, C.; Mishra, C.; Semino, C.; Semino, C.; Mccreath, K. J.; Mccreath, K. J.; de la Vega, H.; de la Vega, H.; Jones, B. J.; Jones, B. J.; Specht, C. a; Specht, C. a; Robbins, P. W.; Robbins, P. W. Cloning and expression of two chitin deacetylase gens of Saccharomyces cerevisiae. **1997**, *13*, 327–336.

190. Tsigos, I.; Bouriotis, V. Purification and characterization of chitin deacetylase from Colletotrichum lindemuthianum. *J. Biol. Chem.* **1995**, *270*, 26286–26291, doi:10.1074/jbc.270.44.26286.

191. Naqvi, S.; Cord-Landwehr, S.; Singh, R.; Bernard, F.; Kolkenbrock, S.; El Gueddari, N. E.; Moerschbacher, B. M. A recombinant fungal chitin deacetylase produces fully defined chitosan oligomers with novel patterns of acetylation. *Appl. Environ. Microbiol.* **2016**, *82*, 6645–6655, doi:10.1128/AEM.01961-16.

192. Moynihan, P. J.; Clarke, A. J. Assay for peptidoglycan O-acetyltransferase: A potential new antibacterial target. *Anal. Biochem.* **2013**, *439*, 73–79, doi:10.1016/j.ab.2013.04.022.

193. Blair, D. E.; Schuttelkopf, A. W.; MacRae, J. I.; van Aalten, D. M. F. Structure and metal-dependent mechanism of peptidoglycan deacetylase, a streptococcal virulence factor. *Proc. Natl. Acad. Sci.* **2005**, *102*, 15429–15434, doi:10.1073/pnas.0504339102.

194. Mere, E.; Aguila, D.; Gomes, L. P.; Andrade, C. T.; Silva, J. T.; Margaret, V.; Paschoalin, F. Biocatalytic production of chitosan polymers from shrimp shells, using a recombinant enzyme produced by Pichia pastoris. *Am. J. Mol. Biol.* **2012**, *2*, 341–350, doi:10.4236/ajmb.2012.24035.

195. Leane, M. M.; Nankervis, R.; Smith, A.; Illum, L. Use of the ninhydrin assay to measure the release of chitosan from oral solid dosage forms. *Int. J. Pharm.* **2004**, *271*, 241–249, doi:10.1016/j.ijpharm.2003.11.023.

196. Cord-Landwehr, S.; Ihmor, P.; Niehues, A.; Luftmann, H.; Moerschbacher, B. M.;

Mormann, M. Quantitative Mass-Spectrometric Sequencing of Chitosan Oligomers Revealing Cleavage Sites of Chitosan Hydrolases. *Anal. Chem.* **2017**, *89*, 2893–2900, doi:10.1021/acs.analchem.6b04183.

197. Hirano, T.; Sugiyama, K.; Sakaki, Y.; Hakamata, W.; Park, S. Y.; Nishio, T. Structure-based analysis of domain function of chitin oligosaccharide deacetylase from Vibrio parahaemolyticus. *FEBS Lett.* **2015**, *589*, 145–151, doi:10.1016/j.febslet.2014.11.039.

198. Datta, S.; Christena, L. R.; Rajaram, Y. R. S. Enzyme immobilization: an overview on techniques and support materials. *3 Biotech* **2013**, *3*, 1–9, doi:10.1007/s13205-012-0071-7.

199. BS, C.; RR, M. Enzyme thermostabilization by bovine serum albumin and other proteins: evidence for hydrophobic interactions. *undefined* **1995**, *22*, 203–214, doi:10.1111/J.1470-8744.1995.TB00346.X.

200. Smith, P. K.; Krohn, R. I.; Hermanson, G. T.; Mallia, A. K.; Gartner, F. H.; Provenzano, M. D.; Fujimoto, E. K.; Goeke, N. M.; Olson, B. J.; Klenk, D. C. Measurement of protein using bicinchoninic acid. *Anal. Biochem.* **1985**, *150*, 76–85, doi:10.1016/0003-2697(85)90442-7.

201. Verma, S. C.; Mahadevan, S. The ChbG gene of the chitobiose (chb) operon of Escherichia coli encodes a chitooligosaccharide deacetylase. *J. Bacteriol.* **2012**, *194*, 4959–4971, doi:10.1128/JB.00533-12.

202. Quan, J.; Tian, J. Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. *Nat. Protoc.* **2011**, *6*, doi:10.1038/nprot.2010.181.

203. Quan, J.; Tian, J. Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. *Nat. Protoc.* **2011**, *6*, 242–251, doi:10.1038/nprot.2010.181.

204. Orrego, A. H.; Romero-Fernández, M.; Del, M.; Millán-Linares, C.; Pedroche, J.; Guisán, J. M.; Rocha-Martin, J. High Stabilization of Enzymes Immobilized on Rigid Hydrophobic Glyoxyl-Supports: Generation of Hydrophilic Environments on Support Surfaces., doi:10.3390/catal10060676.

205. Ellman, G. L. Tissue sulfhydryl groups. *Arch. Biochem. Biophys.* **1959**, *82*, 70–77, doi:10.1016/0003-9861(59)90090-6.

206. Krieger, F.; Mö, A.; Kiefhaber, T. Effect of Proline and Glycine Residues on Dynamics and Barriers of Loop Formation in Polypeptide Chains. **2005**, doi:10.1021/ja042798i.

207. Bagby, S.; Tong, K. I.; Liu, D.; Alattia, J. R.; Ikura, M. The button test: A small scale method using microdialysis cells for assessing protein solubility at concentrations suitable for NMR. *J. Biomol. NMR* **1997**, *10*, 279–282, doi:10.1023/A:1018359305544.

208. Sattler, M.; Schleucher, J.; Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* **1999**, *34*, 93–158, doi:10.1016/S0079-6565(98)00025-9.

209. Hill, R. B.; Bracken, C.; Degrado, W. F.; Palmer, A. G. Molecular motions and protein folding: Characterization of the backbone dynamics and folding equilibrium of α2D using 13C NMR spin relaxation. *J. Am. Chem. Soc.* **2000**, *122*, 11610–11619, doi:10.1021/ja001129b.

210. Ishima, R.; Torchia, D. A. Protein dynamics from NMR. *Nat. Struct. Biol.* 2000, *7*, 740–743.

211. Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science (80-.).* **1996**, *274*, 1531–1534, doi:10.1126/science.274.5292.1531.

212. Zheng, D.; Huang, Y. J.; Moseley, H. N. B.; Xiao, R.; Aramini, J.; Swapna, G. V. T.; Montelione, G. T. Automated protein fold determination using a minimal NMR constraint strategy. *Protein Sci.* **2003**, *12*, 1232–1246, doi:10.1110/ps.0300203.

213. Cai, M.; Huang, Y.; Sakaguchi, K.; Clore, G. M.; Gronenborn, A. M.; Craigie, R. *An efficient and cost-effective isotope labeling protocol for proteins expressed in Escherichia coli*; KLUWER/ESCOM, 1998; Vol. 11;.

214. Schütz, S.; Sprangers, R. Methyl TROSY spectroscopy: A versatile NMR approach to study challenging biological systems. *Prog. Nucl. Magn. Reson. Spectrosc.* 2020, *116*, 56–84.

215. Muchmore, D. C.; McIntosh, L. P.; Russell, C. B.; Anderson, D. E.; Dahlquist, F. W. Expression and nitrogen-15 labeling of proteins for proton and nitrogen-15 nuclear magnetic resonance. *Methods Enzymol.* **1989**, *177*, 44–73, doi:10.1016/0076-6879(89)77005-1.

216. Kainosho, M.; Torizawa, T.; Iwashita, Y.; Terauchi, T.; Mei Ono, A.; Güntert, P. Optimal isotope labelling for NMR protein structure determinations. *Nature* **2006**, *440*, 52–57, doi:10.1038/nature04525.

217. Gardner, K. H.; Kay, L. E. The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins. *Annu. Rev. Biophys. Biomol. Struct.* 1998,

27, 357–406.

218. Bhatwa, A.; Wang, W.; Hassan, Y. I.; Abraham, N.; Li, X. Z.; Zhou, T. Challenges Associated With the Formation of Recombinant Protein Inclusion Bodies in Escherichia coli and Strategies to Address Them for Industrial Applications. *Front. Bioeng. Biotechnol.* 2021, *9*, 65.

219. Klopp, J.; Winterhalter, A.; Gébleux, R.; Scherer-Becker, D.; Ostermeier, C.; Gossert, A. D. Cost-effective large-scale expression of proteins for NMR studies. *J. Biomol. NMR* **2018**, *71*, 247–262, doi:10.1007/s10858-018-0179-0.

220. Cai, M.; Huang, Y.; Yang, R.; Craigie, R.; Clore, G. M. A simple and robust protocol for high-yield expression of perdeuterated proteins in Escherichia coli grown in shaker flasks. *J. Biomol. NMR* **2016**, *66*, 85–91, doi:10.1007/s10858-016-0052-y.

221. Studier, F. W.; Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **1986**, *189*, 113–130, doi:10.1016/0022-2836(86)90385-2.

222. Sambrook, J.; Fritsch, E. F.; Maniatis, T. Molecular cloning: a laboratory manual. *Mol. cloning a Lab. manual.* **1989**.

223. Foster, M. P.; McElroy, C. A.; Amero, C. D. Solution NMR of large molecules and assemblies. *Biochemistry* **2007**, *46*, 331–340, doi:10.1021/bi0621314.

224. Suzuki, H.; Mori, S. Shodex Columns for Size Exclusion Chromatography. In *Column Handbook for Size Exclusion Chromatography*; Elsevier, 1999; pp. 171–217.

225. Ribeiro, A. J. M.; Tyzack, J. D.; Borkakoti, N.; Holliday, G. L.; Thornton, J. M. A global analysis of function and conservation of catalytic residues in enzymes. *J. Biol. Chem.* 2020, *295*, 314–324.

226. Arnold, F. H.; Moore, J. C. Optimizing industrial enzymes by directed evolution. *Adv. Biochem. Eng. Biotechnol.* **1997**, *58*, 1–14.

227. Reetz, M. T.; Carballeira, J. D. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. Protoc.* **2007**, *2*, 891–903, doi:10.1038/nprot.2007.72.

228. Reetz, M. T.; Reetz, M. T. Laboratory Evolution of Stereoselective Enzymes: A Prolific Source of Catalysts for Asymmetric Reactions., doi:10.1002/anie.201000826.

229. Davies, G. J.; Gloster, T. M.; Henrissat, B. Recent structural insights into the expanding world of carbohydrate-active enzymes. *Curr. Opin. Struct. Biol.* **2005**, *15*,

637–645, doi:10.1016/j.sbi.2005.10.008.

230. Davies, G.; Henrissat, B. Structures and mechanisms of glycosyl hydrolases. *Structure* **1995**, *3*, 853–859.

231. Planas, A.; Faijes, M.; Codera, V. When Enzymes Do It Better: Enzymatic Glycosylation Methods. *Carbohydr. Chem. State Art Challenges Drug Dev.* **2015**, 215–245, doi:10.1142/9781783267200_0009.

232. CAZy - GH Available online: http://www.cazy.org/Glycoside-Hydrolases.html (accessed on Oct 10, 2021).

233. Lloyd F. Mackenzie, †; Qingping Wang, †; R. Antony J. Warren, ‡ and; Stephen G. Withers*, † Glycosynthases: Mutant Glycosidases for Oligosaccharide Synthesis. *J. Am. Chem. Soc.* **1998**, *120*, 5583–5584, doi:10.1021/JA980833D.

234. Malet, C.; Planas, A. From β-glucanase to β-glucansynthase: glycosyl transfer to α-glycosyl fluorides catalyzed by a mutant endoglucanase lacking its catalytic nucleophile. *FEBS Lett.* **1998**, *440*, 208–212, doi:10.1016/S0014-5793(98)01448-3.

235. Honda, Y.; Kitaoka, M. The first glycosynthase derived from an inverting glycoside hydrolase. *J. Biol. Chem.* **2006**, *281*, 1426–1431, doi:10.1074/JBC.M511202200.

236. Faijes, M.; Pérez, X.; Pérez, O.; Planas, A. Glycosynthase activity of Bacillus licheniformis 1,3-1,4-beta-glucanase mutants: specificity, kinetics, and mechanism. *undefined* **2003**, *42*, 13304–13318, doi:10.1021/BI030131N.

237. Faijes, M.; Planas, A. In vitro synthesis of artificial polysaccharides by glycosidases and glycosynthases. *Carbohydr. Res.* **2007**, *342*, 1581–1594, doi:10.1016/J.CARRES.2007.06.015.

238. Vallmitjana, M.; Ferrer-Navarro, M.; Planell, R.; Abel, M.; Ausín, C.; Querol, E.; Planas, A.; Pérez-Pons, J. A. Mechanism of the family 1 β-glucosidase from Streptomyces sp: Catalytic residues and kinetic studies. *Biochemistry* **2001**, *40*, 5975–5982, doi:10.1021/BI002947J.

239. Guasch, A.; Vallmitjana, M.; Pérez, R.; Querol, E.; Pérez-Pons, J. A.; Coll, M. Cloning, overexpression, crystallization and preliminary X-ray analysis of a family 1 β-glucosidase from *Streptomyces*. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **1999**, *55*, 679–682, doi:10.1107/S0907444998013833.

240. M, F.; A, P. In vitro synthesis of artificial polysaccharides by glycosidases and glycosynthases. *Carbohydr. Res.* **2007**, *342*, 1581–1594, doi:10.1016/J.CARRES.2007.06.015.

241. Sager, C. P.; Fiege, B.; Zihlmann, P.; Vannam, R.; Rabbani, S.; Jakob, R. P.; Preston, R. C.; Zalewski, A.; Maier, T.; Peczuh, M. W.; Ernst, B. The price of flexibility – a case study on septanoses as pyranose mimetics. *Chem. Sci.* **2018**, *9*, 646–654, doi:10.1039/C7SC04289B.

242. Pote, A. R.; Vannam, R.; Richard, A.; Gascón, J.; Peczuh, M. W. Formation of and Glycosylation with Per- *O* -Acetyl Septanosyl Halides: Rationalizing Complex Reactivity En Route to *p* -Nitrophenyl Septanosides. *European J. Org. Chem.* **2018**, *2018*, 1709–1719, doi:10.1002/ejoc.201800310.

243. Tauss, A.; Steiner, A. J.; Stütz, A. E.; Tarling, C. A.; Withers, S. G.; Wrodnigg, T. M. L-Idoseptanosides: Substrates of D-glucosidases? *Tetrahedron Asymmetry* **2006**, *17*, 234–239, doi:10.1016/j.tetasy.2005.12.007.

polymers

MDPI

Review

Chitin Deacetylases: Structures, Specificities, and Biotech Applications

Laia Grifoll-Romero, Sergi Pascual, Hugo Aragunde, Xevi Biarnés and Antoni Planas *

Laboratory of Biochemistry, Institut Químic de Sarrià, Universitat Ramon Llull, 08017 Barcelona, Spain; laiagrifollr@iqs.edu (L.G.-R.); sergipascualt@iqs.edu (S.P.); hugoaragunde@gmail.com (H.A.); xavier.biarnes@iqs.edu (X.B.)

* Correspondence: antoni.planas@iqs.edu

Received: 19 February 2018; Accepted: 19 March 2018; Published: 22 March 2018

Abstract: Depolymerization and de-*N*-acetylation of chitin by chitinases and deacetylases generates a series of derivatives including chitosans and chitooligosaccharides (COS), which are involved in molecular recognition events such as modulation of cell signaling and morphogenesis, immune responses, and host-pathogen interactions. Chitosans and COS are also attractive scaffolds for the development of bionanomaterials for drug/gene delivery and tissue engineering applications. Most of the biological activities associated with COS seem to be largely dependent not only on the degree of polymerization but also on the acetylation pattern, which defines the charge density and distribution of GlcNAc and $GlcNH_2$ moieties in chitosans and COS. Chitin de-*N*-acetylases (CDAs) catalyze the hydrolysis of the acetamido group in GlcNAc residues of chitin, chitosan, and COS. The deacetylation patterns are diverse, some CDAs being specific for single positions, others showing multiple attack, processivity or random actions. This review summarizes the current knowledge on substrate specificity of bacterial and fungal CDAs, focusing on the structural and molecular aspects of their modes of action. Understanding the structural determinants of specificity will not only contribute to unravelling structure-function relationships, but also to use and engineer CDAs as biocatalysts for the production of tailor-made chitosans and COS for a growing number of applications.

Keywords: chitin deacetylases; chitosan; chitooligosaccharides; carbohydrate esterases; structure; substrate specificity; deacetylation pattern

1. Introduction

Chitin is a linear polysaccharide of β(1→4)-linked *N*-acetylglucosamine monomers. It was first isolated from fungi in 1811 [1] and its structure was determined in 1929 [2]. Chitin is a major structural component of the exoskeletons of arthropods (insects and crustaceans), of the endoskeletons of mollusks, and it is also found in the cell walls of fungi and diatoms [1,3,4]. It is one of the most abundant organic molecules after cellulose, and the most abundant natural amino polysaccharide. Chitin is present as ordered macrofibrils, mainly in two allomorphs: α-chitin, with antiparallel chains [5], is the most abundant and it is isolated from the exoskeleton of crustaceans, particularly from shrimps and crabs; and β-chitin, with parallel chains [6], is present in the cell walls of diatoms and in the skeletal structures of cephalopods, and commonly extracted from squid pens. β-Chitin is easily converted to α-chitin by alkaline treatment followed by flushing in water [5]. Chitin is also found as γ-chitin in fungi and yeast, which is a combination of the α and β allomorphs [7,8].

Depolymerization (hydrolysis) of chitin by chitinases results in chitooligosaccharides (COS), and de-*N*-acetylation of chitin and COS yields chitosan and partially acetylated COS (paCOS) or fully deacetylated glucosamine oligomers (Figure 1). In nature, the deacetylation of chitin is almost never

complete, and chitosan refers to a family of heteropolysaccharides composed of *N*-acetylglucosamine and glucosamine, characterized by their degree of polymerization (DP), degree of acetylation (DA), and pattern of acetylation (PA). Only some fungi of the *Zygomycota*, *Basidiomycota* and *Ascomycota* phyla have been reported to be capable of naturally producing chitosans [9]. The free amino groups of the deacetylated units in the polymer are protonated at slightly acidic pH, thus making chitosans the only known natural polycationic polysaccharides [1,3,9]. They interact with polyanionic biomolecules such as polyanionic phospholipidic membranes, glycosaminoglycans at cell surfaces, proteins, and nucleic acids. Depolymerization by hydrolysis of the β-1,4-linkages of chitin and chitosan yields their respective oligosaccharides (COS, paCOS) [10]. Whereas chitin and chitosans act as structural polymers, their oligomers are involved in molecular recognition events, such as cell signaling and morphogenesis, and act as immune response elicitors and host-pathogen mediators [11–15]. The catabolism of chitin and chitosan is summarized in Figure 1. Chitin and chitosan oligosaccharides are further degraded following different organism-dependent pathways to end up in the central energy metabolism.

The de-*N*-acetylation of chitin and chitooligosaccharides is catalyzed by chitin deacetylases (CDAs), which exhibit different substrate specificities leading to fully or partially deacetylated products with diverse degrees of acetylation (DA) and patterns of acetylation (PA). In addition to the role CDAs play in the biology of their natural organisms, there is a growing interest in the biochemical characterization of CDAs in order to use them as biocatalysts for the production of partially deacetylated chitooligosaccharides (paCOS) as bioactive molecules in different application fields, or to inhibit them since they are potential targets against pathogenic microorganisms. The aim of this review is to analyze the current knowledge on the biochemistry of chitin deacetylases with regard to their substrate specificity. Although a large number of enzymes have been experimentally identified as chitin deacetylases able to deacetylate chitin either in vivo or in vitro, few of them have been analyzed at the protein level. Here we will focus on those CDAs with characterized activity on chitooligosaccharides (COS) to address the issue of substrate specificity and the deacetylation pattern.

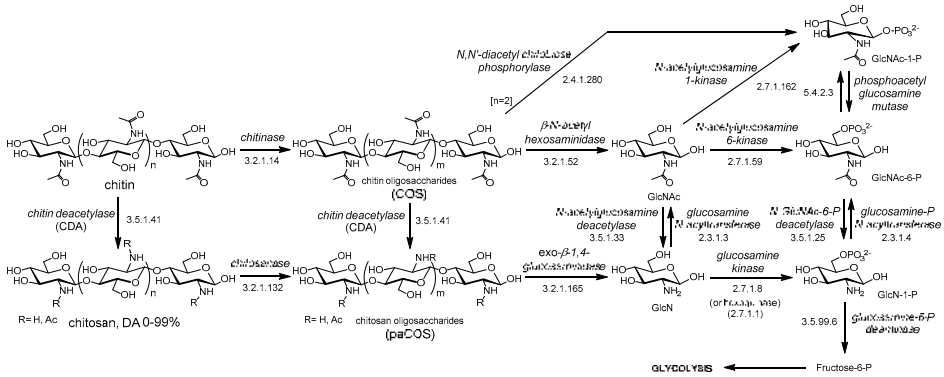


Figure 1. Chitin catabolism. GlcNAc: *N*-acetylglucosamine; GlcN: glucosamine; DA: degree of acetylation. COS: chitooligosaccharides (or chitin oligosaccharides), paCOS: partially acetylated chitooligosaccharides (or chitosan oligosaccharides).

2. Chitin Deacetylases and the Carbohydrate Esterase Family 4 (CE4)

Chitin deacetylases (CDAs, EC 3.5.1.41) and chitooligosaccharides deacetylases (CODs, EC 3.5.1.105) are classified in the carbohydrate esterase family 4 (CE4) in the CAZY database (Carbohydrate Active Enzymes, www.cazy.org) [16]. The CE4 family also contains peptidoglycan *N*-acetylglucosamine deacetylases (EC 3.5.1.104), peptidoglycan *N*-acetylmuramic acid deacetylases (EC 3.5.1.-), poly-β-1,6-*N*-acetylglucosamine deacetylase (EC 3.5.1.-), and some acetyl xylan esterases (EC 3.1.1.72) [17]. These enzymes share a conserved region known as the NodB homology domain

due to its similarity to the NodB oligosaccharide deacetylase, one of the first CE4 enzymes to be characterized [18]. Most currently reported and characterized CDAs and CODs are CE4 enzymes, with the exception of diacetylchitobiose deacetylases (Dacs) from archaea and a COD from *Bacillus cereus* (*Bc*ZBP) that belong to the CE14 family [19,20]. Few other enzymes, such as insect CDAs and a COD from *E. coli* (ChbG) [21] are in the group of "non-classified" in the CAZY database since they do not share sequence similarities to the other CDA families.

The deacetylase activity from extracts of the fungus *Mucor rouxii* was the first active CDA identified and partially purified in the mid 1970s [22,23]. Later on, the NodB from a rhizobium species was the first biochemically characterized COD in 1993 [18]. Many other CDAs and CODs were later identified and purified from very diverse organisms, including archaea, marine bacteria, fungi, and insects [24]. These enzymes are diverse in their biochemical properties: molecular masses in the range from 12 to 150 kDa, acidic isoelectric points (pI from 2.7 to 4.8), optimum pH for activity from 4.5 to 12, and significant thermal stability, with optimum temperatures for activity in the range from 30 to 60 °C. Most CDAs are highly inactive on crystalline chitin due to the inaccessibility of the acetyl groups in the tightly packed chitin structure, and have a preference for soluble forms of chitins such as glycol-chitin or chitin oligomers, as well as partially deacetylated chitin (chitosans). It has recently been shown that CDA activity on crystalline chitin is greatly enhanced by oxidative cleavage of the surface polymer chains by lytic polysaccharide monooxygenases (LPMO) [25]. Some CDAs contain carbohydrate binding modules (CBM) fused to the catalytic domain that seem to enhance the deacetylase activity by increasing the accessibility of the substrate to the catalytic domain [26].

CDAs are localized in different cellular compartments, in the periplasm, in the cytosol, or secreted as extracellular enzymes. Periplasmic fungal CDAs are generally tightly coupled to a chitin synthase to rapidly deacetylate newly synthesized chitin before their maturation and crystallization. Extracellular fungal CDAs are secreted to alter the physicochemical properties of the cell wall, which results in protection against exogenous chitinases, or initiates sporulation or autolysis. In bacteria, CDAs are either intracellular, as those involved in Nod factors biosynthesis in *Rhizobium* species, or extracellular, as those involved in the catabolism of chitin in marine bacteria [24,27,28].

Some CE4 enzymes classified in a specific subfamily also show activity on typical substrates from other subfamilies. Peptidoglycan GlcNAc deacetylases, involved in the de-*N*-acetylation of the bacterial cell wall peptidoglycan with critical functions in the maturation and turnover of peptidoglycan and in bacterial pathogenicity, are also active on COS. Some CDAs have activity on acetylxylan, as well as some acetylxylan esterases are active on COS, which makes their classification doubtful in some cases.

3. Function and Specificity of CE4 Chitin Deacetylases

3.1. Deacetylation Patterns

Chitin deacetylases exhibit diverse deacetylation patterns, reflecting different substrate specificities and pattern recognition on their linear substrates. The mechanisms of action of enzymes that modify in-chain units on a linear polysaccharide are commonly classified as multiple-attack, multiple-chain, and single-chain mechanisms [29]. In the multiple-attack mechanism, binding of the enzyme to the polysaccharide chain is followed by a number of sequential deacetylations, after which the enzyme binds to another region of the polymeric chain. (i.e., *M. rouxii* [30,31]). On a polymeric substrate, this mechanism will result in a block-copolymer structure with blocks of $GlcNH_2$ units within the GlcNAc chain. On COS, it will usually result in full deacetylation of the oligomer. In the multiple-chain mechanism, the enzyme forms an active enzyme-polymer complex and catalyzes the hydrolysis of only one acetyl group before it dissociates and forms a new active complex (i.e., *C. lindemuthianum* CDA [32,33]). It will result in a random distribution of the $GlcNH_2$ and GlcNAc units along the polymeric chain or, in the case of COS substrates, it will render a number of partially deacetylated oligosaccharide intermediates ending in a specific deacetylation pattern or in full deacetylation, depending on the enzyme and the substrate. Finally, a single-chain mechanism includes

processive enzymes in which a number of catalytic events occur on a single substrate molecule leading to sequential deacetylation. Some bacterial chitooligosaccharides deacetylases (CODs), which are specific for a single position leading to mono-deacetylated products (i.e., *Rhizobium* NodB or *Vibrio* CDA or COD) are also included in the last group.

A fundamental question is how these enzymes define their action pattern. This is relevant not only to understand their biological functions but also to use CDAs (native and engineered variants) as biocatalysts for the production of chitosans with non-random deacetylation patterns, and partially deacetylated COS with tailored patterns of acetylation (see Section 6). A structural model on the determinants of substrate specificity is currently emerging from studies on substrate specificity, determination of 3D structures of enzyme-substrate complexes, and multiple sequence alignments. Many CDAs, particularly from fungal origin, have been identified as involved in chitin deacetylation in vivo, but only few of them have been characterized with regard to substrate specificity and mode of action. To the best of our knowledge, Table 1 compiles the CDA enzymes in family CE4 that have been biochemically characterized and have reported activity on COS substrates, some of them with solved three-dimensional structure by X-ray crystallography. Relevant information on their biological function and substrate specificity is summarized below.

CDAs participate in diverse biological processes, which include cell wall morphogenesis and host-pathogen interaction in fungi, generation of signaling molecules in bacteria, and participation in the catabolism of chitin as carbon, nitrogen, and energy sources in marine bacteria and fungi. CDAs were thought to be restricted to fungi and bacteria until a first report in 1986 on their presence in arthropods [34]. CDAs seem to be widely present in insects, in cuticles and the perithrophic midgut matrix, but little is known on the function and properties of insect CDAs [35,36], and they are not included here because scarce information on substrate specificities has been reported.

3.2. Fungal CDAs

Fungal CDAs are involved in fungal nutrition, morphogenesis and development [27,29], participating in cell wall formation and integrity [37], in spore formation [38], germling adhesion [39], fungal autolysis [40], and in defense mechanisms for host infection [41].

Fungi that have chitosan (in addition to chitin) as a structural component of the cell wall, secrete CDAs to the periplasmic space that contribute to chitosan biosynthesis from nascent chitin synthesized by chitin synthases. It occurs during exclusive periods corresponding to their particular biological role in the cell cycle of the fungal species: during cell wall formation (i.e., *M. rouxi* [42] and *A. coerulea* [43]), during sporulation (i.e., *S. cerevisiae* [38,44]), or during vegetative growth (i.e., *C. neoformas* [37]).

Pathogenic fungi secrete CDAs during fungal hyphae penetration to evade plant defense mechanisms and gain access to host tissues. Plants secrete chitinases to break the fungal cell-wall chitin down to chitooligosaccharides (COS), and the released COS are recognized by plant chitin-specific receptors, triggering resistance responses [41]. COS elicitation of resistance mechanisms involve activation of host defense genes [45,46]. There is cumulative evidence that fungi evade plant defense mechanisms by partially deacetylating either their exposed cell wall chitin or the chitooligosaccharides produced by the action of plant chitinases. In both cases, the resulting partially deacetylated oligomers are not well recognized by the specific plant receptors reducing or preventing the elicitation of the defense responses [41,47–49].

Filamentous fungi undergo autolysis by self-digestion of aged hyphal cultures due to carbon starvation [40,50]. During this event there is an increased presence of hydrolytic enzymes, especially those involved in cell wall degradation, and CDAs are secreted to the extracellular medium to deacetylate the chitin oligomers produced by chitinases (i.e., *A. nidulans* [51,52]).

Table 1. CDAs and CODs with characterized activity on COS.

Enzyme [1]	Organism	ID [2]	PDB [3] [Ref.]	COS Substrates [4]	Ref. [5] COS	Metal [6]	PA [7] (on A_n)
*Mr*CDA	*Mucor rouxii*	P50325		≥DP3	[31]	Zn^{2+}	D_n
*Cl*CDA	*Colletotrichum lindemuthianum*	Q6DWK3	2IW0 [53]	DP6 > DP5 > DP4 > DP3 > DP2	[32]	Co^{2+}, Zn^{2+}	D_n
*An*CDA	*Aspergillus nidulans*	Q5AQQ0	2Y8U [25]	DP2 > DP3 > DP4 > DP5	[25]	Co^{2+}	D_n
*Pa*CDA	*Podospora anserine*	XP_001912680.1		≥DP2	[26]	Zn^{2+}	D_n
*Pgt*CDA	*Puccinia graminis*	XP_003323413.1		DP6 > DP5 > DP4	[54]	n.r. [7]	AAD_{n-2}
*Pes*CDA	*Pestolotiopsis* sp.	APH81274.1		DP6-DP5-DP4	[49]	n.r.	$AAD_{n-3}A$
*Pc*CDA	*Pochonia chlamydosporia*			DP5 > DP4	[55]	n.r.	$ADDA_{n-3}$
*Sc*CDA1	*Saccharomyces cerevisiae*	Q06702		DP4, DP6	[56]	n.r.	n.r.
*Sc*CDA2	*Saccharomyces cerevisiae*	Q06703		DP6 > DP5 > DP4 > DP3 > DP2	[57]	Co^{2+}	n.r.
*Mo*CDA *	*Mortierella* sp.			DP7 > DP6 > DP5 > DP4 > DP3 > DP2	[58]	(Co^{2+})	n.r.
*Acoe*CDA *	*Absidia coerulea*			DP5 > DP4 > DP3	[43]	n.r.	n.r.
*Acor*CDA *	*Absidia corymbifera*			DP7 > DP6 > DP5 > DP4 > DP3 > DP2	[59]	(Co^{2+}, Ca^{2+}, Mg^{2+})	n.r.
*Fv*CDA	*Flammulina velutipes*	BAE92728.1		DP5 > DP4 > DP3 > DP2	[60]	(Co^{2+}, Ca^{2+}, Zn^{2+})	n.r.
*Po*CDA *	*Penicillium oxilicum*			DP5 >> DP3 > DP2	[61]	(Co^{2+} Cu^{2+})	n.r.
*Af*CDA *	*Aspergillus flavus*			DP4	[62]	(Zn^{2+}, Mn^{2+})	n.r.
*Sb*CDA *	*Scopulariopsis brevicaulis*			DP6 > DP5 > DP4 > DP3 > DP2	[63]	n.r.	n.r.
*Rc*CDA	*Rhizopus circinans*	A7UMZ0		DP6	[64]	(Mn^{2+}, Mg^{2+})	n.r.
*Rs*CDA	*Rhizopus stolonifer (nigricans)*	Q32XH4		n.r.	[64]		
*Gb*CDA	*Gongronella butleri*	Q8J2N6		n.r.	[65]		
*Pb*CDA	*Phycomyces blakesleeanus*	Q9P4U2		n.r.	[66]		
*Sch*CDA	*Schizophyllum commune*	Q9P453		n.r.	[67]		
*Cn*CDA1, 2, 3	*Cryptococcus neoformans*	Q5KFG8, Q5KIC2, P0CP76		n.r.	[37]		
*Eh*CDA	*Entamoeba histolytica*	XP_656356.1		DP5, DP6	[68]	n.r.	n.r.
NodB	*Sinorhizobium meliloti*	P02963		DP5 > DP2 (DP4, DP3)	[18]	Mn^{2+} Mg^{2+}	DA_{n-1}
*Vc*COD (*Vc*CDA)	*Vibrio cholera*	Q9KSH6	4NY2 [69]	DP2 > DP3 > DP4 > DP5 > DP6	[70]	Zn^{2+}	ADA_{n-2}
*Vp*COD	*Vibrio parahaemolyticus*	Q9KSH6	3WX7 [71]	DP2 > DP3	[72]	Zn^{2+}	n.r.
*Va*COD	*Vibrio alginolyticus*	Q9KSH6		DP2	[73]	Zn^{2+}	AD
*Sw*COD	*Shewanella woodyi*	ACA84860.1		DP2 > DP3 > DP4	[74]	n.r.	AD; [ADA_{n-2}]
*Sb*COD	*Shewanella baltica*	ABN60929.1		DP2 > DP4 > DP3	[75]	n.r.	AD; [ADA_{n-2}]
*Ar*CE4A	*Arthrobacter* sp.	A0A2C8C1T7	5LFZ [76]	DP5 > DP6 ≈ DP4 > DP3 >> DP2	[76]	Ni^{2+} [8]	$D_{n-1}A$

[1] Characterized recombinant enzymes, except those with an asterisk (*) that have been characterized from the native organism and are not included in the sequence alignment, Figure 4. [2] Uniprot or GenBank accession code, [3] PDB accession code and publication of the 3D structure. [4] Activity on chitooligosaccharides (COS) as a function of the degree of polymerization (DP). [5] Selected publication on substrate specificity. [5] Native metal cation or, in parenthesis, metals that enhanced the enzyme activity when added in the reaction buffer. [6] Pattern of acetylation (PA): structure of the main final deacetylated product (A: GlcNAc; D: $GlcNH_2$). Other patterns of acetylation with specific substrates are given in the text. [7] n.r.: not reported. [8] Native metal unknown, Ni^{2+} probably from purification/crystallization.

In general, fungal CDAs deacetylate soluble forms of chitin such as glycol-chitin and chitosans of variable DA, but they are inactive, or show low activity, towards insoluble chitins such as crystalline α- and β-chitin and colloidal chitin. Pretreatment of chitin to make surface fibrils more accessible may result in increased deacetylase activity. It has been recently shown that the activity of an *A. nidulans* CDA on crystalline chitin was enhanced by a lytic polysaccharide monooxygenase (LPMO) that increases substrate accessibility by oxidative cleavage of the chitin chains [25]. Some CDAs also appear to be active on acetylxylan (i.e., *An*CDA), but any of them act on peptidoglycans, typical substrates of other CE4 family members. In addition to polymeric substrate, not many CDA have been analyzed with COS as substrates (Table 1). The analysis of the products from enzymatic deacetylation with regards to the extent and pattern of deacetylated provides information about the specificity and mode of action of these enzymes. The first seven entries in Table 1 correspond to CDAs for which the deacetylation pattern on COS has been reported, whereas the rest of the entries are CDAs active on COS but, to the best of our knowledge, the structure of the deacetylated products has not been analyzed.

Mucor rouxii (*Amylomyces rouxii*) (*Mr*CDA). The dimorphic fungus *M. rouxii* has a cell wall mainly composed of chitin, chitosan, and mucoric acid. While chitin accounts for 10% of the total dry weight of the cell wall, chitosan reaches 30% [77]. The *M. rouxii* CDA enzyme was initially found in the cytosol [23], but it is also secreted into the periplasm where it participates in a tandem synthetic mechanism that involves a chitin synthase and a chitin deacetylase working consecutively and synchronously for synthesis and deposition of chitosan polymers at the outer membrane [78]. Decoupling this mechanism prevents the formation of chitosan [42,79]. *M. rouxii*, like other fungi, has been identified as a suitable microorganism for chitosan production by means of biofermentation processes [80,81]. *Mr*CDA is a monomeric high-mannose-type glycosylated protein with an apparent molecular mass of 75–80 kDa [82]. Kinetic studies indicate that the preferred catalytic metal is Zn^{2+} like many other CDAs. In terms of activity, its optimal pH and temperature values are between 4.5–5.5 and 50 °C, respectively [23,33]. *Mr*CDA deacetylates chitinous polymers such as glycol-chitin, colloidal chitin, chitosan, and chitin, but also deacetylates acetylxylan [17,23]. On chitooligosaccharides, triacetylchitotriose is the smallest substrate and the activity increases with the degree of polymerization (DP) [23,31,42,78]. The enzyme follows a multiple-attack mechanism [30] but the resulting pattern of acetylation (PA) depends on the DP of the substrate: whereas DP3, DP6 and DP7 substrates are not fully deacetylated, leaving the reducing GlcNAc unmodified [$D_{n\_1}A$], DP4 and DP5 substrates are fully deacetylated [D_n]. In all cases, deacetylation starts at the non-reducing end residue and then proceeds to the neighboring monomer towards the reducing end [31].

Colletotrichum lindemuthianum (*Cl*CDA). The deuteromycete *C. lindemuthianum* is a plant pathogen that causes anthracnose, a disease which affects economically important crop species [83]. *Cl*CDA is a heavily glycosylated secreted enzyme allegedly playing a role in the host-pathogen interaction, deacetylating the chitin oligomers resulting from the activity of plant chitinases on the fungal cell wall [83,84]. Less likely is its function in deacetylating the fungal cell wall chitin to evade degradation by plant chitinases, since no chitosan has been observed in the cell wall ultra-structure [85]. Since its discovery in the 1980s it has been purified from its natural host [84,86] as well as expressed in several eukaryotic and bacterial hosts such as *Pichia pastoris* [87,88] and *E. coli* [89,90]. The enzyme has a preference for Co^{2+} and Zn^{2+} as the catalytic metal cation and its activity is substantially inhibited by Cu^{2+} or Ni^{2+}, but not inhibited by EDTA or acetate [53,86]. It is a quite thermostable enzyme with an optimum temperature of 60 °C, and a pH optimum of 8.0. *Cl*CDA is active on both chitin polymers (glycol-chitin) and COS. It fully deacetylates COS with a DP equal to or greater than 3, while it only deacetylates the non-reducing GlcNAc of diacetylchitobiose [32,91]. *Cl*CDA acts by a multiple-chain mechanism following a pathway in which the first residue to be deacetylated is the second from the reducing end [32,33]. The initial mono-deacetylation reaction shows no dependency of k_{cat} on DP and a decrease of K_M with increasing DP [33,53]. However, kinetics of fully deacetylated products formation show an increase in k_{cat} and reduction in K_M that correlate with the increase of DP [86]. It has been reported that this enzyme is reversible, as it is also able to catalyze the acetylation of chitosan oligomers [92–94].

Aspergillus nidulans (*An*CDA). During cell autolysis, *An*CDA is secreted into the extracellular medium to deacetylate the chitin oligomers produced by chitinases [40,50,52,95]. The enzyme has been purified from A. nidulans cultures as a glycosylated enzyme [51]. The recombinant protein has been expressed in E. coli and purified by refolding from inclusion bodies [96] and recently it has been obtained in soluble form [25]. Like ClCDA, *An*CDA is a thermostable protein with an optimal temperature of 50 °C and retaining 68% activity after 1 h at 80 °C. Its optimum pH is 7–8 [51,96]. The enzyme is active on soluble chitins (CM-chitin, glycol-chitin), colloidal chitin, chitosan, acetylxylan, and acetylated glucuronoxylan, but not on peptidoglycan [25,51]. *An*CDA is active on COS with a DP from 2 to 6 [25]. The enzyme catalyzes mono-deacetylation of $(GlcNAc)_2$ and it is inactive on GlcNAc monosaccharide. Longer substrates than DP2 are fully deacetylated. However, the deacetylation rate exhibits a counter-intuitive relationship with the DP of the substrate: odd-numbered COS (DP5, DP3) have higher apparent rate constants than even-numbered oligomers (DP4, DP2). For the DP6 substrate, time-course monitoring of products formation reveals that the first deacetylation event occurs at random positions except for the reducing end, which reacts much slower to yield the fully deacetylated end product $[D_n]$.

Podospora anserina (*Pa*CDA). The filamentous ascomycete *Podospora anserina* lives as a saprophyte on herbivore dung [97]. It has a limited lifespan and it is a model organism in cell aging studies [98]. *Pa*CDA was identified in a search for CDAs containing chitin binding domains. The enzyme has been recombinantly expressed in *Hansenula polymorpha* as a full length protein composed of the CE4 domain flanked by two CBM18 domains [26]. The low activity of the enzyme on colloidal chitin is significantly reduced by deletion of the CBM domains, which supports the hypothesis that the presence of the CBMs helps the enzyme to act on insoluble substrates. *Pa*CDA is active on soluble glycol-chitin, chitosans with a high DA, and COS, with optimum pH and temperature values of 8.0 and 55 °C, respectively. It fully deacetylates COS with a DP $\geq$ 2 and follows a multiple-chain mechanism. With the DP3 substrate, the first deacetylation event has a clear preference for the reducing end, but all possible isomers are found for both mono- and di-deacetylated intermediate products. With DP4 and DP5 substrates, the residue next to the reducing end is preferentially deacetylated first, with the second deacetylation occurring mainly next to the existing $GlcNH_2$ unit on either side. Deacetylation is faster for longer substrates, with deacetylation of the reducing end occurring as a late event [26].

Puccinia graminis f. sp. *Tritici* (*Pgt*CDA). The biotrophic basidiomycete *Puccinia graminis f.* sp. *Tritici* is the causative agent of the stem rust [99]. The appearance of resistant races of *P. graminis* affecting wheat cultivars has been recognized as a serious threat to food security [100,101], boosting the interest in understanding the virulence and defense mechanism of this fungal pathogen. Rust fungi promote the formation of complex structures in order to invade the plant cells but at the same time they must prevent the triggering of immune responses [102]. A main transition during infection is from the ectophytically growing appressorium to the endophytically growing substomatal vesicle; while the former exposes chitin on its surface, the latter exposes chitosan [47,103]. *Pgt*CDA may not only participate in the chitin to chitosan transition, making the cell wall less susceptible to host chitinases [104], but also could deacetylate the chitooligosaccharide products, reducing its elicitor properties [105]. *Pgt*CDA has been recombinantly expressed in *E. coli* as a fusion protein with the maltose binding protein (MBP) [54]. Its optimal pH for activity is between 8 and 9 and its optimal temperature is 50 °C. It is not active on insoluble polymers such as α- or β-chitin, but efficiently deacetylates colloidal chitin, glycol-chitin and chitosans, on which activity increases with the degree of acetylation. With COS substrates, the minimal substrate is tetraacetylchitotetraose (DP4). The structure of the products from enzymatic deacetylation of DP4 to DP6 substrates reveals that the enzyme acts by a multiple-chain mechanism and specifically deacetylates all but the last two GlcNAc units on the non-reducing end $[AA(D)_{n\_2}]$ [54].

Pestalotiopsis sp. (*Pes*CDA). The endophytic fungus *Pestalotiopsis* sp. lives inside the tissues of its plant hosts in tropical areas [106]. To successfully survive in their hosts, endophytes also need to avoid being detected by the plant immune system. A secreted *Pestalotiopsis* CDA has been identified when chitosan was present in the culture medium [49]. The recombinantly expressed *Pes*CDA is active on colloidal chitin as substrate, chitosans with a DA of 10–60% (higher activity with a higher DA), and COS,

but inactive on crystalline α- or β-chitin. When analyzing the activity on COS, tetraacetylchitotetraose is the minimal substrate. With a DP5 substrate, the optimum pH and temperature values are 8.0 and 55 °C, respectively. Through a multiple-chain mechanism, the enzyme deacetylates all residues of the substrates except the reducing end and the last two GlcNAc residues from the non-reducing end, with a pattern of deacetylation [AA(D)$_{n_3}$A] [49]. The chitosan oligomers obtained from deacetylation of a DP6 substrate by *Pes*CDA have shown that, as opposed to the fully acetylated oligomer, they are no longer elicitors of the plant immune system in rice cells [49].

Pochonia chlamydosporia (*Pc*CDA). The ascomycete *Pochonia chlamydosporia* infects females and eggs of cyst or root-knot nematodes. It is used as a biocontrol agent against a number of plant parasitic nematodes in food-security crops [107–109]. *P. chlamydosporia* expresses chitosanases and chitin deacetylases during egg infection. Since chitosan is associated with the sites of fungal penetration, it has been suggested that secreted CDAs are involved in nematode infection [110]. A *Pc*CDA has been recently characterized [55]. The full-length protein contains the CE4 catalytic domain flanked by two CBM18 chitin binding domains. The recombinantly expressed *Pc*CDA catalytic domain deacetylates COS with a DP $\geq$ 4, with preference for longer substrates. It starts deacetylating the penultimate residue from the non-reducing end and continues deacetylating the next residue towards the reducing end, with a pattern of acetylation [ADDA$_{n_3}$] [55].

The above described CDAs are well characterized in terms of their deacetylation mode of action on COS and the structure of their deacetylated products. A number of other fungal CDAs (Table 1, and below) have also been assayed on COS substrates but, to the best of our knowledge, the deacetylation pattern of the products has not been reported.

Saccharomyces cerevisiae (*Sc*CDA1 and 2). The *S. cerevisiae* ascospore walls are well ordered structures with two outer layers that confer spore resistance, one made of 95% chitosan and the outermost proteinaceous layer rich in dityrosine [44]. Two CDAs are expressed exclusively during sporulation and are required for spore wall rigidity [38]. Both CDAs have been cloned and expressed in yeast as glycosylated proteins active on glycol-chitin [38,56], and in *E. coli* [57] as soluble proteins with deacetylase activity on glycol-chitin, chitosan (DA 50%) and COS. More detailed characterization of *Sc*CDA2 expressed in *E. coli* revealed that at least two GlcNAc residues are required for activity on COS, with maximum activity on DP6 [57]. When glycol chitin is used as substrate the optimum temperature for enzyme activity is 50 °C and the pH optimum is 8.0. It has also been shown that the *Sc*CDAs may act on nascent chitin chains in an in vitro assay system with chitin synthase [56].

Mortierella sp. (*Mo*CDA). Some *Mortierella* species live as saprotrophs in soil and other organic materials such as decaying plant leafs, fecal pellets or on the exoskeleton of arthropods, whereas other species are endophytes [111]. An extracellular CDA was identified [58] and purified from a *Mortierella* sp. as a highly glycosylated protein with maximum activity at pH 5.5–6 and 60 °C [112]. *Mo*CDA is active on soluble substrates as chitosans and glycol-chitin but with no detectable activity on β-chitin, colloidal chitin, and CM-chitin. It is active on COS with a DP $\geq$ 2, with higher activity with increasing DP of the substrate. With diacetylchitobiose, only monodeacetylation was observed. The structure of the deacetylated products from larger oligomers has not been reported.

Absidia sp. (*Acoe*CDA, *Acory*CDA). *Absidia* strains of *Zygomycetes* produce chitosan in their cell wall through the tandem action of chitin synthases and deacetylases. In *A. coerulea*, chitosan accounts for 10% of the vegetative cells and the DA reaches 95%. *AcoeCDA* was purified and proven to be active on glycol-chitin with a pH optimum of 5 at 50 °C. When the purified enzyme was incubated with a chitin synthase, it converted 90% of the nascent chitin from UDP-GlcNAc into chitosan. It deacetylates COS with more than two GlcNAc units, with increasing activity with longer substrates [43]. Similarly, *Absidia corymbifera* secretes a CDA active on glycol-chitin and chitosans with optimum pH and temperature of 6.5 and 55 °C, respectively, and active on COS with DP $\geq$ 2 [59].

Flammulina velutipes (*Fv*CDA). The basidiomycete *Flammulina velutipes* (called Enokitake in Japan) is commercially cultivated and fruited to produce foods with high nutritional value. A CDA that is expressed at the early stages of fruity body development was recombinantly expressed in *Pichia pastoris* [60]. *Fv*CDA,

active on glycol-chitin and colloidal chitin, deacetylates COS from dimer to pentamer, with activity increasing with the DP of the substrate. The enzyme exhibits the maximum activity at 60 °C and pH 7.

Penicilium oxilicum (*Po*CDA). An extracellular CDA from *Penicilium oxilicum*, purified from culture supernatants, exhibits deacetylase activity on glycol-chitin at pH 9, a common value for extracellularly secreted CDAs as opposed to intracellular CDAs with typical pH optima in the range of 5 to 7. *Po*CDA is active on COS with activity increasing from DP2 to DP5 [61].

Aspergillus flavus (*Af*CDA). In the search for extracellularly secreted CDAs for industrial applications, optimization of solid substrate fermentation and submerged fermentation of *Aspergillus flavus* has been reported [62,113]. The *Af*CDA enzyme purified from the extracellular medium has optimal activity on glycol-chitin and colloidal chitin at pH 8 and 50 °C. When assayed with COS as substrates, *Af*CDA is active on DP4 but has no activity on shorter substrates [62].

Scopulariopsis brevicaulis (*Sb*CDA). *Scopulariopsis* spp. are common soil saprophytes. Few species have been associated with human diseases, including *S. brevicaulis*. They are dermatomycotic molds and mainly have been associated with onychomycosis [114,115]. *Sb*CDA is an extracellular enzyme that is active on chitin and chitosans. The purified native enzyme is also active on COS with at least two GlcNAc units, and the activity increases with the DP of the substrate. With DP6, optimum conditions for deacetylation are pH 7.5 and 55 °C [63].

Rhizopus sp. (*Rc*CDA, *Rs*CDA). *Rhizopus* species have been screened as CDA producers. A *R. circicans* CDA has been cloned and recombinantly expressed in *Pichia pastoris* [64]. *Rc*CDA has maximum activity on glycol-chitin at pH 5–6 and 37 °C. On COS, only activity on a DP6 substrate has been reported. A CDA from *Rhizopus stolonifer* (or *nigricans*) has also been isolated as an active enzyme on glycol-chitin but no activity on COS has been reported [64,116]. Fermentation conditions of other *Rhizopus* species as CDA producers are being studied for the bioconversion of chitin to chitosan [117].

Other chitosan producers have been identified and studied as a source of chitosans, with many reports on screening and fermentation optimization, but the corresponding chitin deacetylases have not been characterized yet. Some examples include *Gongronella butleri* [65,118], *Phycomyces blakesleeanus* [66,119], and *Schizophyllum commune* [67].

Cryptococcus neoformans (*Cn*CDA). *Cryptococcus neoformans* is a dimorphic basidiomycetous human fungal pathogen that causes cryptococcal meningoencephalitis, particularly in immunocompromised patients [37]. *C. neoformans* has substantial chitosan in its cell wall during vegetative growth that is necessary for virulence and persistence in the mammalian host [120,121]. Three CDAs are predicted to be GPI-anchored to the cell wall, suggesting that they transverse the plasma membrane or attach to the cell wall to deacetylate the chitin generated by a chitin synthase as it is extruded through the plasma membrane [37]. The GPI-anchor of *Cn*CDA2 has proven to be required for membrane association but dispensable for cell wall association [122]. Activity of *C. neoformans* CDAs on COS substrates has not been reported. Interestingly, screening studies to identify cryptococcal antigens that stimulate an immune response on murine T cell hybridomas reactive with cryptococcal proteins, have shown that two of the CDAs are immunogenic [123,124].

3.3. Protozoan CDAs

Entamoeba histolytica (*Eh*CDA). *Entamoeba histolytica* is an anaerobic parasitic amoebozoan that predominantly infects humans and other primates causing amoebiasis [125]. The genome contains two putative CDAs, one of which has been cloned and recombinantly expressed in *Saccharomyces cerevisiae* [68]. *Eh*CDA deacetylates COS, being active on DP5 and DP6, but with no detected activity on DP4 [68].

3.4. Bacterial CDAs

The predominant CE4 deacetylases in bacteria are chitin oligosaccharide deacetylases (CODs), active on low molecular mass COS and essentially inactive on polymeric chitin and chitosans. These include rhizobial NodB deacetylases and CODs from marine bacteria. But bacterial CDAs

other than CODs are being discovered from screening programs and data mining of sequenced genomes and metagenomes, as in the recent case of an *Arthobacter* CDA.

Sinorhizobium meliloti (NodB). Rhizobial NodB is part of the Nod operon involved in the biosynthesis of Nod factors, the morphogenic signal molecules produced by rhizobia, which initiate the development of root nodules in leguminous plants [126]. NodB is active on chitooligosaccharides from DP2 to DP5 with no differences in k_{cat}, but K_M decreases with increasing DP [18,127–129]. Specifically, k_{cat}/K_M is 5-fold higher for DP5 than for DP2 substrates. DP4 or DP5 substrates are the natural substrates depending on the Rhizobial strain. *Sm*NodB optimum activity between pH 7 and 8 at 30 °C [18]. NodB is highly specific deacetylating only the non-reducing end residue [DA_{n-1}] although traces of a second deacetylation event have been observed upon long incubations [18,130,131].

Vibrio species (*Vc*CDA, *Vp*CDA, *Va*CDA). Chitin oligosaccharide deacetylases (COD) from the *Vibrionaceae* family are involved in the chitin degradation cascades occurring in sea water [132–135]. They have been identified in many *Vibrio* species, such as *V. algynolyticus* [73,136], *V. parahaemolyticus* [72,137], *V. cholera* [70], *V. harveyi* [138] and others. The *V. parahaemolyticus* and *Vibrio* sp. SN184 CDAs only deacetylate DP2 and DP3 substrates, whereas the *Vibrio cholera* chitin deacetylase (*Vc*CDA) has a broader specificity, accepting substrate from DP2 to DP6 [69,70]. *VcCDA* has a 10-fold higher activity on DP2 than on DP4 [69], and specifically deacetylates the penultimate residue from the non-reducing end, generating monodeacetylated products with the pattern [ADA_{n-2}] [69,70,130].

Shewanella species (*Sw*COD, *Sb*COD). In addition to the *Vibrio* genus, CODs have been recently identified and characterized from the *Shewanella* genus, marine bacteria found in extreme aquatic habitats (low temperature and high pressure). *Shewanella* sp. CODs share high sequence identity (50–60%) with *Vibrio* CODs, and have essentially the same biochemical properties. The *S. woodyi* enzyme (*Sw*COD) contains two CBM12 chitin binding domains at the C-terminus, deacetylates the reducing end on diacetylchitobiose [AD], and the activity drastically decreases from DP2 to DP4 substrates, with no activity detected on a DP5 substrate [74]. The *S. baltica* enzyme (*Sb*COD) contains a single CBM12 at the C-terminus, it is active of diacetylchitobiose with the same deacetylation pattern [AD] but it is less active on a DP3 than on a DP4 substrate [75].

Arthrobacter sp. (*Ar*CE4). A bioinformatics search for monodomain and extracellular CDAs in annotated genomes and metagenomes identified *Ar*CE4 as a CDA from an *Arthrobacter* species [76], a Gram-positive bacteria known to grow on chitin and secrete chitinases [139–141]. *Ar*CE4 is active on chitosan (DA 64%), acetylxylan, and insoluble chitin. It also deacetylates COS substrates with DP $\geq$ 2. The activity increases with increasing DP, with higher activity against DP5 compared to DP6. As shown with the DP5 substrate, the enzyme follows a multiple-chain mechanism where different mono- and di-deacetylated products are obtained. Whereas the first deacetylation occurs at all three internal positions, di-deacetylation mainly takes place at the GlcNAc unit next to the reducing end and at either of the two other internal units (ADDAA and ADADA). The final products have a pattern of acetylation [$D_{n-1}A$], where the reducing end unit is not deacetylated [76].

4. Structural Determinants of Activity and Specificity

Structural analysis of CE4 enzymes with solved 3D structure have been recently reviewed [142], comparing and highlighting the differences between the different subfamilies based on substrate preferences. Here we focus and summarize the current knowledge on the structure and specificity of CDAs as a subfamily of CE4 enzymes. The closer similarity and activity on the same substrates provides a framework to analyze the structural determinants responsible for the different modes of action that lead to different patterns of deacetylation in their products.

4.1. 3D Structures

Some CDAs are mono-domain proteins and some others have a multi-domain architecture composed of the CE4 catalytic domain (or NodB homology domain), and several other domains, such as

carbohydrate binding modules (CBMs [143]) and domains with unknown function. The function of the CBMs is not clear and might be diverse depending on the biological role of each enzyme in its organism. In extracellular CDAs acting on the cell wall chitin, they may facilitate solubilization and access to the substrate (i.e., *Pa*CDA with two CBMs, where deletion of one or both confirmed their proposed function in supporting the enzymatic conversion of insoluble chitin [26]). In CDAs acting on low molecular weight COS, the CBMs may be involved in enzyme localization. This is the case of COD enzymes from marine bacteria (*Vibrio* and *Shewanella* species), where the small substrate does not span out of the active site, and the CBMs might bind to chitinous material in order to keep the COD activity close to the site where COS are generated by the action of chitinases.

The first CE4 enzymes with 3D structure determined by X-ray crystallography were the peptidoglycan deacetylases *Bs*PdaA [144] and *Sp*PgdA [145], and the first CDA was that from *Colletotrichum lindemuthianum* (*Cl*CDA) [53]. Currently, only five CDAs in the CE4 family have known 3D structure (Figure 2). The CE4 catalytic domain is characterized by a distorted $(\beta/\alpha)_8$ barrel fold. The distorted barrel, which often lacks one of the αβ repeats of regular TIM barrels, creates a groove into which the extended polymer substrate binds [144,146,147]. Seven or eight parallel β-strands form the β-barrel surrounded by α-helices. In addition, a series of loops decorate the β-barrel and make up the majority of the carbohydrate binding pocket as discussed below.

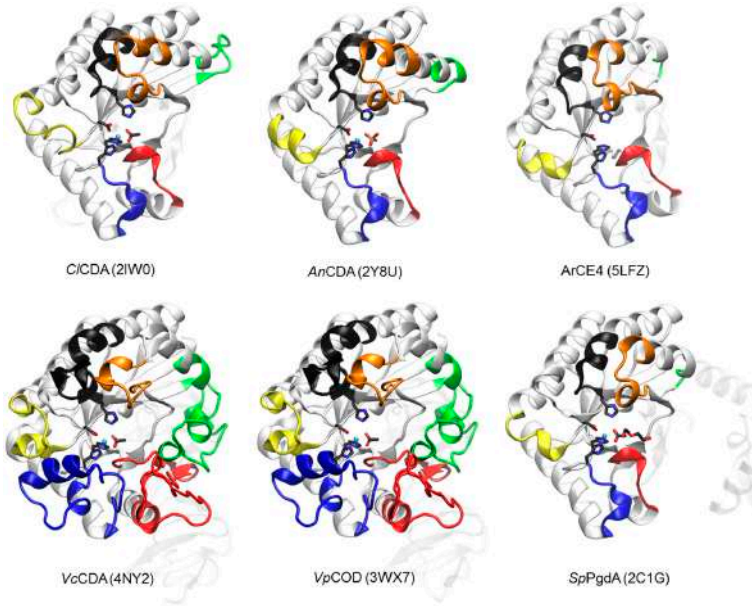


Figure 2. 3D structures of CDAs determined by X-ray crystallography. Loops 1 to 6 colored as in Figure 3. The peptidoglycan deacetylase *Sp*PgdA is also included for comparison (see text). In parenthesis, PDB accession codes.

4.2. The NodB Homology Domain and Conserved Active Site Motifs

The multiple sequence alignment of the CE4 domain for the CDAs listed in Table 1 was guided by the structural superimposition of the available X-ray structures (Figure 2) and is presented in Figure 3. Compared to most of the CDA members, the *Vc*CDA enzyme has substantially longer insertions, and it was key to defining the loops that differentiate CDAs and shape the binding site cleft of these enzymes. Sequences of enzymes without structural date were incorporated into the alignment by means of Hidden Markov Model comparisons. As seen in Figure 3, the conserved motifs and non-conserved insertions are evenly distributed along the sequences of CDAs.

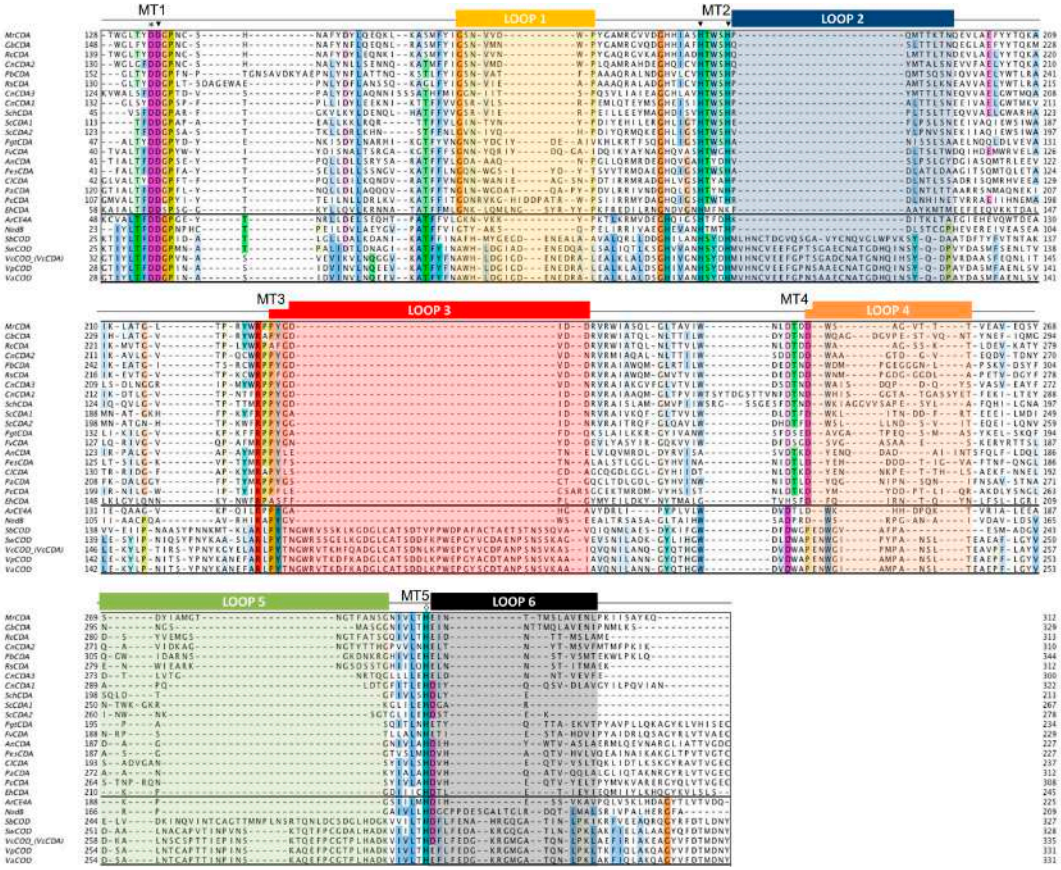


Figure 3. Multiple sequence alignment of the CDA enzymes listed in Table 1. Loops are highlighted with colored boxes according to [69]. Conserved catalytic motifs are labelled MT1-5. The 'His-His-Asp' metal binding triad (▼), catalytic base (*), and catalytic acid (◊) are labelled.

The conserved motifs are related to enzymatic activity (Motifs 1 to 5) and are typically located at the center of the active site structure. The non-conserved insertions correspond to both un-structured and structured loops of variable length, sequence, and geometry that surround the active site. These loops are numbered from Loop 1 to Loop 6 in the sequence alignment (Figure 3). As discussed in Section 4.5, they are key elements in determining the substrate specificity of different CDAs.

As members of the CE4 family, CDAs share the ≈150 aa-long NodB homology domain (CE4 domain). This region is defined by five conserved motifs that, according to the order they appear in the sequence, are named Motif 1 to Motif 5. These consensus motifs were first proposed in 2005 by sequence alignment of representative enzyme members of the CE4 family when the 3D structure of the peptidoglycan deacetylase *Sp*PgdA was solved [144]. Motif 1 (TFDD) is highly conserved in CDAs and contains the general base aspartate (first D) and the metal-binding aspartate (second D). Motif 2 (H(S/T)xxH) is a zinc-binding motif, where the two His residues bind the metal cation and the Ser or Thr residue forms a hydrogen bond with the second His, stabilizing the local conformation of the loop-shaped motif. These two His from Motif 2 plus the metal-binding Asp from Motif 1 are often designated the His-His-Asp metal-binding triad of CE4 enzymes. Motif 3 (RxPY) forms one of the sides of the active site groove and establishes stabilizing interactions with other active site residues. Motif 4 (DxxD(W/Y)) forms the other side of the active site groove, including a hydrophobic residue exposed to the solvent and a buried Asp. Motif 5 (I(V/I)LxHD) contains the catalytic general acid His residues and a Leu, which is part of a hydrophobic pocket that accommodates the acetate methyl group of the substrate.

4.3. Phylogeny of CE4 Chitin Deacetylases

Based on the above multiple sequence alignment, a clustering of the CE4 domain sequences of characterized CDAs based on phylogenetic analysis is presented in Figure 4. This is a reduced phylogenic analysis limited to the CDAs with reported activity on COS as listed in Table 1. Fungal and bacterial CDAs are clearly segregated in two clades, with a protist CDA (*Eh*CDA) located between both groups.

Fungal enzymes from organisms belonging to different phyla (*Zygomycota*, *Basidiomycota*, and *Ascomycota*) are distributed throughout the fungal clade. Within the clade, CDAs appear grouped in two clusters related with their biological function. The first cluster contains orthologous CDAs of different phyla known to have a role in cell wall chitosan biosynthesis at different stages of the fungal cell cycle, such as *Mr*CDA during cell wall formation [42], *Cn*CDAs during vegetative growth [37], or *Sc*CDAs during sporulation [38,44]. Although there is no experimental proof of their biological function, the CDAs from *Gonglonella*, *Phycomyces*, as well as those from *Rhizopus* are likely to be also involved in cell wall formation due to their location in the same cluster of the phylogenetic tree and their taxonomic classification (mucorales inside the *Zygomycota* phylum). Regarding the cellular location of the enzymes in this cluster, most of them are secreted to the periplasm or are GPI-anchored to the cell wall, where they are coupled with chitin synthases for chitosan biosynthesis. The second cluster is mainly composed of extracellular CDAs that participate in host infection, either as a defense mechanism to prevent the elicitation of host defense mechanisms (*Pgt*CDA, *Pes*CDA), or involved in the interaction with the host or as a virulent factor (*Cl*CDA, *Pc*CDA). Extracellular CDAs involved in cell autolysis (*An*CDA) also fall in this group.

Most of the bacterial enzymes included in the alignment are chitin oligosaccharide deacetylases (COD) and are more distantly related to the fungal CDAs. These form a different clade in the phylogenetic tree (Figure 4). The enzymes from marine bacteria (*Vibrio* and *Shewanella* species) are clustered together with high sequence similarity and have similar biological functions and biochemical properties. NodB has a more distant relationship with the other CODs.

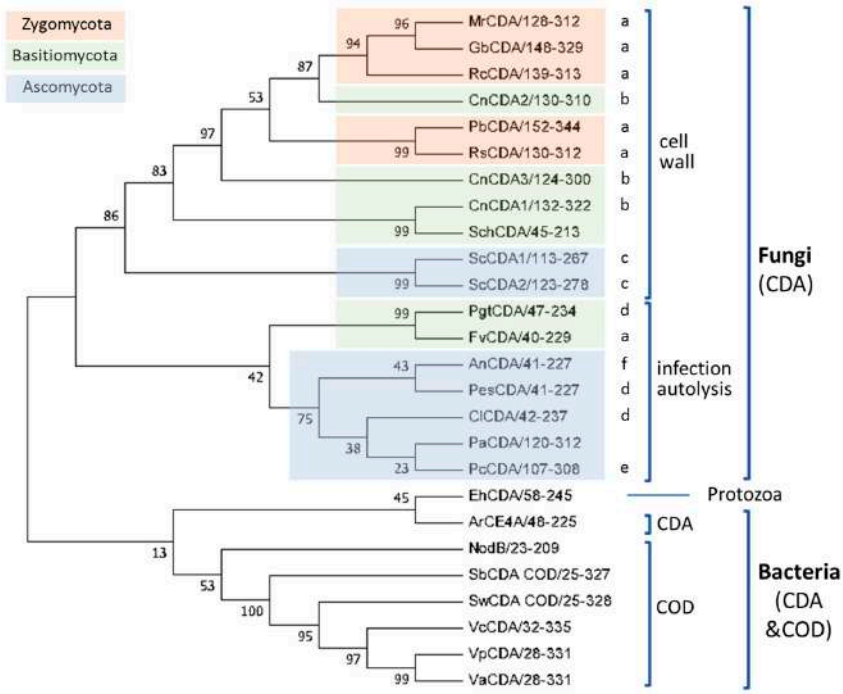


Figure 4. Phylogenetic analysis of CDAs from the multiple sequence alignment presented in Figure 3. A bootstrap analysis with 500 replicates was carried out on the trees inferred from the neighbor joining method. The consensus tree is shown with bootstrap values at each node of the tree. Biological functions: cell wall biosynthesis: (a) cell wall, (b) vegetative growth, (c) sporulation; host infection, (d) defense, (e) interaction/infection, (f) autolysis (see text).

4.4. Catalytic Mechanism

CDA enzymes operate by metal-assisted acid/base catalysis. The general mechanism was first proposed for the peptidoglycan GlcNAC deacetylase *Sp*PgdA when solving its X-ray structure [144] and short after supported by the 3D structures of the acetylxylan esterases *Sl*AxeA and *Ct*AxeA [148]. The catalytic machinery involves the conserved active site motifs containing the metal-binding triad and the general acid and base residues. Only the structures of four different CDA have been solved up to date (*Colletotrichum*, *Aspergillus*, *Vibrio*, and *Arthrobacter* CDAs, Table 1), all consistent with the proposed metal-assisted mechanism. *Vc*CDA was the first CE4 enzyme for which the 3D structure of enzyme-substrate complexes were solved by X-ray crystallography [69]. The structure of complexes of an inactive mutant (at the general base Asp residue) with diacetylchitobiose (DP2) and triacetylchitotriose (DP3) in productive binding for catalysis showed that a sugar hydroxyl group of the substrate also participates in metal coordination. Specifically (Figure 5), the Zn^{2+} cation is coordinated by the imidazole nitrogens of His97 and His101, the carboxylate group of Asp40, and the O7 atom of the *N*-acetyl group and O3 hydroxyl of the GlcNAc ring. The distorted octahedral coordination is completed by a water molecule. Upon activation, this water molecule is proposed to be the nucleophile responsible for removal of the *N*-acetyl group. Just recently, a second structure of an enzyme-substrate complex has been reported for the *Arthrobacter* sp. CDA (*Ar*CE4) [76]. The diacetylchitobiose ligand bound into the active site also shows the same type of interactions with the conserved active site residues.

The proposed mechanism of CDAs and related CE4 enzymes is shown in Figure 5. In the first step, metal coordination polarizes the carbonyl amide of the substrate which reacts with the nucleophilic water molecule activated by the general base (Asp), leading to a tetrahedral oxyanion

intermediate. Next, protonation of the nitrogen group of the intermediate by the general acid (His) facilitates C-N bond breaking with release of acetate and the generation of a free amine in product. Kinetic evidence for an oxyanion tetrahedral intermediate and significant charge development at the first transition state was provided by Hammett linear free energy correlations using the *Cl*CDA enzyme with α-haloacetamido substrate analogues [53]. In most of the enzymes, the catalytic acid and base residues are part of two conserved "charge relay" side chain pairs that may contribute to modulate the pK_a of the catalytic residues [53,144]: the catalytic base (Asp) is tethered by a conserved Arg from MT3 (RxxPY) and the catalytic acid (His) is tethered by a conserved Asp from MT4, DxxD(W/Y).

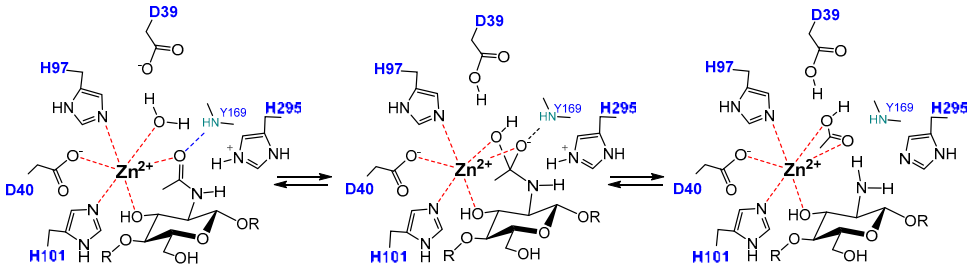


Figure 5. Metal-assisted general acid/base mechanism proposed for CE4 deacetylases. Scheme based on the 3D structure of the enzyme·substrate complex $VcCDA_{D39S}$·DP2 [61]. D39 is the general base and His295 is the general acid.

4.5. Determinants of Substrate Specificity

The series of crystal structures of the *Vibrio cholerae* chitin oligosaccharide deacetylase (*Vc*CDA or *Vc*COD) reported in 2014 [69] were the first 3D structures of a CE4 enzyme in complex with substrates. These data provided a first insight into structure-function relationships for this family of enzymes and highlighted the role of the loops that shape the binding site cleft in substrate binding. Recently by the end of 2017, the 3D structure of an *Arthrobacter* sp. CDA (ArCE4) in complex with substrate has also been reported [76]. Albeit sharing the same molecular function, both enzymes represent two different scenarios regarding the binding site topology and, hence, substrate specificity. *Vc*CDA has a rather closed binding cleft and is highly specific for monodeacetylation of COS, whereas *Ar*CE4 has a more open binding cleft and is able to fully deacetylate their COS substrates (Figure 6). A comparative structural analysis of both enzyme structures has been recently reviewed [142].

4.5.1. *Vc*CDA. Long Loops and High Specificity

Currently available structures of *Vc*CDA include the unliganded form of the enzyme and the binary complexes with *N*-acetylglucosamine (DP1), diacetylchitobiose (DP2), and triacetylchitotriose (DP3). These structures revealed two significant observations: a series of non-conserved loops (labeled Loop 1 to 6 in Figures 3 and 6) that shape the binding cleft, and the dynamics of the loops that assemble the active site for catalysis [69].

In all structures, the substrate is confined in a small binding cleft that is shaped by a series of long loops surrounding the active site (see Figure 6A for the *Vc*CDA·DP3 complex). Given the topology of *Vc*CDA protein surface, the binding of longer COS is prevented because these loops cap both the reducing and non-reducing ends of the substrate. Indeed, the catalytic efficiency of *Vc*CDA drops substantially on oligomers longer than DP2 [69] and attempts to solve the structure of *Vc*CDA bound to substrates longer than DP3 in a catalytically competent mode have been unsuccessful. Another consequence of this constricting topology is the high specificity of *Vc*CDA to exclusively deacetylate the penultimate residue from the non-reducing end of the substrates. There is no room for the ligand to slide along the binding cleft, thus it can only accommodate one GlcNAc unit of the

oligomeric chain at the catalytic site where deacetylation takes place. The binding of substrates induces a conformational change of Loop 4 from an open conformation in the unliganded enzyme to a closed conformation in the enzyme·DP2 complex or a semi-closed conformation in the enzyme·DP3 complex. It is triggered by a staking interaction between a Trp residue located in apical site of Loop 4 and the GlcNAc unit at the catalytic center, locking the substrate in the active site in the proper orientation for catalysis.

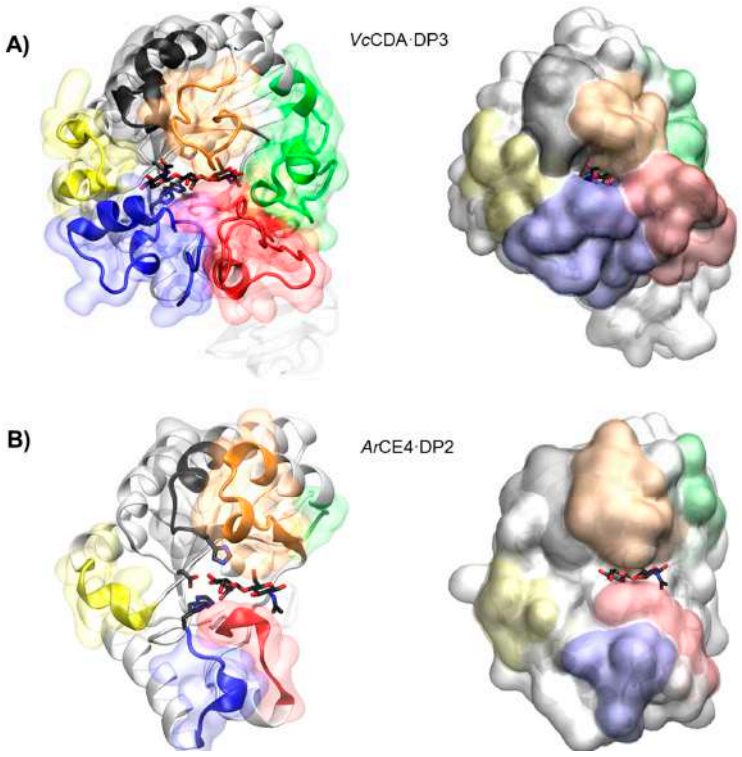


Figure 6. 3D structures of enzyme·substrate complex. (**A**) *Vc*CDA with DP3 substrate and (**B**) *Ar*CE with DP2 substrate. Loops 1 to 6 are colored as in Figure 3.

4.5.2. *Ar*CE4. Short Loops and Broad Specificity

In contrast to *Vc*CDA, the crystal structure of *Ar*CE4 in complex with diacetylchitobiose [76] reveals a flatter protein surface with the substrate bound to a more open binding cleft (Figure 6B). Even though the enzyme was co-crystalized with tetraacetylchitotetraose (DP4), only two GlcNAc units are observed in the structure. This indicates a weak binding of part of the COS substrate on this flat topology of the protein surface. The catalytic center in both enzymes is in the same position with respect to the protein core, being the main difference, the size and shape of the loops surrounding the active site. Since the binding cleft is more open, the enzyme can accommodate longer COS. Indeed, the enzymatic activity of *Ar*CE4 increases as the length of the chitin oligomer chain increases. The lack of protein caps at either the reducing and non-reducing ends of the substrate can also explain the multiple-chain mechanism proposed for this enzyme. Deacetylation takes place at all GlcNAc units of the substrate (except the reducing end) because it can freely bind to *Ar*CE4 in different binding modes exposing different GlcNAc units of the oligomeric chain at the catalytic site.

4.5.3. The Subsite Capping Model

The diversity of deacetylation patterns exhibited by chitin deacetylases and related CE4 enzymes can be attributed to the differential accessibility of the linear chitin oligosaccharide chain to the separate subsites along the substrate binding cleft of their structures. Considering all CE4 enzymes with reported activity on polymeric chitin or COS, these can be classified into two groups. One group is represented by general chitin deacetylases (CDA), and a second group is formed by chitin oligosaccharide deacetylases (COD). The two structures of the enzymes-substrate complexes described above are reference models for the protein surface topologies and substrate binding mechanisms of these two groups: CDAs (*Ar*CE4) and CODs (*Vc*CDA). These two structures provide a unified view of the determinants of substrate specificity in chitin deacetylases in terms of the "subsite capping model" proposed in [69]. According to this model, substrate accessibility is affected by the length, shape, and dynamics of a series of loops surrounding the active site of CE4 enzymes. These loops are numbered from 1 to 6 and their location in the sequences and structures of CDAs and CODs is highlighted in Figures 2 and 3.

The group of CDAs bears short loops, and their structures exhibit a flat and open binding cleft. The substrate binding mechanism in this group of enzymes may be similar to that described for the reference structure of *Ar*CE4. According to the model, the substrate may be able to slide along the binding cleft or to bind in different modes resulting in processive or multiple-chain attack mechanisms of deacetylation. This can already be anticipated for CDA enzymes of known structure (Figure 2) because the flat protein surface is already evident, but also for CDA enzymes of unknown structure given the similar sequence lengths of the loops evidenced in the alignment (Figure 3). This could be an explanation of why CDA enzymes in general are not specific for the deacetylation at a single *N*-acetylglucosamine unit. However, the patterns of deacetylation differ among the different CDAs. The surface charge distribution along the binding cleft and other structural features yet to be disclosed may also participate in defining the mode of action and deacetylation pattern by each particular enzyme.

On the contrary, the group of COD enzymes bears longer loops and their structures have narrower binding pockets and buried active sites. According to the subsite capping model, the substrate is constrained to bind in very specific binding modes resulting in single-site deacetylations. This is the case for the reference structure of *Vc*CDA in complex with substrates, but it can also be anticipated for other COD enzymes for which the 3D structure is still unknown. For instance, Loop 6 in *Rm*NodB is longer than in other CDAs. This loop is located on the non-reducing end site of the binding cleft and may cap the accessibility of the substrate after subsite 0 (the catalytic site) thus defining the deacetylation specificity for the non-reducing end of the substrate. Likewise, the *Shewanella* CODs have a Loop 6 with the same length than the *Vibrio* CODs, but shorter than NodB, and both exhibit the same mono-deacetylation specificity for the penultimate GlcNAc residue from the reducing end of the substrate.

For most CDAs, the reducing end of the substrate is not deacetylated, or it is the least reactive GlcNAc unit. As seen in the *Ar*CE4·DP2 complex 3D structure [64], binding to the +1 subsite seems to be dominated by the stacking interaction of the GlcNAc unit of the substrate with a Trp in Motif 4 at the beginning of Loop 4. This aromatic residue is highly conserved (MT4, DxxD(W/Y), Figure 3). CDA enzymes having this aromatic residue prefer a sugar bound in the +1 subsite; they do not deacetylate the reducing end of their substrates, as it is the case for *Ar*CE4, *Pes*CDA [89] and *Pc*CDA [95], or the reducing end is the slowest position to be deacetylated, as shown for *Cl*CDA [49] and *An*CDA [12]. On the contrary, *Pgt*CDA, which deacetylates the reducing end GlcNAc unit of all substrates from DP4 to DP6, lacks the +1 aromatic residue [89]. Different is the case of *Vibrio* and *Sewanella* CODs that have the equivalent aromatic residue in a slightly different position after a two-amino acid insertion in the MT4 motif, and it is located farther in Loop 4 (Figure 3). In the *Vc*CDA enzyme this loop moves from an open to a closed conformation upon substrate binding, and the same is expected for the other closely related CODs that have the same Loop 4 size. As a consequence of the induced fit, the Trp

residue now establishes a stacking interaction with the GlcNAc unit in subsite 0. DP2 is the preferred substrate for this group of COD enzymes, and it is deacetylated at the reducing end [62].

Gaining further structural information of protein-ligand complexes of CDA and COD enzymes, and other CE4 in general, will contribute further to decipher the structural and sequential determinants of substrate specificity in this family of enzymes. This will pave the way to the rational design or discovery of novel CDA with controlled specificities on the deacetylation of oligomeric and polymeric chitin for the biotechnological production of chitosans and paCOS with defined patterns of acetylation.

5. Application of Chitin Deacetylases

5.1. Targets for Antifungals

Fungal infections have an enormous impact on human health. Fungi are generally opportunistic pathogens affecting immunocompromised individuals including those with AIDS, receiving immunosuppressive drugs or undergoing cancer treatments. The cell wall is a meaningful target for antifungal therapies. Current major classes of antifungal drugs target cell membrane ergosterol biosynthesis (azoles), ergosterol function by disrupting membrane integrity (polyenes), or 1,3-β-glucan synthase preventing the formation of the cell-wall structural polysaccharide 1,3-β-glucan (echinocandins) [149]. New targets to overcome the emerging drug resistance by pathogenic fungi are becoming critical to treat life-threatening fungal infections. Other promising targets are the so called cell wall proteins (CWP) which mediate important cellular processes, including adhesion, invasion, biofilm formation and flocculation [122]. In fungal chitosan producers, chitin deacetylases are a class of CWP and potential targets for drug design. *Cryptococcus neoformans*, one of the most deadly pathogens, requires chitosan for virulence. Lack of chitosan in the cell wall has detrimental consequences in fungal growth and results in the complete loss of sporulation [120,121]. Thus, CDAs represent a promising target for anticryptococcal therapeutics [37,120], but no CDA inhibitors have been reported yet.

In pathogenic plants, major strategies to prevent fungal pathogenesis are related to the inhibition of fungal chitinases, which are required for chitin remodeling in the cell wall [150–152]. Different types of chitinase inhibitors have been reported, including potent natural inhibitors such as allosamidin [153] and the cyclic pentapeptides argifin and argadin [154]. However, inhibition must be selective so as not to interfere with the plant chitinases involved in triggering the plant defense mechanisms. Another potential and promising strategy is the inhibition of extracellularly secreted fungal CDAs since they constitute a defense mechanism to evade the plant immune system, as discussed in Section 3.2. As in the case of human fungal pathogens, no inhibitors have been yet reported against CDAs from plant pathogenic fungi.

5.2. Biocatalysts for the Enzymatic Production of Chitosans and paCOS

Chitosans can be found in a large number of applications in such distant areas as agriculture, cosmetics, water treatment, medicine and the food industry [155–159]. In addition to chitosan polymers, their oligomers (paCOS) have also proven to have relevant potential applications in agriculture and pharmaceutical industries [160]. The physicochemical and biological properties of chitosans and paCOS have been shown to be strongly dependent on their degree of polymerization and their degree of acetylation [161,162]. Many of the identified CDAs arose from screening programs addressed to find efficient biocatalysts to overcome the current industrial chitosan production by highly concentrated alkali treatment of chitin. Some examples include CDAs from *Mortiriella* sp. [58], *Rhizopus* sp. [117], or *Gongronella* sp. [118].

Chemical methods for the production of COS and paCOS are based on chemical depolymerization of chitosan [163,164], total synthesis of chitosan oligomers [165,166], partial chemical deacetylation of fully acetylated COS, or chemical re-*N*-acetylation of glucosamine oligomers based in two-step procedures or one-pot synthesis [167–169]. The drawbacks of chemical strategies are the unwanted

side reactions and the randomness of the chemical reactions. Current efforts are addressed to develop enzymatic routes for COS and paCOS production with defined DP, DA, and PA.

Enzymatic approaches include depolymerization of chitin or chitosan polymers using hydrolytic enzymes, chitinases and chitosanases, enzymatic polymerization by transglycosidation using transglycosylating hydrolases, and enzymatic de-acetylation and re-acetylation of chitin oligomers using chitin deacetylases, strategies recently reviewed in [170]. Based on the current knowledge on the specificity of a number of fungal and bacterial CDAs, recent reports have combined enzymes with different specificities to have access to a large family of paCOS with defined structures. The first proof of concept was to show that two specific CODs, NodB and VcCDA, each accept the monodeacetylated product from the other, leading to specific di-deacetylation, and that both enzymes can work in a one-pot process [130]. Recently, the use of different recombinant CDAs from bacterial and fungal origin to produce all of fourteen possible partially acetylated chitosan tetramers combining different enzymatic deacetylations and enzymatic *N*-acetylations has been reported [171] (Figure 7).

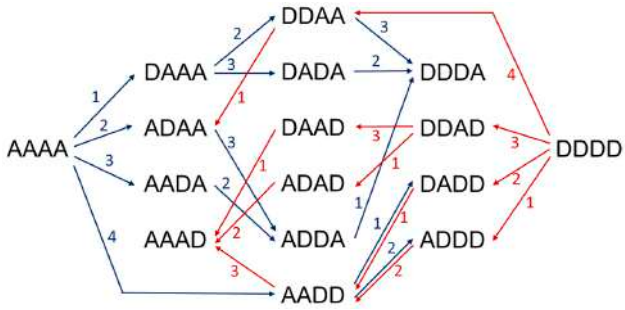


Figure 7. Production routes of all possible chitin and chitosan tetramers using 4 different CDAs to specifically deacetylate or *N*-acetylate paCOS. A: GlcNAc, D: $GlcNH_2$. Blue arrows, deacetylation reactions, red arrows, *N*-acetylation reactions in the presence of excess acetate.

The production of paCOS using in vivo strategies is an alternative to increase the scalability of the process. The first example towards a more general cell factory approach for the in vivo synthesis of paCOS was based on NodB deacetylase. By in vivo studies with *Escherichia coli* expressing different combinations of the nodABCS genes of *Azorhizobium caulinodans*, Nod factor intermediates were identified, as well as the sequence of the biosynthetic steps [172]. The nod gene cluster encodes a series of enzymes, which include the NodC chitin oligosaccharide synthase that produces fully acetylated chitin oligomers, the NodB chitin oligosaccharide deacetylase that deacetylates the non-reducing end unit, the NodA *N*-acyl transferase that transfer a fatty acid chain to the free amine group, and the NodS *N*-methyl transferase. Further transformations by other nod proteins elaborate the final Nod signaling factors [173]. In a first cell factory approach, high density cells of *E. coli* expressing nodC or nodBC genes produced in high yield (up to 2.5 g/L) penta-*N*-acetyl-chitopentaose and its deacetylated derivative tetra-*N*-acetyl-chitopentaose, which were easily purified by charcoal adsorption and ion-exchange chromatography [174]. The strategy was further extended to the production of sulfated and *O*-acetylated derivatives of these two compounds by coexpressing nodC or nodBC with nodH and/or nodL that encode chitooligosaccharide sulfotransferase and chitooligosaccharide *O*-acetyltransferase, respectively [175]. Other Nod analogues have also been generated with further modifications [176–178]. The cell factory approach, currently limited to one deacelylation pattern based on the use of NodB, is a promising technology to be developed by incorporating the diversity of CDAs with different deacetylation patterns in order to access a large family of paCOS and derivatives.

6. Conclusions

We have here summarized the current knowledge on substrate specificity of fungal and bacterial chitin deacetylases, their modes of action, and their use as biocatalysts for the production of chitosans and chitosan oligosaccharides with defined pattern of acetylation. By combining multiple sequence alignments and 3D structures of enzyme·substrate complexes of representative enzymes, a unified view of the determinants of substrate specificity is proposed in terms of the "subsite capping model." According to this model, substrate accessibility is affected by the length, shape, and dynamics of a series of loops surrounding the active site of CE4 enzymes. The group of CDAs active on polymeric substrates and COS bear short loops, and their structures exhibit a flat and open binding cleft. The substrate may be able to slide along the binding cleft or to bind in different modes, resulting in processive or multiple-chain attack mechanisms of deacetylation. Other structural features not yet disclosed, such as the charge distribution along the binding cleft may also participate in defining the mode of action and deacetylation pattern by each particular enzyme. The group of COD enzymes active on low molecular mass COS bear longer loops and their structures have narrower binding pockets and buried active sites. The substrate is constrained to bind in very specific binding modes resulting in single-site deacetylations.

But a deeper knowledge on substrate specificity requires further structural information of protein-ligand complexes of CDA and COD enzymes in order to decipher the structural and sequential determinants of substrate specificity in this family of enzymes aimed at the rational design or discovery of novel CDAs with controlled specificities on the deacetylation of oligomeric and polymeric chitin for biotechnological applications.

Although CDAs have been proposed as targets for antifungal drugs, no specific inhibitors have been yet reported. This is an open field that deserves attention not only for drug design but also to probe the signaling function of CDAs and CODs through their specific deacetylation of COS substrates.

Applications of CDAs and CODs as biocatalyst are currently being developed as a novel methodology to produce partially acetylated COS with tailored patterns of acetylation. Since not all patterns for COS of different sizes are yet available, enzyme discovery and protein engineering offer new opportunities for the biotechnological production of chitosans and paCOS with defined patterns of acetylation.

Acknowledgments: Work supported by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement n°613931, and grant BFU2016-77427-C2-1-R from MINECO, Spain. Laia Grifoll and Sergi Pascual acknowledge a predoctoral contract from the NANO3BIO project. Hugo Aragunde acknowledges a predoctoral fellowship from Generalitat de Catalunya.

Author Contributions: Antoni Planas conceived and designed the review; Laia Grifoll, Sergi Pascual, Hugo Aragunde, Xevi Biarnés, and Antoni Planas compiled the references and prepared the figures. Xevi Biarnés and Antoni Planas wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

AXE	Acetylxylan esterase
A_n	$(GlcNAc)_n$
CDA	Chitin deacetylase
COS	Chitooligosaccharides
D_n	$(GlcNH_2)_n$
DA	Degree of acetylation
DP	Degree of polymerization
GlcNAc	*N*-acetylglucosamine
PA	Pattern of acetylation
paCOS	Partially acetylated chiton oligosaccharides
PDB	Protein data bank

References

1. Peniche Covas, C.A.; Argüelles-Monal, W.; Goycoolea, F.M. Chitin and chitosan: Major sources, properties and applications. In *Monomers, Polymers and Composites from Renewable Resources*; Belgacem, M.N., Gandini, A., Eds.; Elsevier: Amsterdam, the Netherlands, 2008; Volume 1, pp. 517–542. ISBN 9780080453163.
2. Karrer, P.; Hofmann, A. Über den enzymatischen Abbau von Chitin und Chitosan I. *Helv. Chim. Acta* **1929**, *12*, 616–637. [CrossRef]
3. Rinaudo, M. Chitin and chitosan: Properties and applications. *Prog. Polym. Sci.* **2006**, *31*, 603–632. [CrossRef]
4. Dutta, P.K.; Duta, J.; Tripathi, V.S. Chitin and Chitosan: Chemistry, properties and applications. *J. Sci. Ind. Res. (India)* **2004**, *63*, 20–31. [CrossRef]
5. Noishiki, Y.; Takami, H.; Nishiyama, Y.; Wada, M.; Okada, S.; Kuga, S. Alkali-induced conversion of β-chitin to α-chitin. *Biomacromolecules* **2003**, *4*, 896–899. [CrossRef] [PubMed]
6. Jang, M.K.; Kong, B.G.; Jeong, Y.I.; Lee, C.H.; Nah, J.W. Physicochemical characterization of α-chitin, β-chitin, and γ-chitin separated from natural resources. *J. Polym. Sci. Part A Polym. Chem.* **2004**, *42*, 3423–3432. [CrossRef]
7. Kumirska, J.; Czerwicka, M.; Kaczyński, Z.; Bychowska, A.; Brzozowski, K.; Thöming, J.; Stepnowski, P. Application of spectroscopic methods for structural analysis of chitin and chitosan. *Mar. Drugs* **2010**, *8*, 1567–1636. [CrossRef] [PubMed]
8. Kaya, M.; Mujtaba, M.; Ehrlich, H.; Salaberria, A.M.; Baran, T.; Amemiya, C.T.; Galli, R.; Akyuz, L.; Sargin, I.; Labidi, J. On chemistry of γ-chitin. *Carbohydr. Polym.* **2017**, *176*, 177–186. [CrossRef] [PubMed]
9. Dhillon, G.S.; Kaur, S.; Brar, S.K.; Verma, M. Green synthesis approach: Extraction of chitosan from fungus mycelia. *Crit. Rev. Biotechnol.* **2013**, *33*, 379–403. [CrossRef] [PubMed]
10. Hoell, I.A.; Vaaje-Kolstad, G.; Eijsink, V.G.H. Structure and function of enzymes acting on chitin and chitosan. *Biotechnol. Genet. Eng. Rev.* **2010**, *27*, 331–366. [CrossRef]
11. Xia, W.; Liu, P.; Zhang, J.; Chen, J. Biological activities of chitosan and chitooligosaccharides. *Food Hydrocoll.* **2011**, *25*, 170–179. [CrossRef]
12. Yu, R.; Liu, W.; Li, D.; Zhao, X.; Ding, G.; Zhang, M.; Ma, E.; Zhu, K.Y.; Li, S.; Moussian, B.; et al. Helicoidal organization of chitin in the cuticle of the migratory locust requires the function of the chitin deacetylase2 enzyme (LmCDA2). *J. Biol. Chem.* **2016**, *291*, 24352–24363. [CrossRef] [PubMed]
13. Winkler, A.J.; Dominguez-Nuñez, J.A.; Aranaz, I.; Poza-Carrión, C.; Ramonell, K.; Somerville, S.; Berrocal-Lobo, M. Short-chain chitin oligomers: Promoters of plant growth. *Mar. Drugs* **2017**, *15*, 40. [CrossRef] [PubMed]
14. Li, X.; Min, M.; Du, N.; Gu, Y.; Hode, T.; Naylor, M.; Chen, D.; Nordquist, R.E.; Chen, W.R. Chitin, chitosan, and glycated chitosan regulate immune responses: The novel adjuvants for cancer vaccine. *Clin. Dev. Immunol.* **2013**, *2013*. [CrossRef] [PubMed]
15. Varki, A.; Sharon, N. *Essentials of Glycobiology*, 2nd ed.; Varki, A., Cummings, R., Esko, J., Freeze, H., Stanley, P., Bertozzi, C.R., Hart, G., Etzler, M.E., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 2009.
16. Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P.M.; Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **2014**, *42*, 490–495. [CrossRef] [PubMed]
17. Caufrier, F.; Martinou, A.; Dupont, C.; Bouriotis, V. Carbohydrate esterase family 4 enzymes: Substrate specificity. *Carbohydr. Res.* **2003**, *338*, 687–692. [CrossRef]
18. John, M.; Rohrig, H.; Schmidt, J.; Wieneke, U.; Schell, J. Rhizobium NodB protein involved in nodulation signal synthesis is a chitooligosaccharide deacetylase. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 625–629. [CrossRef] [PubMed]
19. Mine, S.; Niiyama, M.; Hashimoto, W.; Ikegami, T.; Koma, D.; Ohmoto, T.; Fukuda, Y.; Inoue, T.; Abe, Y.; Ueda, T.; et al. Expression from engineered *Escherichia coli* chromosome and crystallographic study of archaeal *N*,*N*-diacetylchitobiose deacetylase. *FEBS J.* **2014**, *281*, 2584–2596. [CrossRef] [PubMed]
20. Fadouloglou, V.E.; Deli, A.; Glykos, N.M.; Psylinakis, E.; Bouriotis, V.; Kokkinidis, M. Crystal structure of the BcZBP, a zinc-binding protein from *Bacillus cereus*. *FEBS J.* **2007**, *274*, 3044–3054. [CrossRef] [PubMed]
21. Verma, S.C.; Mahadevan, S. The ChbG gene of the chitobiose (chb) operon of *Escherichia coli* encodes a chitooligosaccharide deacetylase. *J. Bacteriol.* **2012**, *194*, 4959–4971. [CrossRef] [PubMed]
22. Araki, Y.; Ito, E. A pathway of chitosan formation in *Mucor rouxii*: Enzymatic deacetlation of chitin. *Biochem. Biophys. Res. Commun.* **1974**, *56*, 669–675. [CrossRef]

23. Araki, Y.; Ito, E. A Pathway of Chitosan Formation in *Mucor rouxii* Enzymatic Deacetylation of Chitin. *Eur. J. Biochem.* **1975**, *55*, 71–78. [CrossRef] [PubMed]
24. Zhao, Y.; Ju, W.; Jo, G.; Jung, W.; Park, R. Perspectives of Chitin Deacetylase Research. In *Biotechnology of Biopolymers*; InTech: London, UK, 2011; pp. 131–145. [CrossRef]
25. Liu, Z.; Gay, L.M.; Tuveng, T.R.; Agger, J.W.; Westereng, B.; Mathiesen, G.; Horn, S.J.; Vaaje-Kolstad, G.; van Aalten, D.M.F.; Eijsink, V.G.H. Structure and function of a broad-specificity chitin deacetylase from *Aspergillus nidulans* FGSC A4. *Sci. Rep.* **2017**, *7*, 1746. [CrossRef] [PubMed]
26. Hoßbach, J.; Bußwinkel, F.; Kranz, A.; Wattjes, J.; Cord-Landwehr, S.; Moerschbacher, B.M. A chitin deacetylase of *Podospora anserina* has two functional chitin binding domains and a unique mode of action. *Carbohydr. Polym.* **2018**, *183*, 1–10. [CrossRef] [PubMed]
27. Ghormade, V.; Kulkarni, S.; Doiphode, N.; Rajamohanan, P.R.; Deshpande, M.V. Chitin deacetylase: A comprehensive account on its role in nature and its biotechnological applications. In *Current Research, Technology and Education Topics in Applied Microbiology and Microbial Biotechnology*; Méndez-Vilas, A., Ed.; Formatex Research Center: Badajoz, Spain, 2010; pp. 1054–1066.
28. Tsigos, I.; Martinou, A.; Kafetzopoulos, D.; Bouriotis, V. Chitin deacetylases: New, versatile tools in biotechnology. *Trends Biotechnol.* **2000**, *18*, 305–312. [CrossRef]
29. Zhao, Y.; Park, R.D.; Muzzarelli, R.A.A. Chitin deacetylases: Properties and applications. *Mar. Drugs* **2010**, *8*, 24–46. [CrossRef] [PubMed]
30. Martinou, A.; Bouriotis, V.; Stokke, B.T.; Vårum, K.M. Mode of action of chitin deacetylase from *Mucor rouxii* on partially *N*-acetylated chitosans. *Carbohydr. Res.* **1998**, *311*, 71–78. [CrossRef]
31. Tsigos, I.; Zydowicz, N.; Martinou, A.; Domard, A.; Bouriotis, V. Mode of action of chitin deacetylase from *Mucor rouxii* on *N*-acetylchitooligosaccharides. *Eur. J. Biochem.* **1999**, *261*, 698–705. [CrossRef] [PubMed]
32. Tokuyasu, K.; Mitsutomi, M.; Yamaguchi, I.; Hayashi, K.; Mori, Y. Recognition of chitooligosaccharides and their *N*-acetyl groups by putative subsites of chitin deacetylase from a Deuteromycete, *Colletotrichum lindemuthianum*. *Biochemistry* **2000**, *39*, 8837–8843. [CrossRef] [PubMed]
33. Hekmat, O.; Tokuyasu, K.; Withers, S.G. Subsite structure of the endo-type chitin deacetylase from a deuteromycete, *Colletotrichum lindemuthianum*: An investigation using steady-state kinetic analysis and MS. *Biochem. J.* **2003**, *374*, 369–380. [CrossRef] [PubMed]
34. Gooday, G.W. Chitin deacetylases in invertebrates. In *Chitin in Nature and Technology*; Springer: Boston, MA, USA, 1986; pp. 263–267.
35. Muthukrishnan, S.; Merzendorfer, H.; Arakane, Y.; Yang, Q. Chitin Metabolic Pathways in Insects and Their Regulation. In *Extracellular Composite Matrices in Arthropods*; Cohen, E., Moussian, B., Eds.; Springer: Berlin, Germany, 2016; pp. 31–65. ISBN 9783319407401.
36. Dixit, R.; Arakane, Y.; Specht, C.A.; Richard, C.; Kramer, K.J.; Beeman, R.W.; Muthukrishnan, S. Domain organization and phylogenetic analysis of proteins from the chitin deacetylase gene family of Tribolium castaneum and three other species of insects. *Insect Biochem. Mol. Biol.* **2008**, *38*, 440–451. [CrossRef] [PubMed]
37. Baker, L.G.; Specht, C.A.; Donlin, M.J.; Lodge, J.K. Chitosan, the deacetylated form of chitin, is necessary for cell wall integrity in *Cryptococcus neoformans*. *Eukaryot. Cell* **2007**, *6*, 855–867. [CrossRef] [PubMed]
38. Christodoulidou, A.; Bouriotis, V.; Thireos, G. Two sporulation-specific chitin deacetylase-encoding genes are required for the ascospore wall rigidity of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **1996**, *271*, 31420–31425. [CrossRef] [PubMed]
39. Geoghegan, I.A.; Gurr, S.J. Chitosan Mediates Germling Adhesion in Magnaporthe oryzae and Is Required for Surface Sensing and Germling Morphogenesis. *PLoS Pathog.* **2016**, *12*, 1–34. [CrossRef] [PubMed]
40. White, S.; McIntyre, M.; Berry, D.R.; McNeil, B. The autolysis of industrial filamentous fungi. *Crit. Rev. Biotechnol.* **2002**, *22*, 1–14. [CrossRef] [PubMed]
41. Sánchez-Vallet, A.; Mesters, J.R.; Thomma, B.P. The battle for chitin recognition in plant-microbe interactions. *FEMS Microbiol. Rev.* **2015**, *39*, 171–183. [CrossRef] [PubMed]
42. Davis, L.L.; Bartnicki-Garcia, S. Chitosan Synthesis by the Tandem Action of Chitin Synthetase and Chitin Deacetylase from *Mucor rouxii*. *Biochemistry* **1984**, *23*, 1065–1073. [CrossRef]
43. Gao, X.-D.; Katsumoto, T.; Onodera, K. Purification and characterization of chitin deacetylase from *Absidia coerulea*. *J. Biochem.* **1995**, *117*, 257–263. [CrossRef] [PubMed]

44. Christodoulidou, A.; Briza, P.; Ellinger, A.; Bouriotis, V. Yeast ascospore wall assembly requires two chitin deacetylase isozymes. *FEBS Lett.* **1999**, *460*, 275–279. [CrossRef]
45. Hadwiger, L.A. Anatomy of a nonhost disease resistance response of pea to Fusarium solani: PR gene elicitation via DNase, chitosan and chromatin alterations. *Front. Plant Sci.* **2015**, *6*, 373. [CrossRef] [PubMed]
46. Hadwiger, L.A. Pea-Fusarium solani interactions contributions of a system toward understanding disease resistance. *Phytopathology* **2008**, *98*, 372–379. [CrossRef] [PubMed]
47. El Gueddari, N.E.; Rauchhaus, U.; Moerschbacher, B.M.; Deising, H.B. Developmentally regulated conversion of surface-exposed chitin to chi-tosan in cell walls of plant pathogenic fungi. *New Phytol.* **2002**, *156*, 103–112. [CrossRef]
48. Liu, T.; Liu, Z.; Song, C.; Hu, Y.; Han, Z.; She, J.; Fan, F.; Wang, J.; Jin, C.; Chang, J.; et al. Chitin-Induced Dimerization Activates a Plant Immune Receptor. *Science* **2012**, *336*, 1160–1164. [CrossRef] [PubMed]
49. Cord-Landwehr, S.; Melcher, R.L.J.; Kolkenbrock, S.; Moerschbacher, B.M. A chitin deacetylase from the endophytic fungus *Pestalotiopsis* sp. efficiently inactivates the elicitor activity of chitin oligomers in rice cells. *Sci. Rep.* **2016**, *6*, 38018. [CrossRef] [PubMed]
50. Emri, T.; Molnár, Z.; Szilágyi, M.; Pócsi, I. Regulation of autolysis in *Aspergillus nidulans*. *Appl. Biochem. Biotechnol.* **2008**, *151*, 211–220. [CrossRef] [PubMed]
51. Alfonso, C.; Nuero, O.M.; Santamaría, F.; Reyes, F. Purification of a heat-stable chitin deacetylase from *Aspergillus nidulans* and its role in cell wall degradation. *Curr. Microbiol.* **1995**, *30*, 49–54. [CrossRef] [PubMed]
52. Reyes, F.; Calatayud, J.; Martinez, M.J. Endochitinase from *Aspergillus nidulans* implicated in the autolysis of its cell wall. *FEMS Microbiol. Lett.* **1989**, *51*, 119–124. [CrossRef] [PubMed]
53. Blair, D.E.; Hekmat, O.; Schüttelkopf, A.W.; Shrestha, B.; Tokuyasu, K.; Withers, S.G.; van Aalten, D.M. Structure and Mechanism of Chitin Deacetylase from the Fungal Pathogen Colletotrichium lindemuthianum. *Biochemistry* **2006**, *45*, 9416–9426. [CrossRef] [PubMed]
54. Naqvi, S.; Cord-Landwehr, S.; Singh, R.; Bernard, F.; Kolkenbrock, S.; El Gueddari, N.E.; Moerschbacher, B.M. A recombinant fungal chitin deacetylase produces fully defined chitosan oligomers with novel patterns of acetylation. *Appl. Environ. Microbiol.* **2016**, *82*, 6645–6655. [CrossRef] [PubMed]
55. Aranda-Martinez, A.; Grifoll-Romero, L.; Aragunde Pazos, H.; Enea Sancho-Vaello; Biarnés, X.; Lopez-Llorca, L.V.; Planas, A. Expression and specificity of a chitin deacetylase catalytic doain from the nematophagous fungus *Pochonia chlamydosporia* potentially involved in pathogenicity. *Sci. Rep.* **2018**, *8*, 2170. [CrossRef] [PubMed]
56. Mishra, C.; Mishra, C.; Semino, C.; Semino, C.; Mccreath, K.J.; Mccreath, K.J.; de la Vega, H.; de la Vega, H.; Jones, B.J.; Jones, B.J.; et al. Cloning and expression of two chitin deacetylase gens of *Saccharomyces cerevisiae*. *Yeast* **1997**, *13*, 327–336. [CrossRef]
57. Martinou, A.; Koutsioulis, D.; Bouriotis, V. Cloning and expression of a chitin deacetylase gene (CDA2) from *Saccharomyces cerevisiae* in *Escherichia coli*: Purification and characterization of the cobalt-dependent recombinant enzyme. *Enzyme Microb. Technol.* **2003**, *32*, 757–763. [CrossRef]
58. Kim, Y.J.; Zhao, Y.; Oh, K.T.; Nguyen, V.N.; Park, R.D. Enzymatic deacetylation of chitin by extracellular chitin deacetylase from a newly screened *Mortierella* sp. DY-52. *J. Microbiol. Biotechnol.* **2008**, *18*, 759–766. [PubMed]
59. Zhao, Y.; Kim, Y.J.; Oh, K.T.; Nguyen, V.N.; Park, R.D. Production and characterization of extracellular chitin deacetylase from *Absidia corymbifera* DY-9. *J. Appl. Biol. Chem.* **2010**, *53*, 119–126. [CrossRef]
60. Yamada, M.; Kurano, M.; Inatomi, S.; Taguchi, G.; Okazaki, M.; Shimosaka, M. Isolation and characterization of a gene coding for chitin deacetylase specifically expressed during fruiting body development in the basidiomycete *Flammulina velutipes* and its expression in the yeast *Pichia pastoris*. *FEMS Microbiol. Lett.* **2008**, *289*, 130–137. [CrossRef] [PubMed]
61. Pareek, N.; Vivekanand, V.; Saroj, S.; Sharma, A.K.; Singh, R.P. Purification and characterization of chitin deacetylase from *Penicillium oxalicum* SAEM-51. *Carbohydr. Polym.* **2012**, *87*, 1091–1097. [CrossRef]
62. Karthik, N.; Binod, P.; Pandey, A. SSF production, purification and characterization of chitin deacetylase from *Aspergillus flavus*. *Biocatal. Biotransform.* **2017**. [CrossRef]
63. Cai, J.; Yang, J.; Du, Y.; Fan, L.; Qiu, Y.; Li, J.; Kennedy, J.F. Purification and characterization of chitin deacetylase from *Scopulariopsis brevicaulis*. *Carbohydr. Polym.* **2006**, *65*, 211–217. [CrossRef]

64. Gauthier, C.; Clerisse, F.; Dommes, J.; Jaspar-Versali, M.F. Characterization and cloning of chitin deacetylases from *Rhizopus circinans*. *Protein Expr. Purif.* **2008**, *59*, 127–137. [CrossRef] [PubMed]
65. Maw, T.; Tan, T.K.; Khor, E.; Wong, S.M. Complete cDNA sequence of chitin deacetylase from *Gongronella butleri* and its phylogenetic analysis revealed clusters corresponding to taxonomic classification of fungi. *J. Biosci. Bioeng.* **2002**, *93*, 376–381. [CrossRef]
66. Mélida, H.; Sain, D.; Stajich, J.E.; Bulone, V. Deciphering the uniqueness of Mucoromycotina cell walls by combining biochemical and phylogenomic approaches. *Environ. Microbiol.* **2015**, *17*, 1649–1662. [CrossRef] [PubMed]
67. Smirnou, D.; Krcmar, M.; Prochazkova, E.V.A. Chitin-Glucan complex production by *Schizophyllum commune* submerged cultivation. *Pol. J. Microbiol.* **2011**, *60*, 223–228. [PubMed]
68. Das, S.; Van Dellen, K.; Bulik, D.; Magnelli, P.; Cui, J.; Head, J.; Robbins, P.W.; Samuelson, J. The cyst wall of *Entamoeba invadens* contains chitosan (deacetylated chitin). *Mol. Biochem. Parasitol.* **2006**, *148*, 86–92. [CrossRef] [PubMed]
69. Andrés, E.; Albesa-Jové, D.; Biarnés, X.; Moerschbacher, B.M.; Guerin, M.E.; Planas, A. Structural basis of chitin oligosaccharide deacetylation. *Angew. Chem. Int. Ed.* **2014**, *53*, 6882–6887. [CrossRef] [PubMed]
70. Li, X.; Wang, L.X.; Wang, X.; Roseman, S. The chitin catabolic cascade in the marine bacterium *Vibrio cholerae*: Characterization of a unique chitin oligosaccharide deacetylase. *Glycobiology* **2007**, *17*, 1377–1387. [CrossRef] [PubMed]
71. Hirano, T.; Sugiyama, K.; Sakaki, Y.; Hakamata, W.; Park, S.Y.; Nishio, T. Structure-based analysis of domain function of chitin oligosaccharide deacetylase from *Vibrio parahaemolyticus*. *FEBS Lett.* **2015**, *589*, 145–151. [CrossRef] [PubMed]
72. Kadokura, K.; Rokutani, A.; Yamamoto, M.; Ikegami, T.; Sugita, H.; Itoi, S.; Hakamata, W.; Oku, T.; Nishio, T. Purification and characterization of *Vibrio parahaemolyticus* extracellular chitinase and chitin oligosaccharide deacetylase involved in the production of heterodisaccharide from chitin. *Appl. Microbiol. Biotechnol.* **2007**, *75*, 357–365. [CrossRef] [PubMed]
73. Ohishi, K.; Yamagishi, M.; Ohta, T.; Motosugi, M.; Izumida, H.; Sano, H.; Adachi, K.; Miwa, T. Purification and Properties of Two Deacetylases Produced by *Vibrio alginolyticus* H-8. *Biosci. Biotechnol. Biochem.* **1997**, *61*, 1113–1117. [CrossRef]
74. Hirano, T.; Uehara, R.; Shiraishi, H.; Hakamata, W.; Nishio1, T. Chitin Oligosaccharide Deacetylase from *Shewanella woodyi* ATCC51908. *J. Appl. Glycosci.* **2015**, *62*, 153–157. [CrossRef]
75. Hirano, T.; Shiraishi, H.; Ikejima, M.; Uehara, R.; Hakamata, W.; Nishio, T. Chitin oligosaccharide deacetylase from *Shewanella baltica* ATCC BAA-1091. *Biosci. Biotechnol. Biochem.* **2017**, *81*, 547–550. [CrossRef] [PubMed]
76. Tuveng, T.R.; Rothweiler, U.; Udatha, G.; Vaaje-Kolstad, G.; Smalås, A.; Eijsink, V.G.H. Structure and function of a CE4 deacetylase isolated from a marine environment. *PLoS ONE* **2017**, *12*, e0187544. [CrossRef] [PubMed]
77. Bartnicki-Garcia, S.; Nickerson, W.J. Isolation, composition, and structure of cell walls of filamentous and yeast-like forms of *Mucor rouxii*. *Biochim. Biophys. Acta* **1962**, *58*, 102–119. [CrossRef]
78. Kafetzopoulos, D.; Martinou, A.; Bouriotis, V. Bioconversion of chitin to chitosan: Purification and characterization of chitin deacetylase from *Mucor rouxii*. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 2564–2568. [CrossRef] [PubMed]
79. Davis, L.L.; Bartnicki-Garcia, S. The co-ordination of chitosan and chitin synthesis in *Mucor rouxii*. *J. Gen. Microbiol.* **1984**, *130*, 2095–2102. [CrossRef] [PubMed]
80. Chatterjee, S.; Adhya, M.; Guha, A.K.; Chatterjee, B.P. Chitosan from *Mucor rouxii*: Production and physico-chemical characterization. *Process Biochem.* **2005**, *40*, 395–400. [CrossRef]
81. Synowiecki, J.; Al-Khateeb, N.A.A.Q. Mycelia of *Mucor rouxii* as a source of chitin and chitosan. *Food Chem.* **1997**, *60*, 605–610. [CrossRef]
82. Martinou, A.; Kafetzopoulos, D.; Bouriotis, V. Isolation of chitin deacetylase from *Mucor rouxii* by immunoaffinity chromatography. *J. Chromatogr. A* **1993**, *644*, 35–41. [CrossRef]
83. Kauss, H.; Bauch, B. Chitin deacetylase from *Colletotrichum lindemuthianum*. *Methods Enzymol.* **1988**, *161*, 518–523. [CrossRef]
84. Tsigos, I.; Bouriotis, V. Purification and characterization of chitin deacetylase from *Colletotrichum lindemuthianum*. *J. Biol. Chem.* **1995**, *270*, 26286–26291. [CrossRef] [PubMed]

85. O'Connell, R.J.; Ride, J.P. Chemical detection and ultrastructural localization of chitin in cell walls of *Colletotrichum lindemuthianum*. *Physiol. Mol. Plant Pathol.* **1990**, *37*, 39–53. [CrossRef]
86. Tokuyasu, K.; Ohnishi-Kameyama, M.; Hayashi, K. Purification and characterization of extracellular chitin deacetylase from *Colletotrichum lindemuthianum*. *Biosci. Biotechnol. Biochem.* **1996**, *60*, 1598–1603. [CrossRef] [PubMed]
87. Shrestha, B.; Blondeau, K.; Stevens, W.F.; Hegarat, F.L. Expression of chitin deacetylase from *Colletotrichum lindemuthianum* in *Pichia pastoris*: Purification and characterization. *Protein Expr. Purif.* **2004**, *38*, 196–204. [CrossRef] [PubMed]
88. Kang, L.; Chen, X.; Zhai, C.; Ma, L. Synthesis and high expression of chitin deacetylase from *Colletotrichum lindemuthianum* in *Pichia pastoris* GS115. *J. Microbiol. Biotechnol.* **2012**, *22*, 1202–1207. [CrossRef] [PubMed]
89. Tokuyasu, K.; Kaneko, S.; Hayashi, K.; Mori, Y. Production of a recombinant chitin deacetylase in the culture medium of *Escherichia coli* cells. *FEBS Lett.* **1999**, *458*, 23–26. [CrossRef]
90. Tokuyasu, K.; Ohnishi-Kameyama, M.; Hayashi, K.; Mori, Y. Cloning and expression of chitin deacetylase gene from a deuteromycete, *Colletotrichum lindemuthianum*. *J. Biosci. Bioeng.* **1999**, *87*, 418–423. [CrossRef]
91. Tokuyasu, K.; Ono, H.; Ohnishi-Kameyama, M.; Hayashi, K.; Mori, Y. Deacetylation of chitin oligosaccharides of dp 2-4 by chitin deacetylase from *Colletotrichum lindemuthianum*. *Carbohydr. Res.* **1997**, *303*, 353–358. [CrossRef]
92. Tokuyasu, K.; Ono, H.; Hayashi, K.; Mori, Y. Reverse hydrolysis reaction of chitin deacetylase and enzymatic synthesis of β-D-GlcNAc-(1→4)-GlcN from chitobiose. *Carbohydr. Res.* **1999**, *322*, 26–31. [CrossRef]
93. Tokuyasu, K.; Ono, H.; Mitsutomi, M.; Hayashi, K.; Mori, Y. Synthesis of a chitosan tetramer derivative, β-D-GlcNAc-(1→4)-β-D-GlcNAc-(1→4)-β-D-GlcNAc-(1→4)-D-GlcN through a partial *N*-acetylation reaction by chitin deacetylase. *Carbohydr. Res.* **2000**, *325*, 211–215. [CrossRef]
94. Kang, L.X.; Liang, Y.X.; Ma, L.X. Novel characteristics of chitin deacetylase from *Colletotrichum lindemuthianum*: Production of fully acetylated chitooligomers, and hydrolysis of deacetylated chitooligomers. *Process Biochem.* **2014**, *49*, 1936–1940. [CrossRef]
95. Reyes, F.; Calatayud, J.; Vazquez, C.; Martínez, M.J. β-*N*-Acetylglucosaminidase from *Aspergillus nidulans* which degrades chitin oligomers during autolysis. *FEMS Microbiol. Lett.* **1990**, *65*, 83–87. [CrossRef]
96. Wang, Y.; Song, J.Z.; Yang, Q.; Liu, Z.H.; Huang, X.M.; Chen, Y. Cloning of a heat-stable chitin deacetylase gene from *Aspergillus nidulans* and its functional expression in *Escherichia coli*. *Appl. Biochem. Biotechnol.* **2010**, *162*, 843–854. [CrossRef] [PubMed]
97. Espagne, E.; Lespinet, O.; Malagnac, F.; Da Silva, C.; Jaillon, O.; Porcel, B.M.; Couloux, A.; Aury, J.-M.; Ségurens, B.; Poulain, J.; Anthouard, V.; et al. The genome sequence of the model ascomycete fungus *Podospora anserina*. *Genome Biol.* **2008**, *9*, R77. [CrossRef] [PubMed]
98. Lorin, S.; Dufour, E.; Sainsard-Chanet, A. Mitochondrial metabolism and aging in the filamentous fungus *Podospora anserina*. *Biochim. Biophys. Acta Bioenerg.* **2006**, *1757*, 604–610. [CrossRef] [PubMed]
99. Figueroa, M.; Upadhyaya, N.M.; Sperschneider, J.; Park, R.F.; Szabo, L.J.; Steffenson, B.; Ellis, J.G.; Dodds, P.N. Changing the Game: Using Integrative Genomics to Probe Virulence Mechanisms of the Stem Rust Pathogen *Puccinia graminis* f. sp. tritici. *Front. Plant Sci.* **2016**, *7*, 1–10. [CrossRef] [PubMed]
100. Singh, R.P.; Hodson, D.P.; Jin, Y.; Lagudah, E.S.; Ayliffe, M.A.; Bhavani, S.; Rouse, M.N.; Pretorius, Z.A.; Szabo, L.J.; Huerta-Espino, J.; et al. Emergence and Spread of New Races of Wheat Stem Rust Fungus: Continued Threat to Food Security and Prospects of Genetic Control. *Phytopathology* **2015**, *105*, 872–884. [CrossRef] [PubMed]
101. Jin, Y.; Singh, R.P. Resistance in U.S. Wheat to Recent Eastern African Isolates of *Puccinia graminis* f. sp. tritici with Virulence to Resistance Gene Sr31. *Plant Dis.* **2006**, *90*, 476–480. [CrossRef]
102. Mendgen, K.; Hahn, M. Plant infection and the establishment of fungal biotrophy. *Trends Plant Sci.* **2002**, *7*, 352–356. [CrossRef]
103. Broeker, K.; Fehser, S.; Tenberge, K.B.; Moerschbacher, B.M. Two class III chitin synthases specifically localized in appressoria and haustoria of *Puccinia graminis* f. sp. tritici. *Physiol. Mol. Plant Pathol.* **2011**, *76*, 27–33. [CrossRef]
104. Ride, J.P.; Barber, M.S. Purification and characterization of multiple forms of endochitinase from wheat leaves. *Plant Sci.* **1990**, *71*, 185–197. [CrossRef]

105. Vander, P.; Vårum, K.M.; Domard, A.; El Gueddari, N.E.; Moerschbacher, B.M. Comparison of the Ability of Partially *N*-Acetylated Chitosans and Chitooligosaccharides to Elicit Resistance Reactions in Wheat Leaves. *Plant Physiol.* **1998**, *118*, 1353–1359. [CrossRef] [PubMed]
106. Maharachchikumbura, S.S.N.; Hyde, K.D.; Groenewald, J.Z.; Xu, J.; Crous, P.W. *Pestalotiopsis* revisited. *Stud. Mycol.* **2014**, *79*, 121–186. [CrossRef] [PubMed]
107. Lopez-Llorca, L.V.; Olivares-Bernabeu, C.; Salinas, J.; Jansson, H.-B.; Kolattukudy, P.E. Pre-penetration events in fungal parasitism of nematode eggs. *Mycol. Res.* **2002**, *106*, 499–506. [CrossRef]
108. Manzanilla-Lopez, R.H.; Esteves, I.; Finetti-Sialer, M.M.; Hirsch, P.R.; Ward, E.; Devonshire, J.; Hidalgo-Diaz, L. *Pochonia chlamydosporia*: Advances and Challenges to Improve Its Performance as a Biological Control Agent of Sedentary Endo-parasitic Nematodes. *J. Nematol.* **2013**, *45*, 1–7. [PubMed]
109. Larriba, E.; Jaime, M.D.L.A.; Carbonell-Caballero, J.; Conesa, A.; Dopazo, J.; Nislow, C.; Martín-Nieto, J.; Lopez-Llorca, L.V. Sequencing and functional analysis of the genome of a nematode egg-parasitic fungus, *Pochonia chlamydosporia*. *Fungal Genet. Biol.* **2014**, *65*, 69–80. [CrossRef] [PubMed]
110. Aranda-Martinez, A.; Lenfant, N.; Escudero, N.; Zavala-Gonzalez, E.A.; Henrissat, B.; Lopez-Llorca, L.V. CAZyme content of *Pochonia chlamydosporia* reflects that chitin and chitosan modification are involved in nematode parasitism. *Environ. Microbiol.* **2016**, *18*, 4200–4215. [CrossRef] [PubMed]
111. Wani, Z.A.; Kumar, A.; Sultan, P.; Bindu, K.; Riyaz-Ul-Hassan, S.; Ashraf, N. *Mortierella* alpina CS10E4, an oleaginous fungal endophyte of *Crocus sativus* L. enhances apocarotenoid biosynthesis and stress tolerance in the host plant. *Sci. Rep.* **2017**, *7*, 8598. [CrossRef] [PubMed]
112. Zhao, Y.; Jo, G.-H.; Ju, W.-T.; Jung, W.-J.; Park, R.-D. A Highly *N*-Glycosylated Chitin Deacetylase Derived from a Novel Strain of *Mortierella* sp. DY-52. *Biosci. Biotechnol. Biochem.* **2011**, *75*, 960–965. [CrossRef] [PubMed]
113. Narayanan, K.; Parameswaran, B.; Pandey, A. Production of chitin deacetylase by *Aspergillus flavus* in submerged conditions. *Prep. Biochem. Biotechnol.* **2016**, *46*, 501–508. [CrossRef] [PubMed]
114. Cuenca-Estrella, M.; Gomez-Lopez, A.; Mellado, E.; Buitrago, M.J.; Monzón, A.; Rodriguez-Tudela, J.L. *Scopulariopsis brevicaulis*, a fungal pathogen resistant to broad-spectrum antifungal agents. *Antimicrob. Agents Chemother.* **2003**, *47*, 2339–2341. [CrossRef] [PubMed]
115. Tosti, A.; PiracciniI, B.M.; Stinchi, C.; Lorenzi, S. Onychomycosis due to *Scopulariopsis brevicaulis*: Clinical features and response to systemic antifungals. *Br. J. Dermatol.* **1996**, *135*, 799–802. [CrossRef] [PubMed]
116. El Ghaouth, A.; Arul, J.; Grenier, J.; Asselin, A. Effect of chitosan and other polyions on chitin deacetylase in *Rhizopus stolonifer*. *Exp. Mycol.* **1992**, *16*, 173–177. [CrossRef]
117. Zhang, H.; Yang, S.; Fang, J.; Deng, Y.; Wang, D.; Zhao, Y. Optimization of the fermentation conditions of *Rhizopus japonicus* M193 for the production of chitin deacetylase and chitosan. *Carbohydr. Polym.* **2014**, *101*, 57–67. [CrossRef] [PubMed]
118. Maw, T.; Tan, T.K.; Khor, E.; Wong, S.M. Selection of *Gongronella butleri* strains for enhanced chitosan yield with UV mutagenesis. *J. Biotechnol.* **2002**, *95*, 189–193. [CrossRef]
119. Yonemura, A.; Nagashima, T.; Murayama, T. *Expression of Chitin Deacetylase Gene from Phycomyces blakesleeanus in Aspergillus oryzae and Neurospora crassa*; The Society for Bioscience and Bioengineering: Osaka, Japan, 2007; p. 129. Available online: http://dl.ndl.go.jp/view/download/digidepo_10529404_po_ART0009175183.pdf?contentNo=1&alternativeNo= (accessed on 19 February 2018).
120. Baker, L.G.; Specht, C.A.; Lodge, J.K. Cell wall chitosan is necessary for virulence in the opportunistic pathogen *Cryptococcus neoformans*. *Eukaryot. Cell* **2011**, *10*, 1264–1268. [CrossRef] [PubMed]
121. Doering, T.L. How Sweet it is! Cell Wall Biogenesis and Polysaccharide Capsule Formation in *Cryptococcus neoformans*. *Annu. Rev. Microbiol.* **2009**, *63*, 223–247. [CrossRef] [PubMed]
122. Gilbert, N.M.; Baker, L.G.; Specht, C.A.; Lodge, J.K. A glycosylphosphatidylinositol anchor is required for membrane localization but dispensable for cell wall association of chitin deacetylase 2 in *Cryptococcus neoformans*. *mBio* **2012**, *3*, e00007–e00012. [CrossRef] [PubMed]
123. Levitz, S.M.; Nong, S.-H.; Mansour, M.K.; Huang, C.; Specht, C.A. Molecular characterization of a mannoprotein with homology to chitin deacetylases that stimulates T cell responses to *Cryptococcus neoformans*. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10422–10427. [CrossRef] [PubMed]
124. Biondo, C.; Beninati, C.; Delfino, D.; Oggioni, M.; Mancuso, G.; Midiri, A.; Tomaselli, G.; Teti, G.; Biondo, C.; Beninati, C.; et al. Identification and Cloning of a Cryptococcal Deacetylase That Produces Protective Immune Responses. *Infect. Immun.* **2002**, *70*, 2383–2391. [CrossRef] [PubMed]

125. Loftus, B.; Anderson, I.; Davies, R.; Alsmark, U.C.M.; Samuelson, J.; Amedeo, P.; Roncaglia, P.; Berriman, M.; Hirt, R.P.; Mann, B.J.; et al. The genome of the protist parasite *Entamoeba histolytica*. *Nature* **2005**, *433*, 865–868. [CrossRef] [PubMed]
126. Roche, P.; Maillet, F.; Plazanet, C.; Debellé, F.; Ferro, M.; Truchet, G.; Promé, J.C.; Dénarié, J. The common nodABC genes of *Rhizobium meliloti* are host-range determinants. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 15305–15310. [CrossRef] [PubMed]
127. Egelhoff, T.T.; Long, S.R. *Rhizobium meliloti* nodulation genes: Identification of nodDABC gene products, purification of nodA protein, and expression of nodA in *Rhizobium meliloti*. *J. Bacteriol.* **1985**, *164*, 591–599. [PubMed]
128. Spaink, H.P.; Wijfjes, A.H.M.; der van Drift, K.M.G.M.; Haverkamp, J.; Thomas-Oates, J.E.; Lugtenberg, B.J.J. Structural identification of metabolites produced by the NodB and NodC proteins of *Rhizobium* leguminosarum. *Mol. Microbiol.* **1994**, *13*, 821–831. [CrossRef] [PubMed]
129. Chambon, R.; Pradeau, S.; Fort, S.; Cottaz, S.; Armand, S. High yield production of *Rhizobium* NodB chitin deacetylase and its use for in vitro synthesis of lipo-chitinoligosaccharide precursors. *Carbohydr. Res.* **2017**, *442*, 25–30. [CrossRef] [PubMed]
130. Hamer, S.N.; Cord-Landwehr, S.; Biarnés, X.; Planas, A.; Waegeman, H.; Moerschbacher, B.M.; Kolkenbrock, S. Enzymatic production of defined chitosan oligomers with a specific pattern of acetylation using a combination of chitin oligosaccharide deacetylases. *Sci. Rep.* **2015**, *5*, 8716. [CrossRef] [PubMed]
131. Röhrig, H.; Schmidt, J.; Wieneke, U.; Kondorosi, E.; Barlier, I.; Schell, J.; John, M. Biosynthesis of lipooligosaccharide nodulation factors: *Rhizobium* NodA protein is involved in *N*-acylation of the chitooligosaccharide backbone. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 3122–3126. [CrossRef] [PubMed]
132. Keyhani, N.O.; Roseman, S. Physiological aspects of chitin catabolism in marine bacteria. *Biochim. Biophys. Acta Gen. Subj.* **1999**, *1473*, 108–122. [CrossRef]
133. Zobell, C.; Rittenberg, S. The occurrence and characteristics of chitinoclastic bacteria in the sea. *J. Bacteriol.* **1937**, *35*, 275–287.
134. Meibom, K.L.; Li, X.B.; Nielsen, A.T.; Wu, C.-Y.; Roseman, S.; Schoolnik, G.K. The *Vibrio cholerae* chitin utilization program. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 2524–2529. [CrossRef] [PubMed]
135. Li, X.; Roseman, S. The chitinolytic cascade in Vibrios is regulated by chitin oligosaccharides and a two-component chitin catabolic sensor/kinase. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 627–631. [CrossRef] [PubMed]
136. Ohishi, K.; Murase, K.; Ohta, T.; Etoh, H. Cloning and sequencing of the deacetylase gene from *Vibrio alginolyticus* H-8. *J. Biosci. Bioeng.* **2000**, *90*, 561–563. [CrossRef]
137. Kadokura, K.; Sakamoto, Y.; Saito, K.; Ikegami, T.; Hirano, T.; Hakamata, W.; Oku, T.; Nishio, T. Production of a recombinant chitin oligosaccharide deacetylase from *Vibrio parahaemolyticus* in the culture medium of *Escherichia coli* cells. *Biotechnol. Lett.* **2007**, *29*, 1209–1215. [CrossRef] [PubMed]
138. Hirano, T.; Maebara, Y.; Uehara, R.; Sakaki, Y.; Shiraishi, H.; Ichimura, S.; Hakamata, W.; Nishio, T. Chitin oligosaccharide deacetylase from Vibrio harveyi ATCC BAA-1116: Gene cloning, overexpression, purification, and characterization. *Chitin Chitosan Res.* **2012**, *19*, 321–324.
139. Jacquiod, S.; Franqueville, L.; Cécillon, S.; Vogel, T.M.; Simonet, P. Soil bacterial community shifts after Chitin enrichment: An integrative metagenomic approach. *PLoS ONE* **2013**, *8*. [CrossRef] [PubMed]
140. Dsouza, M.; Taylor, M.W.; Turner, S.J.; Aislabie, J. Genomic and phenotypic insights into the ecology of *Arthrobacter* from Antarctic soils. *BMC Genom.* **2015**, *16*, 36. [CrossRef] [PubMed]
141. Lonhienne, T.; Mavromatis, K.; Vorgias, C.E.; Buchon, L.; Gerday, C.; Bouriotis, V. Cloning, sequences, and characterization of two chitinase genes from the Antarctic *Arthrobacter* sp. strain TAD20: Isolation and partial characterization of the enzymes. *J. Bacteriol.* **2001**, *183*, 1773–1779. [CrossRef] [PubMed]
142. Aragunde-pazos, H.; Biarnés, X.; Planas, A. Substrate recognition and specificity of chitin deacetylases and related family 4 carbohydrate esterases. *Int. J. Mol. Sci.* **2018**, *19*, 412. [CrossRef] [PubMed]
143. Boraston, A.B.; Bolam, D.N.; Gilbert, H.J.; Davies, G.J. Carbohydrate-Binding modules: Fine-Tuning polysaccharide recognition. *Biochem. J.* **2004**, *382*, 769–781. [CrossRef] [PubMed]
144. Blair, D.E.; Schuttelkopf, A.W.; MacRae, J.I.; van Aalten, D.M.F. Structure and metal-dependent mechanism of peptidoglycan deacetylase, a streptococcal virulence factor. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15429–15434. [CrossRef] [PubMed]

145. Blair, D.E.; Van Aalten, D.M.F. Structures of *Bacillus subtilis* PdaA, a family 4 carbohydrate esterase, and a complex with *N*-acetyl-glucosamine. *FEBS Lett.* **2004**, *570*, 13–19. [CrossRef] [PubMed]
146. Nakamura, A.M.; Nascimento, A.S.; Polikarpov, I. Structural diversity of carbohydrate esterases. *Biotechnol. Res. Innov.* **2017**, *1*, 35–51. [CrossRef]
147. Nishiyama, T.; Noguchi, H.; Yoshida, H.; Park, S.Y.; Tame, J.R.H. The structure of the deacetylase domain of *Escherichia coli* PgaB, an enzyme required for biofilm formation: A circularly permuted member of the carbohydrate esterase 4 family. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2013**, *69*, 44–51. [CrossRef] [PubMed]
148. Hernick, M.; Fierke, C.A. Zinc hydrolases: The mechanisms of zinc-dependent deacetylases. *Arch. Biochem. Biophys.* **2005**, *433*, 71–84. [CrossRef] [PubMed]
149. Xie, J.L.; Polvi, E.J.; Shekhar-Guturja, T.; Cowen, L.E. Elucidating drug resistance in human fungal pathogens. *Future Microbiol.* **2014**, *9*, 523–542. [CrossRef] [PubMed]
150. Takaya, N.; Yamazaki, D.; Horiuchi, H.; Ohta, A.; Takagi, M. Cloning and characterization of a chitinase-encoding gene (chiA) from *Aspergillus nidulans*, disruption of which decreases germination frequency and hyphal growth. *Biosci. Biotechnol. Biochem.* **1998**, *62*, 60–65. [CrossRef] [PubMed]
151. Hartl, L.; Zach, S.; Seidl-Seiboth, V. Fungal chitinases: Diversity, mechanistic properties and biotechnological potential. *Appl. Microbiol. Biotechnol.* **2012**, *93*, 533–543. [CrossRef] [PubMed]
152. Aoun, M. Host defense mechanisms during fungal pathogenesis and how these are overcome in susceptible plants: A review. *Int. J. Bot.* **2017**, *13*, 82–102. [CrossRef]
153. Huang, G. Chitinase Inhibitor Allosamidin and Its Analogues: An Update. *Curr. Org. Chem.* **2012**, *16*, 115–120. [CrossRef]
154. Rao, F.V.; Houston, D.R.; Boot, R.G.; Aerts, J.M.F.G.; Hodkinson, M.; Adams, D.J.; Shiomi, K.; Omura, S.; Van Aalten, D.M.F. Specificity and affinity of natural product cyclopentapeptide inhibitors against *A. fumigatus*, human, and bacterial chitinases. *Chem. Biol.* **2005**, *12*, 65–76. [CrossRef] [PubMed]
155. Younes, I.; Rinaudo, M. Chitin and chitosan preparation from marine sources. Structure, properties and applications. *Mar. Drugs* **2015**, *13*, 1133–1174. [CrossRef] [PubMed]
156. Cheung, R.C.F.; Ng, T.B.; Wong, J.H.; Chan, W.Y. Chitosan: An update on potential biomedical and pharmaceutical applications. *Mar. Drugs* **2015**, *13*, 5156–5186. [CrossRef] [PubMed]
157. Anitha, A.; Sowmya, S.; Kumar, P.T.T.S.; Deepthi, S.; Chennazhi, K.P.; Ehrlich, H.; Tsurkan, M.; Jayakumar, R. Chitin and chitosan in selected biomedical applications. *Prog. Polym. Sci.* **2014**, *39*, 1644–1667. [CrossRef]
158. Abdul Khalil, H.P.S.; Saurabh, C.K.; Adnan, A.S.; Nurul Fazita, M.R.; Syakir, M.I.; Davoudpour, Y.; Rafatullah, M.; Abdullah, C.K.; Haafiz, M.K.M.; Dungani, R. A review on chitosan-cellulose blends and nanocellulose reinforced chitosan biocomposites: Properties and their applications. *Carbohydr. Polym.* **2016**, *150*, 216–226.
159. Pestov, A.; Bratskaya, S. Chitosan and Its Derivatives as Highly Efficient Polymer Ligands. *Molecules* **2016**, *21*, 330. [CrossRef] [PubMed]
160. Das, S.N.; Madhuprakash, J.; Sarma, P.V.S.R.N.; Purushotham, P.; Suma, K.; Manjeet, K.; Rambabu, S.; Gueddari, N.E.E.; Moerschbacher, B.M.; Podile, A.R. Biotechnological approaches for field applications of chitooligosaccharides (COS) to induce innate immunity in plants. *Crit. Rev. Biotechnol.* **2015**, *35*, 29–43. [CrossRef] [PubMed]
161. Sorlier, P.; Denuzière, A.; Viton, C.; Domard, A. Relation between the degree of acetylation and the electrostatic properties of chitin and chitosan. *Biomacromolecules* **2001**, *2*, 765–772. [CrossRef] [PubMed]
162. Omura, Y.; Shigemoto, M.; Akiyama, T.; Saimoto, H.; Shigemasa, Y.; Nakamura, I.; Tsuchido, T. Antimicrobial Activity of Chitosan with Different Degrees of Acetylation and Molecular Weights. *Biocontrol Sci.* **2003**, *8*, 25–30. [CrossRef]
163. Domard, A.; Cartier, N. Glucosamine oligomers: 1. Preparation and characterization. *Int. J. Biol. Macromol.* **1989**, *11*, 297–302. [CrossRef]
164. Einbu, A.; Varum, K.M. Depolymerization and de-*N*-acetylation of chitin oligomers in hydrochloric acid. *Biomacromolecules* **2007**, *8*, 309–314. [CrossRef] [PubMed]
165. Kuyama, H.; Nakahara, Y.; Nukada, T.; Ito, Y.; Nakahara, Y.; Ogawa, T. Stereocontrolled synthesis of chitosan dodecamer. *Carbohydr. Res.* **1993**, *243*, C1–C7. [CrossRef]
166. Barroca-Aubry, N.; Pernet-Poil-Chevrier, A.; Domard, A.; Trombotto, S. Towards a modular synthesis of well-defined chitooligosaccharides: Synthesis of the four chitodisaccharides. *Carbohydr. Res.* **2010**, *345*, 1685–1697. [CrossRef] [PubMed]

167. Weinhold, M.X.; Sauvageau, J.C.M.; Kumirska, J.; Thöming, J. Studies on acetylation patterns of different chitosan preparations. *Carbohydr. Polym.* **2009**, *78*, 678–684. [CrossRef]
168. Abla, M.; Marmuse, L.; Delolme, F.; Vors, J.P.; Ladavière, C.; Trombotto, S. Access to tetra-*N*-acetyl-chitopentaose by chemical *N*-acetylation of glucosamine pentamer. *Carbohydr. Polym.* **2013**, *98*, 770–777. [CrossRef] [PubMed]
169. Trombotto, S.; Ladavière, C.; Delolme, F.; Domard, A. Chemical preparation and structural characterization of a homogeneous series of chitin/chitosan oligomers. *Biomacromolecules* **2008**, *9*, 1731–1738. [CrossRef] [PubMed]
170. Naqvi, S.; Moerschbacher, B.M. The cell factory approach toward biotechnological production of high-value chitosan oligomers and their derivatives: An update. *Crit. Rev. Biotechnol.* **2017**, *37*, 11–25. [CrossRef] [PubMed]
171. Hembach, L.; Cord-Landwehr, S.; Moerschbacher, B.M. Enzymatic production of all fourteen partially acetylated chitosan tetramers using different chitin deacetylases acting in forward or reverse mode. *Sci. Rep.* **2017**, *7*, 17692. [CrossRef] [PubMed]
172. Mergaert, P.; D'Haeze, W.; Geelen, D.; Promé, D.; Van Montagu, M.; Geremia, R.; Promé, J.C.; Holsters, M. Biosynthesis of *Azorhizobium caulinodans* Nod factors: Study of the activity of the nodABCS proteins by expression of the genes in *Escherichia coli*. *J. Biol. Chem.* **1995**, *270*, 29217–29223. [CrossRef] [PubMed]
173. Poinsot, V.; Crook, M.B.; Erdn, S.; Maillet, F.; Bascaules, A.; Ané, J.M. New insights into Nod factor biosynthesis: Analyses of chitooligomers and lipo-chitooligomers of *Rhizobium* sp. IRBG74 mutants. *Carbohydr. Res.* **2016**, *434*, 83–93. [CrossRef] [PubMed]
174. Samain, E.; Drouillard, S.; Heyraud, A.; Driguez, H.; Geremia, R.A. Gram-scale synthesis of recombiant chitooligosaccharides in *Escherichia coli*. *Carbohydr. Res.* **1997**, *302*, 35–42. [CrossRef]
175. Samain, E.; Chazalet, V.; Geremia, R.A. Production of O-acetylated and sulfated chitooligosaccharides by recombinant *Escherichia coli* strains harboring different combinations of nod genes. *J. Biotechnol.* **1999**, *72*, 33–47. [CrossRef]
176. Cottaz, S.; Samain, E. Genetic engineering of *Escherichia coli* for the production of N^{I}, N^{II}-diacetylchitobiose (chitinbiose) and its utilization as a primer for the synthesis of complex carbohydrates. *Metab. Eng.* **2005**, *7*, 311–317. [CrossRef] [PubMed]
177. Bettler, E.; Samain, E.; Chazalet, V.; Bosso, C.; Heyraud, A.; Joziasse, D.H.; Wakarchuk, W.W.; Imberty, A.; Geremia, R.A. The living factory: In Vivo production of *N*-acetyllactosamine containing carbohydrates in *E. coli*. *Glycoconj. J.* **1999**, *16*, 205–212. [CrossRef] [PubMed]
178. Southwick, A.M.; Wang, L.X.; Long, S.R.; Lee, Y.C. Activity of *Sinorhizobium meliloti* NodAB and nodH enzymes on thiochitooligosaccharides. *J. Bacteriol.* **2002**, *184*, 4039–4043. [CrossRef] [PubMed]

analytical chemistry

Letter

Cite This: *Anal. Chem.* 2018, 90, 10654−10658

pubs.acs.org/ac

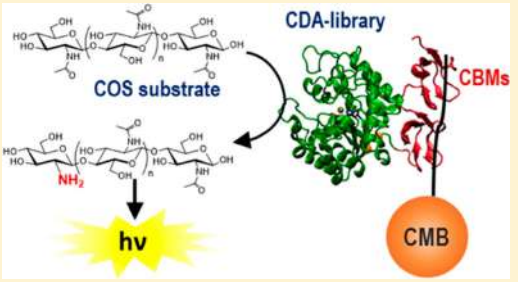
Screening Assay for Directed Evolution of Chitin Deacetylases: Application to *Vibrio cholerae* Deacetylase Mutant Libraries for Engineered Specificity

Sergi Pascual and Antoni Planas*

Laboratory of Biochemistry, Institut Químic de Sarrià, University Ramon Llull, 08017 Barcelona, Spain

Supporting Information

ABSTRACT: Not only the degree of acetylation but also the pattern of acetylation of chitosans and chitooligosaccharides (COS) appear to be critical for their biological activities. Protein engineering may expand the toolbox of chitin deacetylases (CDAs) with defined specificities for the enzymatic production of partially deacetylated COS for biotech and biomedical applications. A high-throughput screening (HTS) assay for screening directed evolution libraries is reported. It is based on a fluorescence monitoring assay of the deacetylase activity on COS substrates after capturing the expressed enzyme variants fused to a chitin binding module with chitin-coated magnetic beads. The assay is applied to the screening of random libraries of a *Vibrio cholera* CDA for increased activity on longer COS substrates.



Deacetylation of chitin and chitooligosaccharides (COS) renders chitosans (with variable degrees of acetylation (DA)) and partially deacetylated oligosaccharides (paCOS) with considerable industrial interest as biocompatible, biodegradable, and nontoxic functional materials with a wide variety of biotechnological and biomedical applications.[1−5] Their polycationic nature due to partial protonation of the free amino groups at physiological pH values enable them to interact with polyanionic biomolecules, such as membrane phospholipids, DNA, and proteins. The DA strongly influences the physicochemical and biological properties of chitosans and paCOS, but the role of the pattern of acetylation (PA), which defines the distribution of charged D-glucosamine ($GlcNH_2$) residues along the N-acetyl-D-glucosamine (GlcNAc) polymeric/oligomeric chain, on the specific interactions with biological receptors and target structures remains largely unknown.[5,6] Because current chemical deacetylation methods essentially yield random patterns of deacetylation, there is a growing interest in the use of enzymes, chitinases, and chitin deacetylases, as biocatalysts to produce chitosans and paCOS with defined structures to evaluate their biological functions and develop new applications.[6−11]

Most chitin deacetylases (CDAs, EC 3.5.1.41) and chitin oligosaccharide deacetylases (CODs, EC 3.5.1.105) are classified in the carbohydrate esterase family 4 (CE4) of the CAZY database (www.cazy.org). They hydrolyze the N-acetamido group of β-1,4-linked GlcNAc residues of chitin and COS by a metal-assisted general acid−base mechanism and exhibit diverse substrate specificities leading to fully or partially deacetylated products.[12−14] CDAs acting on polymeric chitin show different modes of action leading to block copolymers (with stretches of $GlcNH_2$ units within the GlcNAc linear chain, named multiple-attack mechanism with processive deacetylation behavior) or to randomly deacetylated polymers (named multiple-chain mechanism, with random distribution of $GlcNH_2$ units in the chitosan products). When acting on COS (with degrees of polymerization (DP) from 2 to 8), CDAs and CODs show diverse deacetylation patterns, some fully deacetylating their COS substrates, but few recently characterized CDAs show specific deacetylation patterns (see Table S1;[15] for a recent review). Ongoing research programs search for new CDAs from fungal and bacterial origin with the aim of discovering novel specificities and expanding the toolbox of CDAs and CODs for the enzymatic production of tailored paCOS and chitosans for biotech and biomedical application. The structural determinants of CDA specificities are poorly understood, with some new insights coming from the recently solved X-ray structures of enzyme/substrate complexes (*Vc*CDA[14] and *Ar*CE4[16]) and the proposed "subsite capping model"[14,17] to rationalize specificity and deacetylation patterns. Protein engineering will also offer the opportunity to engineer substrate specificity, by either rational or directed evolution approaches. In this context, an efficient assay for deacetylase activity implemented in high-throughput screening (HTS) format is needed to screen either natural CDAs or

Received: June 17, 2018
Accepted: August 23, 2018
Published: August 23, 2018

DOI: 10.1021/acs.analchem.8b02729
Anal. Chem. 2018, 90, 10654−10658

directed evolution libraries in the search of novel specificities and improved efficiencies.

Here, we report the design and development of an HTS assay for the screening of directed evolution libraries of CDA and other CE4 enzymes active on COS and its proof-of-concept application to the engineering of substrate specificity of a *Vibrio cholerae* chitin deacetylase (*Vc*CDA).

Enzyme assays for deacetylation activity of CDAs and CODs are diverse. Whereas unspecific substrates for general amidase activity (4-methylumbelliferyl acetate,[18] 4-nitrophenyl acetate, or *p*-nitroacetanilide[19]) are often used to monitor protein purifications, assays using COS as substrates are specific for kinetic characterizations. These are based on monitoring acetate release (UV absorbance changes,[20] radiolabeled substrates,[21] and coupled enzymatic assays[22]) or formation of free amino groups with chromogenic or fluorogenic reagents such as fluorescamine,[12] *o*-phthalaldehyde,[23] or ninhydrin.[24] All of these methods use purified enzymes and are not directly applicable to an HTS format with cell extracts as required for screening in directed evolution approaches. More complex HPLC-MS methods have been developed to monitor deacetylation and sequence the acetylation pattern of the products.[25]

The HTS assay, implemented in 96-well microtiter plates, is based on fusion of the target CE4 catalytic domain to be evolved to a chitin binding module (CBM), capture of the expressed proteins from cell-free extracts with chitin-coated magnetic beads, and evaluation of the deacetylase activity of the immobilized enzyme variants on natural COS substrates by monitoring product formation with a coupled assay leading to a fluorescence readout (Figure 1).

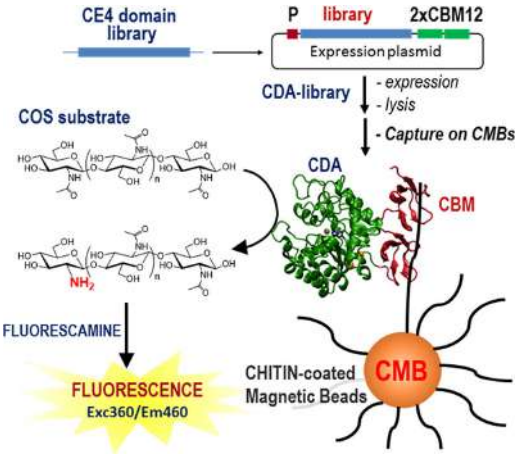


Figure 1. HTS assay principle for deacetylase activity on COS.

For the HTS assay development, different parameters were first optimized using the *Vibrio cholera* chitin deacetylase (wt *Vc*CDA),[14,26] either as purified enzyme or as cell-free extracts of recombinant *E. coli* expressing the enzyme. *Vc*CDA is a multidomain protein composed of a CE4 catalytic domain with chitooligosaccharide deacetylase activity and two family 12 chitin binding modules (CBM12) at the C-terminus, which have been shown to bind insoluble chitin.[27] This tandem of two CBM12 domains will be used in the general HTS protocol; the target CE4 catalytic domain (library of mutants to be screened) will be fused to the CBMs acting as a C-terminus tag for binding to the chitin magnetic beads (CMB). The buffer that was selected for conditioning and binding to the CMB did not contain detergents and EDTA (as commonly used with commercial CMB).[28] Although some CDAs are not inhibited by EDTA, indicating strong metal cation coordination,[13,14] it cannot be a general assumption when applying the screening to directed evolution mutant libraries. Therefore, we chose a simple PBS buffer where the efficiency of binding was about 50−60% in the range of concentrations tested (Figure S1). Binding was shown to be specific when cell extracts were used, not detecting any other bound proteins (Figure S2). BSA, if required as stabilizer, had some affinity for the beads, but it did not displace the bound CDA−CBM fusion protein (Figure S2). Kinetics of binding determined that maximum binding was achieved in less than 15 min at 4 °C with shaking (Figure S3). Immobilization of *Vc*CDA on the CMBs resulted in 40% reduction of its activity relative to the free enzyme in solution in the same reaction buffer, as determined by monitoring product formation from a diacetylchitobiose substrate by HPLC-MS (Figure S4). *E. coli* BL21 (DE3) was used as host for *Vc*CDA expression. Since *E. coli* expresses an endogenous chitin deacetylase, a ΔChib strain (knockout of the *chib* operon)[29] was also tested as expression host to analyze protein binding and specific activity. The same binding behavior to the CMBs as well as the same specific activity were observed with cell-free extract from both strains (Figure S5). Therefore, standard BL21 (DE3) cells were further used to express the CDA libraries as they showed a slightly higher protein expression level. Cell lysis of the cultures grown on 2 mL deep-well plates was optimized using a commercial lysis cocktail (Figure S6).

Next, a sensitive activity assay for the CMB-immobilized CDA enzymes using COS substrates with DP from 2 to 6 in HTS format in 96-well microtiter plates was studied. Methods to quantify acetate release or amine formation were evaluated for sensitivity, reproducibility, and automation on a liquid handling platform. Amine quantification by reaction with fluorescamine and fluorescence readout proved to be the best method for HTS implementation. *Vc*CDA has maximum activity on diacetylchitobiose (DP2) as substrate and a 10-fold lower activity on tetraacetylchitotetraose (DP4).[14] Since the assay will be used to screen libraries for gain-of-function or increase of activity, assay optimization was performed with the slow-reacting DP4 substrate to define a wider dynamic range. Using a fixed excess concentration of the fluorescamine reagent, the volumes of cell-free extract and CMB were varied to set the working conditions (Figure 2).

The final method is able to discriminate between positive (active wt enzyme) and negative (void plasmid) levels with a calculated Z'-factor of 0.5 (Figure 2D), which reflects a good quality for an HTS assay.[30] The reproducibility of the assay had a coefficient of variation (CV) in the range of 6−12% for different independent samples (Figure S7). The final optimized and validated protocol is given in the Supporting Information.

The HTS assay was applied to screen mutant libraries of the model *Vibrio cholera* enzyme in the framework of a directed evolution program. *Vc*CDA is highly specific for deacetylating the penultimate GlcNAc residue from the nonreducing end of COS, with DP2 being the most active substrate and activity rapidly decreasing with increasing DP. For applications in biocatalysis, we were interested in engineering the enzyme to

DOI: 10.1021/acs.analchem.8b02729
Anal. Chem. 2018, 90, 10654−10658

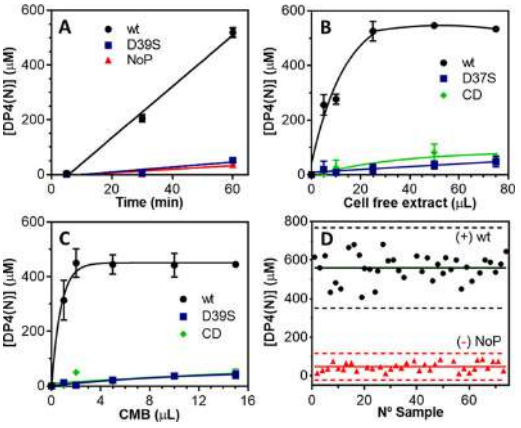


Figure 2. Deacetylase activity by the fluorescamine assay using the HTS protocol. Cell-free extracts of cultures expressing wt *Vc*CDA, an inactive D39S mutant,[14] wt CD (catalytic domain with no CBMs), and a void plasmid (NoP, no CDA expression) were incubated with conditioned CMBs in PBS buffer; DP4 substrate (2 mM) was added, and deacetylated product formation was quantified by reaction with fluorescamine. (A) Kinetics of deacetylated product (DP4(N)) formation. (B) DP4(N) formation at 60 min of reaction, varying the cell-free extract (5−75 μL) at fixed CMB (10 μL). (C) DP4(N) formation at 60 min of reaction, varying CMBs (1−15 μL) at fixed cell-free extract (25 μL). (D) Z′-factor calculation. DP4(N) formation at 60 min of reaction using final conditions (10 μL of CMB, 25 μL of cell-free extract) for 75 colonies of wt and NoP. Solid lines, average value; dotted lines, ±3 SD (standard deviation). All deacetylation reactions were done at 37 °C, pH 8.5.

have higher k_{cat} values for longer substrates (DP4 and DP5) but maintaining the high specificity for the deacetylated position. To this end, random mutant libraries of the catalytic domain of *Vc*CDA (CE4 deacetylase domain) were prepared by error-prone PCR (epPCR) at low mutation frequency (1−2 bp per kb) and subcloned into a recipient plasmid consisting of a pET vector containing the sequence of the two CBM12 modules to be fused at the C-terminus of the catalytic domain (Figure 1). After expression of the library, selective pressure to screen for more active mutants on the DP4 substrate was introduced by performing the activity step at a higher temperature. The temperature profile of the CMB-immobilized wt *Vc*CDA (Figure S8) showed that the enzyme retained less than 40% residual activity at 50 °C compared to its optimal activity at 37 °C. By carrying out the activity step at 50 °C in the HTS protocol, the wt enzyme becomes the negative level and active mutants will show up because they are either more thermostable or more active with higher residual activity at 50 °C.

The first generation of random mutagenesis rendered few active mutants from which mutant G1 had a significantly higher activity (Figure 3A). Gene sequencing showed that the G1 mutant had a single amino acid substitution, K275E. It was then used as a template for a second round of epPCR, this time selecting for active mutants at a higher temperature of 60 °C (Figure 3B). Of the few active mutants, the most active (G2) contained two mutations (K275E/H127R). Both G1 and G2 mutants were recovered and expressed, and the proteins were purified for biochemical characterization.

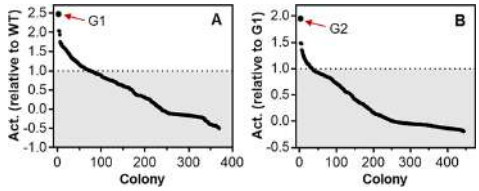


Figure 3. Random libraries by epPCR. (A) First generation, template: wt enzyme, screening for activity at 50 °C. (B) Second generation, template: G1 mutant, screening for activity at 60 °C.

Michaelis−Menten kinetics were evaluated with diacetylchitobiose (DP2), tetraacetylchitotetraose (DP4), and pentaacetylchitopentaose (DP5) at 37 °C and pH 8.5 (Table 1, Figure S9). The first generation K275E mutant has a slightly reduced k_{cat} for DP2 (2-fold lower) but has significantly higher k_{cat} values for the longer substrates DP4 and DP5 (6- and 4.5-fold relative to the wt enzyme). The K_M values have increased about 2- and 4-fold for the DP2 and DP5 substrates but less than 2-fold for the DP4 substrate. The main effect achieved with this first generation K275E mutant is an improved activity on longer substrates in terms of k_{cat} without compromising the activity on the natural DP2 substrate of *Vc*CDA, with a significantly modified k_{cat} ratio of DP2/DP4/DP5 from 10:1:1 for the wt to 10:13:8 for the mutant (Figure 4A). The second round of mutagenesis introduced an additional mutation to afford the K275E/H127R mutant as best hit. Remarkably, k_{cat} values for the DP4 and DP5 substrates have further increased up to 9-fold relative to the initial wt enzyme but retain the wt activity on DP2. The better binding reflected by the lower K_M values relative to the first mutant also resulted in improved k_{cat}/K_M values. Structural characterization of the reaction products confirmed that the pattern of acetylation has not been modified (Figure S12).

Table 1. Kinetics (A) and Stability (B) of wt and Mutant *Vc*CDA Enzymes Selected from the HTS Assay

		(A)		
	Subs[a]	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M ($M^{-1}\cdot s^{-1}$)
WT	DP2	2.02 ± 0.19	0.46 ± 0.10	4.4×10^3
	DP4	0.20 ± 0.01	0.64 ± 0.07	3.1×10^2
	DP5	0.17 ± 0.01	0.36 ± 0.03	4.7×10^2
K275E	DP2	0.94 ± 0.11	1.05 ± 0.25	8.9×10^2
	DP4	1.24 ± 0.01	0.95 ± 0.14	1.3×10^3
	DP5	0.77 ± 0.04	1.42 ± 0.18	5.4×10^2
K275E/H127R	DP2	2.19 ± 0.06	0.60 ± 0.04	3.7×10^3
	DP4	1.87 ± 0.10	0.77 ± 0.10	2.4×10^3
	DP5	1.42 ± 0.06	0.89 ± 0.10	1.6×10^3

	(B)	
	T_m (°C)	$t_{1/2}$ (min) at 50 °C[b]
WT	56.0 ± 0.2	5.4
K275E	56.5 ± 0.1	16.8
K275E/H127R	55.7 ± 0.1	24.1

[a]Conditions: 0.1−4 mM substrate, 0.05−0.8 μM enzyme, 50 mM phosphate buffer pH 8.5, 300 mM NaCl, 37 °C. [b]Thermotolerance at 50 °C. Half-life monitoring residual activity with DP4 substrate.

The temperature profile of both mutants showed a similar optimum temperature (37−40 °C, Figure 4B) and a similar

DOI: 10.1021/acs.analchem.8b02729
Anal. Chem. 2018, 90, 10654−10658

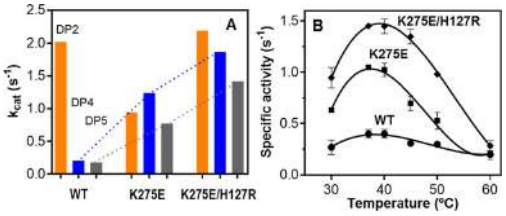


Figure 4. (A) k_{cat} values of wt and mutants on DP2, DP4, and DP5 substrates at 37 °C. (B) Temperature profiles with DP4 substrate.

thermal stability (T_m 56 ± 0.5 °C, Table 1, Figure S10) compared to those of the wt enzyme. However, the mutations conferred higher thermotolerance to the proteins, with an increase of their half-life ($t_{1/2}$) at 50 °C relative to the wt enzyme (Table 1, Figure S11). Therefore, the selective pressure imposed by screening the mutant libraries at high temperatures, where the wt enzyme just retained residual activity, resulted in the selection of mutants with combined higher specific activity and higher thermotolerance.

Preliminary inspection of the mutants location in the 3D structure of *Vc*CDA (PDB: 4NZ1, 4OUI)[14] suggests that the Glu for Lys replacement at position 275 disrupts a salt bridge with Glu270 that may result in a higher flexibility of an active site loop that blocks the access of longer substrates than DP2/DP3 to the reducing-end subsites of the binding side cleft in the wt enzyme. Therefore, the mutation may allow the partial opening of the loop to accommodate the DP4 and DP5 substrates with the concomitant increase of activity (Figure S13). The second mutation (H127R), however, is surprising since His127 is solvent-exposed on the surface and farther away from the binding site. This second mutation does not significantly alter the specificity relative to the single mutant but increases the overall activity of the enzyme. Further work (other mutations and molecular dynamics simulations) will be conducted to understand these effects on specificity and activity.

In conclusion, we have implemented a convenient and sensitive HTS assay to screen directed evolution libraries of CDAs and other CE4 enzymes active on COS. The assay easily allows the adjustment of the dynamic range to different activity levels by modifying the reaction time, temperature, and/or volume of the cell extract in the activity step to adapt the signal intensities to the desired positive and negative levels when using different COS substrates or even partially deacetylated substrates in the search for different specificities. The same assay principle would be applicable to the screening of inhibitors of CE4 enzymes (including chitin and peptidoglycan deacetylases) as they are antimicrobial targets.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.8b02729.

Protocols, materials, figures of binding efficiency, specificity, kinetics, deacetylase activity, *E. coli* expression hosts, cell lysis, reproducibility of the HTS assay, temperature profile, determination of Michaelis-Menten parameters, thermal stability, thermotolerance, deacetylation specificity, HTS assay development, and characterization of mutants (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: antoni.planas@iqs.edu.

ORCID

Antoni Planas: 0000-0001-7073-3320

Author Contributions

A.P. conceived the work and designed the experiments; S.P. performed the experiments. A.P. and S.P. wrote the paper.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

S.P. acknowledges a contract from NANO3BIO project. Work was supported by the EU 7FP Grant No. 613931 (NANO3-BIO) and Grant BFU2016-77427-C2-1-R from MINECO, Spain.

■ REFERENCES

(1) Zhao, Y.; Park, R. D.; Muzzarelli, R. A. A. *Mar. Drugs* **2010**, *8*, 24−46.
(2) Li, X.; Min, M.; Du, N.; Gu, Y.; Hode, T.; Naylor, M.; Chen, D.; Nordquist, R. E.; Chen, W. R. *Clin. Dev. Immunol.* **2013**, *2013*, 1.
(3) Das, S. N.; Madhuprakash, J.; Sarma, P. V. S. R. N.; Purushotham, P.; Suma, K.; Manjeet, K.; Rambabu, S.; Gueddari, N. E. El; Moerschbacher, B. M.; Podile, A. R. *Crit. Rev. Biotechnol.* **2015**, *35*, 29−43.
(4) Aam, B. B.; Heggset, E. B.; Norberg, A. L.; Sørlie, M.; Vårum, K. M.; Eijsink, V. G. H. *Mar. Drugs* **2010**, *8*, 1482−1517.
(5) Omura, Y.; Shigemoto, M.; Akiyama, T.; Saimoto, H.; Shigemasa, Y.; Nakamura, I.; Tsuchido, T. *Biocontrol Sci.* **2003**, *8*, 25−30.
(6) Santos-Carballal, B.; Fernández Fernández, E.; Goycoolea, F. *Polymers (Basel, Switz.)* **2018**, *10*, 444.
(7) Liaqat, F.; Eltem, R. *Carbohydr. Polym.* **2018**, *184*, 243−259.
(8) Hamer, S. N.; Cord-Landwehr, S.; Biarnés, X.; Planas, A.; Waegeman, H.; Moerschbacher, B. M.; Kolkenbrock, S. *Sci. Rep.* **2015**, *5*, 8716.
(9) Naqvi, S.; Moerschbacher, B. M. *Crit. Rev. Biotechnol.* **2017**, *37*, 11−25.
(10) Hembach, L.; Cord-Landwehr, S.; Moerschbacher, B. M. *Sci. Rep.* **2017**, *7*, 17692.
(11) Chavan, S. B.; Deshpande, M. V. *Biotechnol. Prog.* **2013**, *29*, 833−846.
(12) Blair, D. E.; Schuttelkopf, A. W.; MacRae, J. I.; van Aalten, D. M. F. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 15429−15434.
(13) Blair, D. E.; Hekmat, O.; Schüttelkopf, A. W.; Shrestha, B.; Tokuyasu, K.; Withers, S. G.; van Aalten, D. M. *Biochemistry* **2006**, *45*, 9416−9426.
(14) Andrés, E.; Albesa-Jové, D.; Biarnés, X.; Moerschbacher, B. M.; Guerin, M. E.; Planas, A. *Angew. Chem., Int. Ed.* **2014**, *53* (27), 6882−6887.
(15) Grifoll-Romero, L.; Pascual, S.; Aragunde, H.; Biarnés, X.; Planas, A. *Polymers (Basel).* **2018**, *10*, 352.
(16) Tuveng, T. R.; Rothweiler, U.; Udatha, G.; Vaaje-Kolstad, G.; Smalås, A.; Eijsink, V. G. H. *PLoS One* **2017**, *12*, e0187544.
(17) Aragunde, H.; Biarnés, X.; Planas, A. *Int. J. Mol. Sci.* **2018**, *19*, 412.
(18) Shao, W.; Wiegel, J. *Appl. Environ. Microbiol.* **1995**, *61*, 729−733.
(19) Liu, J.; Jia, Z.; Li, S.; Li, Y.; You, Q.; Zhang, C.; Zheng, X.; Xiong, G.; Zhao, J.; Qi, C.; Yang, J. *Gene* **2016**, *590*, 79−84.
(20) Hekmat, O.; Tokuyasu, K.; Withers, S. G. *Biochem. J.* **2003**, *374*, 369−380.

DOI: 10.1021/acs.analchem.8b02729
Anal. Chem. 2018, 90, 10654−10658

(21) Tsigos, I.; Bouriotis, V. *J. Biol. Chem.* **1995**, *270*, 26286−26291.
(22) Naqvi, S.; Cord-Landwehr, S.; Singh, R.; Bernard, F.; Kolkenbrock, S.; El Gueddari, N. E.; Moerschbacher, B. M. *Appl. Environ. Microbiol.* **2016**, *82*, 6645−6655.
(23) Aguila, E. M. D.; Gomes, L. P.; Andrade, C. T.; Silva, J. T.; Paschoalin, V. M. F. *Am. J. Mol. Biol.* **2012**, *2*, 341−350.
(24) Leane, M. M.; Nankervis, R.; Smith, A.; Illum, L. *Int. J. Pharm.* **2004**, *271*, 241−249.
(25) Cord-Landwehr, S.; Ihmor, P.; Niehues, A.; Luftmann, H.; Moerschbacher, B. M.; Mormann, M. *Anal. Chem.* **2017**, *89*, 2893−2900.
(26) Li, X.; Wang, L. X.; Wang, X.; Roseman, S. *Glycobiology* **2007**, *17*, 1377−1387.
(27) Hirano, T.; Sugiyama, K.; Sakaki, Y.; Hakamata, W.; Park, S. Y.; Nishio, T. *FEBS Lett.* **2015**, *589*, 145−151.
(28) New England BioLabs. *Chitin Magnetic Beads*; https://www.neb.com/-/media/catalog/datacards-or-manuals/e8036datasheet-lot0011312.pdf (accessed Apr 19, 2018).
(29) Verma, S. C.; Mahadevan, S. *J. Bacteriol.* **2012**, *194*, 4959−4971.
(30) Zhang, J. H.; Chung, T. D. Y.; Oldenburg, K. R. *J. Biomol. Screening* **1999**, *4*, 67−73.

DOI: 10.1021/acs.analchem.8b02729
Anal. Chem. 2018, 90, 10654−10658

International Journal of
Molecular Sciences

MDPI

Article

Indolyl Septanoside Synthesis for In Vivo Screening of Bacterial Septanoside Hydrolases

Aditya R. Pote [1,†], Sergi Pascual [2,†], Antoni Planas [2,*] and Mark W. Peczuh [1,*]

[1] Department of Chemistry, University of Connecticut, 55 N. Eagleville Road U3060, Storrs, CT 06269, USA; aditya.pote@gmail.com
[2] Laboratory of Biochemistry, Institute Químic de Sarrià, University Ramon Llull, 08017 Barcelona, Spain; sergipascualt@iqs.url.edu
* Correspondence: antoni.planas@iqs.edu (A.P.); mark.peczuh@uconn.edu (M.W.P.)
† These authors made equal contributions.

Abstract: Building-up and breaking-down of carbohydrates are processes common to all forms of life. Glycoside hydrolases are a broad class of enzymes that play a central role in the cleavage of glycosidic bonds, which is fundamental to carbohydrate degradation. The large majority of substrates are five- and six-membered ring glycosides. Our interest in seven-membered ring septanose sugars has inspired the development of a way to search for septanoside hydrolase activity. Described here is a strategy for the discovery of septanoside hydrolases that uses synthetic indolyl septanosides as chromogenic substrates. Access to these tool compounds was enabled by a route where septanosyl halides act as glycosyl donors for the synthesis of the indolyl septanosides. The screening strategy leverages the known dimerization of 3-hydroxy-indoles to make colored dyes, as occurs when the β-galactosidase substrate X-Gal is hydrolyzed. Because screens in bacterial cells would enable searches in organisms that utilize heptoses or from metagenomics libraries, we also demonstrate that septanosides are capable of entering *E. coli* cells through the use of a BODIPY-labeled septanoside. The modularity of the indolyl septanoside synthesis should allow the screening of a variety of substrates that mimic natural structures via this general approach.

Keywords: septanoside; indolyl glycoside; glycosidase; septanoside hydrolase

Citation: Pote, A.R.; Pascual, S.; Planas, A.; Peczuh, M.W. Indolyl Septanoside Synthesis for In Vivo Screening of Bacterial Septanoside Hydrolases. *Int. J. Mol. Sci.* **2021**, *22*, 4497. https://doi.org/10.3390/ijms22094497

Academic Editor: Vladimír Křen

Received: 30 March 2021
Accepted: 23 April 2021
Published: 26 April 2021

1. Introduction

Glycoside hydrolases (GH) are a broad superfamily of carbohydrate active enzymes that perform important functions in glycobiology—degradation of structural polysaccharides, remodeling of cell-surface proteoglycans and glycolipids, and even activation of small molecule natural products from latent precursor compounds [1–5]. Glycoside hydrolases (also referred to as glycosidases) cleave acetal linkages (glycosides) between a sugar and an aglycone moiety. Classification of GH families is based on the configuration (α- vs. β-) and location of the hydrolyzed bond (i.e., *endo*- vs. *exo*-), the identity of the sugar residue that is cleaved (e.g., D-glucose and D-galactose), and the three-dimensional shape of the hydrolase. The widely used repository of Carbohydrate Active Enzymes (CAZy database, www.cazy.org (accessed on 1 March 2021)) [6] classifies CAZymes in families based on sequence similarity and three-dimensional (3D) folds. Among glycoside hydrolases, there are currently (March 2021) 170 GH families. Members of each family share the same reaction mechanism and 3D fold, but different substrate specificities are found in most of the families, meaning that substrate specificity is dictated by subtle structural differences at the active site. Substrates typically consist of a pyranose or a furanose ring attached to a variety of aglycone species. Seven-membered ring septanose glycosides are all but unprecedented in natural systems [7,8]. While they are not true septanosides because the seven-membered ring is not linked through a glycosidic linkage, the recently reported natural products portulasoid **1** and 20-hydroxy-ecdysone septanoside **2** from

Atriplex portulacoides roots (Figure 1a) [9] open the provocative possibility that other similar structures, in which the polyhydroxy oxepane might be a septanoside (e.g., **4** in Figure 1a), could be discovered. In principle, it might be possible that septanoside hydrolases are present in nature or that promiscuous glycoside hydrolases might also accept septanoside substrates. Indeed, it is known that the seven-member ring iminosugar 1,6-dideoxy-1,6-imino-L-iditol was a glycosidase inhibitor [10], and it was reported that 4-nitrophenyl L-idoseptanosides, septanose analogs from hexoses, were substrates of some glycoside hydrolases, albeit weak ones [11].

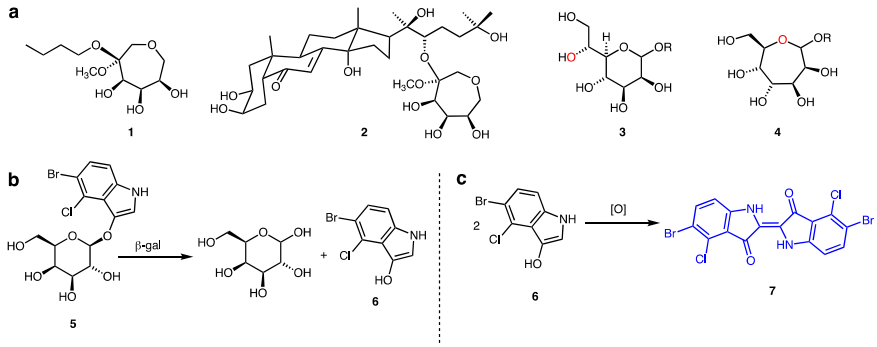


Figure 1. (**a**) Portulasoid **1** and 20-hydroxy-ecdysone septanoside **2** from *Atriplex portulacoides* roots; D-*glycero*-D-*manno*-heptose in its pyranose (**3**) and septanose (**4**) ring forms. (**b**) Hydrolysis of X-Gal **5**, giving rise to D-galactose and 3-hydroxyindole **6**. (**c**) Oxidative dimerization of 3-hydroxyindole **6** to indigo dye **7**.

The identification of a naturally occurring septanoside hydrolase or other GH sufficiently promiscuous to hydrolyze septanosides would be an important milestone. Such an enzyme, or mutants thereof, would open up whole new areas of glycobiology research. For example, discovery of a septanoside hydrolase would motivate the search for septanoside-containing glycoconjugates in the producing organism. Plants are a candidate kingdom in this search because of the discovery of compounds **1** and **2**. An even more promising source, however, might be Gram-negative bacteria. There are numerous examples of heptose-containing glycoconjugates present in Gram-negative species [12,13], with the sedoheptulose/D-*glycero*-D-*manno*-heptose (**3**) biosynthetic pathway being a representative [14]. Cyclization of **3** through its C6 hydroxyl group (red) instead of the C5 hydroxyl could ultimately give rise to septanose glycosides **4**. Biosynthetic pathways involving heptoses are still in the early days of being thoroughly characterized, which also adds an incentive to a search for new glycosyl hydrolases. A first step in any search for septanoside hydrolases, therefore, is the development of tools—an assay strategy and matching substrate—that could be used to identify these enzymes.

One straightforward approach to search for septanoside hydrolases borrows from the β-galactosidase assay widely used in molecular biology. Upon glycoside hydrolysis of a pro-chromogenic substrate such as 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside **5** (X-Gal) [15], D-galactose is liberated along with 3-hydroxyindole **6** (Figure 1b). Oxidative dimerization of **6** results in the formation of an easily detected indigo dye **7**. Development of a blue color in a bacterial colony or culture of when X-Gal is present therefore reports on the presence of the β-galactosidase enzyme. In practice, the "blue-white colony screen" is frequently applied to the preparation of a protein expression vector [16]. The multiple cloning site where a DNA construct encoding the protein of interest will be inserted is positioned within the β-galactosidase gene on the plasmid. Plasmids lacking the insert retain a functioning β-galactosidase and hence lead to blue colonies after transformation. Successful insertion of the construct into the plasmid, on the other hand, disrupts the β-galactosidase gene and the associated hydrolase activity, giving rise to white colonies after transformation. In the context of our search, we envisioned using indolyl septanosides

as candidate substrates of putative septanoside hydrolases. If hydrolytic activity toward such a substrate were present in a bacterial colony, we would anticipate the appearance of a characteristic blue color. To that end, we report here the syntheses of indolyl septanoside substrates for the discovery of septanoside hydrolases. Their structures mimic X-Gal, allowing them to be used for in vivo screening of bacteria. We further demonstrate that a related BODIPY–septanoside conjugate is transported into the cytoplasm of *E. coli*, probably using any of its numerous sugar transporters. Development of these tools constitutes the first steps toward the identification of septanoside hydrolase enzymes.

2. Results

2.1. Synthesis of Indolyl Septanosides

Indolyl septanosides **8** and **9** (Figure 2a) became the initial targets of substrate synthesis for a few reasons. First among them were the structural analogy of **8** and **9** to X-Gal **5** and reports on methods for the synthesis and utilization of indolyl glycosides in glycosidase assays [17]. The substitution pattern on the 5-bromo-3-indolyl aglycone was based on the ready availability of the 5-bromo-anthranilic acid starting material and the color of tyrian purple dye (an analog of **7** lacking the chlorine atoms) that arises from its oxidative dimerization [18,19]. Further, the stereogenic centers from C3-C6 in β-configured D-*glycero*-D-*gulo*-septanoside **8** and D-*glycero*-D-*ido*-septanoside **9** have identical relative configurations as D-glucose from C2-C5. The "phased" correspondence between these centers was important to recognition of septanosides by lectins [20–22]; it refers to the fact that C2 of the pyranose is the same as the C3 of the septanose and so forth. We also considered the possible promiscuity of glucosidases as giving us a higher probability of identifying active GHs in our search. The synthetic strategy drew from reported methods for indolyl glycoside synthesis [18,23–25] and our previous experience at preparing septanose glycosides via nucleophilic displacement on anomeric bromides [26]. It leveraged a synthesis of the per-*O*-acetyl septanose precursors of the anomeric bromides that began from natural D-pyranosides such as D-glucose, suggesting that the route could be extended to other sugars.

Preparation of indolyl septanoside **8** leveraged the known conversion of per-*O*-acetyl septanose **10** to its corresponding α-configured anomeric bromide **11** (93%, Figure 2b) [26]. Attempted glycosylation of **11** via direct displacement (S_N2) conditions with 5-bromo-indoxyl (i.e., the analog of **12** lacking the C2 carboxylate group) was unsuccessful due to rapid oxidative dimerization of the indoxyl species under the basic reaction conditions (**11** + **12**, potassium *tert*-butoxide in acetonitrile, 0 °C to rt). We therefore resorted to a strategy that used 2-carbomethoxy indoxyl **12** as acceptor. The carbomethoxy moiety at the 2-position of the indoxyl ring is a blocking group that prevents oxidative dimerization during the glycosylation [18]. In the event, glycosylation of **11** with **12** gave protected septanoside **13** in 48% yield; the yield for the glycosylation is modest but was not extensively optimized. Recompense for successful glycosylation in this instance was the multi-step deprotection that had to be undertaken to arrive at the target. Compound **13** was therefore converted to key intermediate **14** by a sequence that included removal of the acetate protecting groups and the methyl esters in a two-step process. Re-acetylation of the hydroxyl groups in this species was concomitant with decarboxylation of the carboxy group yielding **13** in 73% over the three steps. It proved convenient to purify **13** at this stage before the final de-acetylation. Removal of the acetates under Zemplén conditions then provided 5-bromo-3-indolyl septanoside **8** in 75% yield. Overall, the yield of **8** over the six-step sequence was 24% with an average yield per step of 73%. Compound **9**, the C2′-epimer of **8**, was prepared from per-*O*-acetyl septanose **15** using the same set of reactions, and shown in Figure 2b, in 35% yield over the six-step sequence.

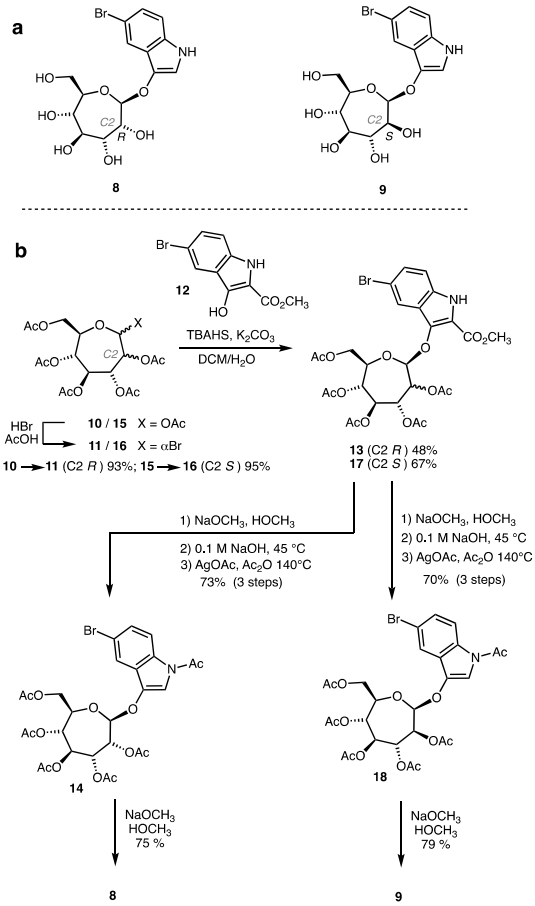


Figure 2. (**a**) Indolyl septanosides **8** and **9**. (**b**) Synthetic routes, including yields, used to prepare indolyl septanosides **8** and **9**.

Indolyl septanoside **8** was characterized both structurally and functionally. NMR spectroscopic data for **8** were consistent with a β-configured anomeric center (H1 δ = 4.93 ppm, C1 δ = 110.1 ppm) and a trans disposition of the groups at C1 and C2 ($^3J_{H1,H2}$ = 5.5 Hz). These values were comparable to other 1,2-trans β-septanosides we have synthesized, including *p*-nitrophenyl (pNP) septanoside **19** (vide infra). Values for the 1,2-cis indolyl septanoside **9** (H1 δ = 5.09 ppm, C1 δ = 105.5 ppm, and $^3J_{H1,H2}$ = 2.3 Hz) were also consistent with its proposed structure [8,26]. In a separate line of experimentation, septanoside **9** was exposed to buffer solutions at pH 5, 6, and 7, as well as a 1 M solution of HCl. With standing at room temperature over 8 h, color developed only in the 1 M HCl samples. We interpreted the results to be a demonstration of the stability of the indolyl glycosides at pH values near neutrality and in a range where many GH enzymes are active. The results also demonstrated the susceptibility to acid hydrolysis. In a preparative scale experiment, the indigo precipitate that arose from 1M HCl hydrolysis of **9** was collected, redissolved in *N*,*N*-dimethylformamide, and its UV–vis spectrum compared to a sample of tyrian purple synthesized independently by a different method [27–29].

2.2. Evaluation of Indolyl Septanoside **8** *as a Substrate of Common Exo-Glycoside Hydrolases*

Compound **8** was assayed in vivo with *E. coli* cells under conditions that are typical for assays that use X-Gal substrate **5**. *E. coli* BL21(DE3) Star cells harboring an inducible

plasmid overexpressing the *E. coli* β-galactosidase (BL21-pGal$^+$) and cell harboring an empty plasmid (BL21-pGal$^-$) were grown on Petri plates containing compound **8** with isopropyl 1-thio-β-D-galactopyranoside (IPTG) as inducer. No blue colonies were observed (as opposed to the control BL21-pGal$^+$ with **7**) (Figure 3). At first glance, the apparent lack of activity, as evidenced by the lack of appearance of blue colonies, may reflect that *E. coli* β-galactosidase did not recognize septanoside **8**. It may also have indicated that the substrate was not internalized into the *E. coli* cells. To test the first scenario, we assayed the purified *E. coli* β-galactosidase with compound **8** under the same experimental conditions where it shows maximum activity with its cognate substrate (Table 1). Once again, no hydrolytic activity was detected, confirming that septanoside **8** is not a β-galactosidase substrate.

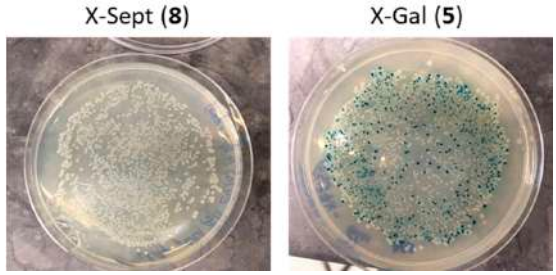


Figure 3. Blue-white colony screen of *E. coli* BL21(DE3) Star cells with indolyl septanoside **8**. BL21(DE3) Star cells harboring a plasmid expressing *E. coli* β-galactosidase: (**Left**) with X-Sept (**8**) substrate; and (**Right**) with substrate X-Gal (**5**).

Table 1. In vitro screen of selected glycosidases.

Enzyme	Substrate	S.A. (U/mg) [a]	Conditions
β-Galactosidase (*Escherichia coli*)	pNP-Gal **8** **9**	35 n.d. [b] n.d.	Phosphate buffer (100 mM) pH 6.5 40 °C
β-Galactosidase (*Aspergillus niger*)	pNP-Gal **8** **9**	170 <0.001 n.d.	Acetate buffer (100 mM) pH 4.5 40 °C
β-Glucosidase (*Streptomyces* sp.)	pNP-Glc **8** **9**	3.3 <0.005 n.d.	Phosphate buffer (50 mM) pH 6.5 50 °C
β-Glucosidase (*Almonds*)	pNP-Glc **8** **9**	2 n.d. n.d.	Phosphate buffer (100 mM) pH 5.0 37 °C
β-Glucosidase (*Thermatoga maritima*)	pNP-Glc **8** **9**	70 n.d. n.d.	Maleate buffer (50 mM) pH 6.5 40 °C
β-Glucosidase (*Phanerochaete chyrosporium*)	pNP-Glc **8** **9**	100 n.d. n.d.	Acetate buffer (100 mM) pH 5.0 40 °C
β-Mannosidase (*Cellulomonas fimi*)	pNP-Man **8** **9**	10 <0.001 <0.001	Maleate buffer (100 mM) pH 6.5 35 °C
β-Glucuronidase (*Escherichia coli*)	pNP-GlcA **8** **9**	110 n.d. n.d.	Tris·HCl buffer (100 mM) pH 7.5 37 °C
β-Xylosidase (*Bacillus pumilus*)	pNP-Xyl **8** **9**	18 n.d. n.d.	Phosphate buffer (50 mM) pH 7.5 35 °C

[a] Specific activity (U/mg) at 1 mM pNP-substrates. [b] n.d., not detected.

With the aim of potentially finding an enzyme that was able to recognize septanoside substrates, indolyl septanoside **8** was subjected to a panel of glycosidases to test in vitro for their ability to hydrolyze the substrate (Table 1). Each enzyme was assayed at high enzyme concentration (25–50 U/mL, activity units with its cognate substrate) with compounds **8** and **9** (1 mM) at the pH optimum of each enzyme. None of them showed measurable activity (by colorimetry of released indigo dye), including the *E. coli* β-galactosidase as indicated above. Only the *Streptomyces* sp. β-glucosidase (Bgl3) resulted in faint activity. This is a rather promiscuous enzyme reported to hydrolyze aryl glycosides of D-glucose, D-galactose, D-mannose, D-xylose and L-fucose [30]. We then analyzed the activity of Bgl3 against *p*-nitrophenyl septanoside **19**, prepared as previously reported [26], in greater detail. At a fixed Bgl3 concentration (9.2 μM), initial rates at increasing concentrations of **19** (Figure 4) were fit to give the following kinetics parameters: k_{cat} = 5.9 × 10^{-7} min^{-1}; K_M = 3 mM; h = 3.8. As compared to an efficient Bgl3 substrate (pNP-β-D-glucopyranoside, k_{cat} = 13.9 × 10^{-2} min^{-1}, K_M = 0.8 mM, K_i = 0.6 mM), the data show that **19** is a poor substrate for the glycosidase with a turnover of approximately five orders of magnitude slower than the glucoside.

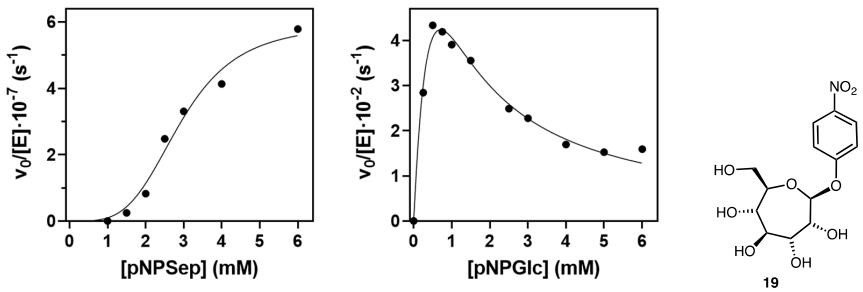


Figure 4. Kinetics of *Streptomyces* sp. β-galactosidase (Bgl3) with pNP-Sept (**19**) and pNP-Glc substrates. Conditions: 50 mM phosphate buffer, pH 6.5, and 50 °C.

2.3. Transport of a BODIPY-Labeled Septanoside into E. coli Cells

In search of septanoside hydrolases, we intend to use compound **8** for the screening of plant, yeast, and bacterial cell extracts as well as for in vivo "blue-white" screening of bacterial cells harboring libraries of glycoside hydrolases (i.e., metagenomics libraries) [31]. It was necessary, therefore, to return to the question of whether or not **8** was capable of entry into *E. coli* cells. To that end, we investigated the transport of septanosides into the bacterial cytoplasm. As a proof-of-concept, a BODIPY-labeled septanoside (i.e., **23**, Figure 5) was synthesized as a surrogate of compound **8** to be used in confocal microscopy studies. BODIPY is a commonly used fluorophore tag that can be used to analyze uptake by cells [32,33] and for cell sorting by fluorescence-activated cell sorting (FACS) [34,35]. A Huisgen click reaction [33,36] between commercially available BODIPY alkyne **20** and the known septanosyl azide **21** [26] gave protected conjugate **22** in 48% yield. Deprotection of compound **22** under Zemplén conditions then afforded BODIPY-labeled septanoside **23**.

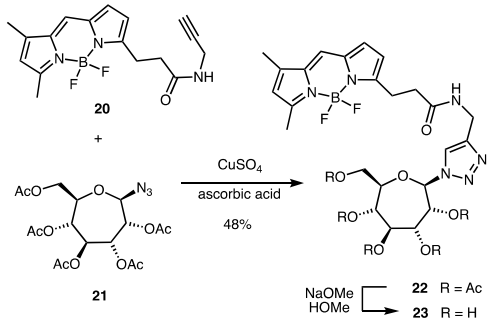


Figure 5. Synthesis of BODIPY-labeled septanoside **23**.

Two different *E. coli* strains, MG1655 and TOP10, were used to analyze cellular uptake by confocal microscopy (Figure 6). Prior to microscopy, cultures were grown to OD 1.0, washed, and then suspended in 10 mM MES buffer at pH 5.5. An aliquot of this culture was incubated with FM4-64 (localizes into bacterial membrane), BODIPY septanoside **23**, and DAPI (DNA stain), and then spun down and resuspended prior to observation. With MG1655 cells, the BODIPY fluorescence clearly localized in the cytoplasm with high intensity, confirming internalization of the septanoside. A lack of BODIPY fluorescence intensity at the cell periphery indicated that the labeled septanoside was not attached to the outer membrane. The merged image in Figure 6 shows the colocalization of the blue and green images, consistent with transport of **23**, which was considered a surrogate for indolyl septanoside **8** in these experiments. Similar results were obtained in experiments with the TOP10 cells, where **23** also localized in the cytoplasm of this *E. coli* strain (see Supplementary Materials).

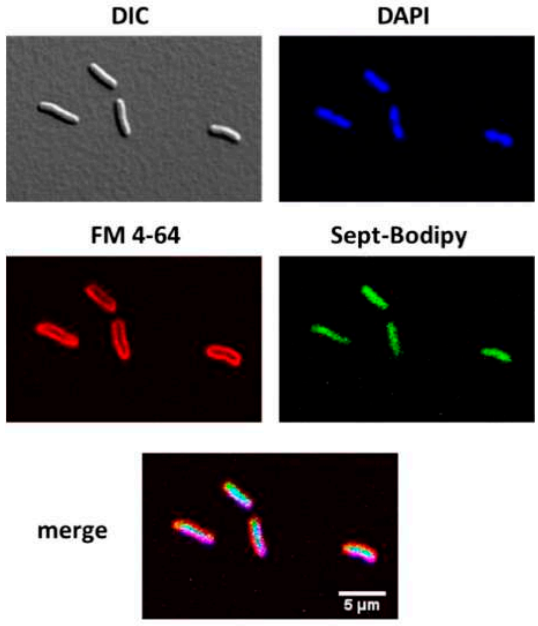


Figure 6. Confocal fluorescence microscopy of *E. coli* MG1655 cells with labeling stains. DIC, optical microscopy image; DAPI, DNA staining (bacterial chromosome); FM 4-64, membrane staining; Sept-BODIPY, BODIPY-labeled septanoside (**23**). Scale bar (5 µm) is the same in all images.

3. Discussion

The investigation reported here has established a conceptual framework behind the search for septanoside hydrolases from natural sources. Such a search will allow the identification of catabolic enzymes from organisms such as *Atriplex portulacoides* that produce polyhydroxy oxepanes (the source of compounds **1** and **2**) or Gram-negative organisms that utilize heptoses. Additionally, metagenomics libraries expressed in bacteria such as *E. coli* can be interrogated as well for hydrolytic activity. Our approach is fundamentally molecular—small molecule synthesis was used to prepare tools that enable searching both in vitro and, critically, in vivo for enzymes. Indolyl septanosides are useful reagents because they enable the colorimetric characterization of glycoside hydrolysis in a manner analogous to X-Gal. Preliminary assays in *E. coli* to characterize the reactivity of indolyl D-*glycero*-D-*gulo*-septanoside **8** as a substrate of β-galactosidase showed no hydrolysis. The apparent lack of reactivity of this septanoside substrate was subsequently explored along two different lines. First, in vitro experiments with a panel of commercially available glycosidases also showed no significant hydrolysis of the unnatural substrates **8** and **9**. Significantly, Bgl3 (β-glucosidase from *Streptomyces* sp.) showed a suggestive activity with light-pale blue colonies when using substrate **8**. Further analysis with *p*-nitrophenyl septanoside **19** confirmed the low activity, in the range of 5–6 orders of magnitude lower than its *p*-nitrophenyl D-glucoside substrate. In a second line of inquiry, BODIPY-labeled septanoside **23**, prepared via click reaction with a septanosyl azide, was observed inside *E. coli* cells by confocal microscopy. It suggested that septanosides traverse the bacterial envelope. Taken together, these results illustrate an approach that opens the door to a broader search for septanoside hydrolase activity in bacterial systems.

4. Materials and Methods

4.1. General Experimental

Commercially available starting materials, reagents, and solvents (Sigma-Aldrich, St. Louis, MO, USA and Acros/Thermofisher, Waltham, MA, USA) were used without further purification. Septanosyl azide **21** was prepared as previously reported [26]. Reactions were performed under nitrogen atmosphere unless otherwise noted and were monitored by TLC using silica gel HL plates w/UV254, 250 μm (SiliCycle, Quebec City, PQ, Canada), visualized either under UV lamp or by charring with 2.5% *p*-anisaldehyde in H_2SO_4, AcOH and EtOH solutions. Reverse phase C18 TLC plates w/UV254, 200 μm (Sorbtech, Atlanta, GA, USA) were also used and visualized by UV. Flash chromatography was performed on silica gel (60 Å, 40–63 μm). ^{1}HNMR spectra were collected on Bruker NMR instruments at either 400 and 500 MHz with chemical shift referenced to $(CH_3)_4Si$ (δ_H 0.00 ppm) or the residual peak in $CDCl_3$ (δ_H 7.24 ppm) or CD_3OD (δ_H 3.31 ppm). ^{13}C NMR spectra were collected at 100 MHz and referenced to residual peak in $CDCl_3$ (δ_C 77.2 ppm) or CD_3OD (δ_C 49.1 ppm). High resolution mass spectrometry data were collected on a JEOL DART Electrospray-Time-of-Flight (AccuTOF-DART) instrument. Enzymes were purchased from Megazyme Ltd. (Bray, Ireland). β-Glucosidase from *Streptomyces* was expressed and purified as reported in [30]. DAPI (4′,6-diamidine-2′-phenylindole dihydrochloride) and FM 4-64 (*N*-(3-triethylammoniumpropyl)-4-(6-(4-(tiethylamino)phenyl)hexatrienyl)pyridinium dibromide) were purchased from Sigma-Aldrich and Thermofisher, respectively.

4.2. Synthesis of Indolyl Septanosides **8** and **9**

Phase-Transfer Glycosylation (General Procedure). Freshly prepared per-*O*-acetyl septanosyl bromide (i.e., **10**) [26] (0.306 g, 0.63 mmol), tetra-*N*-butylammonium hydrogen sulfate (TBAS, 0.214 g, 0.63 mmol), and 2-methoxycarbonyl-3-hydroxy-5-bromo-indole **11** (0.210 g, 0.79 mmol) were mixed in DCM (8 mL). To this solution was added 12 mL of a 1 M aqueous solution of K_2CO_3. The biphasic (organic:aqueous) reaction mixture was stirred vigorously at room temperature until consumption of the bromide was complete as determined by TLC (1:1 Hex:EtOAc). Afterwards, the organic phase was separated from the aqueous phase; the aqueous phase was extracted with additional DCM (1 × 12 mL). The

combined organic layers were dried with Na_2SO_4, filtered, and the and the solvent was removed under reduced pressure. The residue was then purified by column chromatography in the solvent mixtures stated.

5-bromo-2-methoxycarbonyl-indol-3-yl 2,3,4,5,7-penta-*O*-acetyl-β-D-*glycero*-D-*gulo*-septanoside (**13**). Obtained as brownish oil from **11** and **12** to yield 0.210 g (48%) of **13**. R_f 0.4 (1:1 Hex: EtOAc); ^{1}H NMR (400 MHz, $CDCl_3$) δ 9.28 (s, 1H), 7.93 (broad s, 1H), 7.36 (dd, *J* = 8.8, 1.8 Hz, 1H), 7.23 (d, *J* = 8.8 Hz, 1H), 5.74 (dd, *J* = 4.7, 2.2 Hz, 1H), 5.64 (dd, *J* = 8.8, 2.2 Hz, 1H), 5.49–5.45 (m, 1H), 5.30 (d, *J* = 4.8 Hz, 1H), 5.22 (dd, *J* = 9.4, 6.0 Hz, 1H), 4.13 (dd, *J* = 14.3, 7.1 Hz, 1H), 4.08–4.02 (m, 1H), 3.90 (s, 3H), 3.88–3.85 (m, 1H), 2.22 (s, 3H), 2.12 (s, 3H), 2.11 (s, 3H), 2.01 (s, 3H), 1.72 (s, 3H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 170.6, 169.5, 169.4, 169.3, 169.1, 161.4, 139.0, 132.5, 129.2, 123.4, 123.2, 117.0, 113.8, 113.7, 105.4, 73.3, 72.7, 71.2, 70.9, 69.3, 63.5, 60.5, 52.1, 21.0, 20.7, 20.7(2), 20.3, 14.2; HRMS (DART-TOF) *m/z* calcd. for $C_{27}H_{30}BrNO_{12}Na$ $[M+Na]^+$ 694.0747, obs. 694.0743.

5-bromo-2-methoxycarbonyl-indol-3-yl2,3,4,5,7-penta-*O*-acetyl-β-D-*glycero*-D-*ido*-septanoside (**17).** Reaction of **12** with septanosyl bromide **16** using the general procedure gave 0.160 g (67%) of **17** as a brownish syrup. R_f 0.3 (3:2 Hex: EtOAc) ^{1}H NMR (400 MHz, $CDCl_3$) δ 9.16 (s, 1H), 7.96 (d, J = 1.7 Hz, 1H), 7.37 (dd, J = 8.8, 1.8 Hz, 1H), 7.23 (d, J = 8.8 Hz, 1H), 5.66 (dd, J = 7.4, 2.1 Hz, 1H), 5.61 (m, 1H), 5.52 (d, J = 2.2 Hz, 1H), 5.40–5.33 (m, 2H), 4.09 (dd, J = 12.2, 5.7 Hz, 1H), 3.98 (s, 3H), 3.94 (dd, J = 12.0, 2.2 Hz, 1H), 3.92 (ddd, J = 11.0, 5.8, 2.2 Hz, 1H), 2.18 (s, 3H), 2.10 (s, 3H), 2.04 (s, 3H), 1.98 (s, 3H), 1.68 (s, 3H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 170.8, 170.0, 169.5, 169.2, 169.0, 161.4, 139.6, 132.6, 129.5, 123.4, 123.1, 116.5, 114.0, 113.7, 103.2, 76.4, 72.1, 72.0, 70.5(2), 63.4, 60.6, 52.4, 20.9, 20.7, 20.6(2), 20.3, 14.3; HRMS (DART-TOF) m/z calcd. for $C_{27}H_{31}BrNO_{12}$ $[M+H]^+$ 672.0928, obs. 672.0923.

Multistep De-acetylation, Decarboxylation, and Acetylation (General Procedure). At room temperature, a solution of approximately 0.3 mmol of the protected indolyl septanoside (i.e., **13** or **17**) in 5 mL MeOH was treated with a catalytic amount of sodium methoxide (5 mol %). After starting material disappeared from the TLC (1:1 Hex:EtOAc), indicating that deacetylation was complete, the solvent was removed, and NaOH (0.1 M aq.; 14 mL) was added to the residue. The mixture was heated to 40–45 °C and stirred until C18 reverse-phase TLC indicated that ester hydrolysis was complete, approximately 4 h (R_f of the product is 0.7 and starting material is 0.3 in 3:1 water:acetonitrile). The mixture was lyophilized, then AgOAc (3.0 equiv.), K_2CO_3. (6.0 equiv.), and Ac_2O (10.0 mL) were added. The mixture was heated to 110 °C for 1 h. Then the mixture was cooled to room temperature and diluted with 20 mL water and 20 mL CH_2Cl_2. The organic phase was washed with water (2 × 20 mL) and dilute aqueous $NaHCO_3$ (~0.1 M, 1 × 20 mL). The organic phase was dried with Na_2SO_4, filtered, and the eluent was removed under reduced pressure. The residue was then purified by column chromatography in the solvent stated.

N-Acetyl-5-bromoindol-3-yl 2,3,4,5,7-penta-*O*-acetyl-β-D-*glycero*-D-*gulo*-septanoside (**14**). Obtained as a yellowish solid from **13** using the general procedure (73% yield over 3 steps). R_f 0.3 (1:1 Hex:EtOAc) ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.25 (broad s, 1H), 7.69 (d, *J* = 1.8 Hz, 1H), 7.46 (dd, *J* = 8.8, 1.8 Hz, 1H), 7.30 (broad s, 1H), 5.57–5.54 (m, 2H), 5.44 (dd, *J* = 8.0, 4.7 Hz, 1H), 5.15 (d, *J* = 5.4 Hz, 1H), 4.33–4.31 (m, 1H), 4.17–4.09 (m, 3H), 2.62 (s, 3H), 2.20 (s, 3H), 2.14 (s, 3H), 2.13 (s, 3H), 2.06 (s, 3H), 2.03 (s, 3H), 1.93 (s, 3H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 170.6, 169.3(2), 169.2, 169.0, 168.5, 140.0, 132.3, 129.1, 126.1, 120.8, 118.1, 117.2, 112.2, 103.5, 74.0, 72.2, 71.6, 70.6, 69.2, 63.7, 60.5, 24.0, 21.2, 21.0, 20.8(2), 20.7, 14.3; HRMS (DART-TOF) *m/z* calcd. for $C_{27}H_{30}BrNO_{13}Na$ $[M+Na]^+$ 678.0798, obs. 678.0785.

N-Acetyl-5-bromoindol-3-yl 2,3,4,5,7-penta-O-acetyl-β-D-*glycero*-D-*ido*-septanoside (**18**). Obtained as yellow solid from **17** using the general procedure (70% yield over 3 steps) R_f 0.2 (3:2 Hex: EtOAc) ^{1}H NMR (400 MHz, CD_3OD) δ 8.20 (d, J = 8.8 Hz, 1H), 7.62 (d, J = 1.9 Hz, 1H), 7.41 (dd, J = 8.8, 1.9 Hz, 1H), 7.37 (broad s, 1H), 5.64 (d, J = 1.5 Hz, 1H), 5.59–5.55 (m, 2H), 5.49–5.45 (m, 1H), 5.36 (app t, J = 9.3 Hz, 1H) 4.33 (ddd, J = 9.2, 7.4, 3.6Hz, 1H), 4.23 (app d, J = 3.6 Hz, 2H), 2.60 (s, 3H), 2.15 (s, 3H), 2.06 (s, 3H), 2.02 (s, 3H), 2.00 (s, 3H), 1.76 (s, 3H); ^{13}C NMR (100 MHz, CD_3OD) δ 172.1, 171.3, 171.0, 170.9, 170.7, 170.6, 140.9, 133.4, 129.8, 127.3, 121.5, 119.0, 117.7, 114.9, 112.2, 101.6, 76.7, 73.2, 72.7, 71.3,

71.2, 64.4, 23.9, 20.6(2), 20.5; HRMS (DART-TOF) *m/z* calcd. for $C_{27}H_{30}BrNO_{13}Na$ $[M+Na]^+$ 678.0798, obs. 678.0816.

Zemplén De-acetylation (General Procedure). A solution of the starting material (1.00 mmol) in MeOH (15.0–20.0 mL) was treated with a catalytic amount of sodium methoxide (5 mol%). The solution was stirred at room temperature overnight (~16 h). If the product had precipitated during this time it was collected by filtration, otherwise the solution was neutralized with Amberlite IR-120 (H+) resin (pH 7 by pH paper) and then the solution was concentrated to near dryness. The residue was then re-dissolved in a minimum amount of water ($\leq$5.0 mL) and subjected to lyophilization.

5-Bromoindol-3-yl β-D-*glycero*-D-*gulo*-septanoside (**8**). Obtained from **14** as tan powder after lyophilization using the general procedure (75% yield). R_f 0.4 (4:1 DCM: MeOH) ^{1}H NMR (400 MHz, CD_3OD) δ 7.75 (s, 1H), 7.23 (s, 1H), 7.19 (d, *J* = 8.7 Hz, 1H), 7.14 (d, *J* = 8.7 Hz, 1H), 4.99 (s, 6H), 4.93 (d, *J* = 5.5 Hz, 1H), 4.21 (app t, *J* = 4.4 Hz, 1H), 3.89 (dd, *J* = 9.1, 3.3 Hz, 1H), 3.86–3.79 (m, 2H), 3.66–3.60 (m, 2H), 3.46 (app t, *J* = 7.9 Hz, 1H); ^{13}C NMR (100 MHz, CD3OD) δ 137.9, 133.9, 125.6, 123.2, 121.0, 114.2(2), 112.7, 110.1, 82.3, 75.6, 75.5, 73.9, 72.2, 64.4; HRMS (DART-TOF) *m/z* calcd. for $C_{15}H_{19}BrNO_7$ $[M+H]^+$ 404.0345, obs. 404.0338.

5-Bromoindol-3-yl β-D-*glycero*-D-*ido*-septanoside (**9**). Obtained as light brown solid from **18** using the general procedure (79% yield) R_f 0.4 (4:1 DCM: MeOH) ^{1}H NMR (400 MHz, CD_3OD) δ 7.87 (s, 1H), 7.78 (d, *J* = 1.8 Hz, 1H), 7.20 (m, 1H), 7.16 (dd, *J* = 8.6, 1.8 Hz, 1H), 5.09 (d, *J* = 2.3 Hz, 1H), 4.88 (s, 6H), 4.00 (dd, *J* = 6.3, 2.2 Hz, 1H), 3.86–3.83 (m, 1H), 3.71–3.66 (m, 2H), 3.61 (dd, *J* = 14.1, 7.0 Hz, 1H), 3.52–3.50 (m, 2H); ^{13}C NMR (100 MHz, CD_3OD) δ 137.7, 133.9, 125.7, 123.3, 121.1, 114.2(2), 112.8, 105.5, 83.2, 76.7, 75.7, 75.5, 74.1, 64.3; HRMS (DART-TOF) *m/z* calcd. for $C_{15}H_{18}BrNO_7Na$ $[M+Na]^+$ 426.0164, obs. 426.0204.

4.3. Synthesis of BODIPY-Labeled Septanoside **23**

BODIPY:2,3,4,5,7-Penta-*O*-acetyl-β-D-*glycero*-D-*gulo*-septanosyl azide conjugate (**22**). BODIPY-FL alkyne **20** (0.0025 g, 0.008 mmol) and septanosyl azide **21** (0.0085 g, 0.02 mmol) and were dissolved in a mixture of THF:water (1 mL of 3:1 mixture). A solution of $CuSO_4 \cdot 5H_2O$ (0.001 g, 0.5 eq.) and sodium ascorbate (0.0012 g, 0.8 eq.) in an additional 1 mL of the THF:water was sonicated for 10 min and then added to the initial alkyne-azide solution. The mixture was allowed to react at room temperature with stirring for 12 h. After, the reaction mixture was diluted with 5 mL EtOAc and the aqueous and organic layers separated; the aqueous layer was then extracted with additional EtOAc (2 × 5 mL). The combined organic layers were dried over Na_2SO_4, filtered, and the solvent was evaporated under reduced pressure. The crude product was then purified by column chromatography to deliver **20** as brownish oil (0.0028 g, 48% yield) R_f 0.4 (3:2 Hex: EtOAc); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.09 (s, 1H), 6.88 (d, *J* = 3.9 Hz, 1H), 6.28 (d, *J* = 4.0 Hz, 1H), 6.12 (s, 1H), 5.82 (broad s, 1H), 5.40 (dd, *J* = 8.0, 1.9 Hz, 2H), 5.34 (dd, *J* = 8.0, 3.9 Hz, 2H), 5.10 (dd, *J* = 6.5, 1.9 Hz, 2H), 5.05–5.00 (m, 2H), 4.18–4.12 (m, 3H), 3.28 (t, *J* = 7.5 Hz, 2H), 2.67–2.63 (m, 2H), 2.57 (s, 3H), 2.26 (s, 3H), 2.12 (s, 3H), 2.11 (s, 9H), 2.09 (s, 3H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 171.5, 170.8, 169.3, 169.1 (2), 128.5, 124.0, 120.7, 117.7, 91.3, 79.7, 77.4, 76.0, 72.1, 71.8, 71.6, 70.6, 69.4, 64.3, 36.1, 29.9, 29.4, 25.0, 20.9 (3), 15.2, 11.5;HRMS (DART-TOF) *m/z* calcd. for $C_{34}H_{41}BF_2N_6O_{12}Na$ $[M+Na]^+$ 797.2730, obs. 797.2725.

BODIPY: β-D-*glycero*-D-gulo-septanosyl azide conjugate (**23**). A solution of starting per-*O*-acetylated BODIPY conjugate **22** (0.0028 g, 0.004 mmol) was dissolved in dry methanol (0.75 mL) and kept under an inert atmosphere and dark conditions. To this solution was added 5 mol % of NaOMe in MeOH. The reaction was stirred at room temperature for 12 h and monitored by TLC (R_f of the product is 0.1 and starting material is 0.5 in 1:1 Hex/EtOAc). Upon complete disappearance of the starting material, the reaction solvent was evaporated under the reduced pressure and the residue was re-dissolved in de-ionized water (1.00 mL) and subjected to lyophilization to obtain **23** as a brownish solid. The crude product was used in confocal microscopy experiments without further purification.

4.4. Glycosidase Activity Assay on E. coli Cells

Twenty microliters of 100 mM IPTG solution in water and 20 µL of a 20 mg/mL X-Gal (5) or X-Sept (**8**) solution in DMF were spread on LB agar Petri plates (20 mL medium) containing kanamycin (100 µg/mL) and allowed to dry for 20 min at 37 °C. Competent BL21 (DE3) Star cells were transformed (chemically competent cells [37]) with plasmid pRSF-β-gal encoding the *E. coli* β-galactosidase and spread onto the agar plates. Color development was analyzed after 24 h incubation at 37 °C.

4.5. Probing Glycoside Hydrolases for Septanoside Hydrolysis

Activity assays were performed at the optimal pH and temperature for each enzyme as indicated in Table 1. Indolyl substrates **8** and **9** (1 mM) and enzymes (concentration ranging from 25 to 50 U/mL) in buffer were incubated in 1 mL cuvettes and the absorbance monitored at 630 nm. Specific activities of the commercial enzymes with pNP substrates are from the manufacturer, and β-glucosidase Bgl3 from *Streptomyces* sp. as reported [30].

4.6. Kinetics of Streptomyces β-Glucosidase with p-Nitrophenyl Septanoside **19**

Kinetics were performed by monitoring *p*NP release by absorbance at 400 nm. Reactions were done in thermostated cuvettes at 50 °C in 50 mM phosphate buffer pH 6.5 with 0.1–10 mM substrates and 46 nM enzyme (Bgl3) for *p*NP-Glc or 9.2 µM Bgl3 for *p*NP-septanoside (**19**). Rates were obtained from the initial slopes after subtracting the blank rates (in the absence of enzyme) and kinetic constants were calculated from data fitted to a Michaelis–Menten equation with substrate inhibition (Equation (1)) for *p*NP-Glc and to a sigmoidal equation (Equation (2)) for *p*NP-Sept (**19**) using GraphPad software (Prism, San Diego, CA, USA).

$$v = k_{cat}[E]_0[S]/(K_M + [S] + [S]^2/K_i) \quad (1)$$

$$v = k_{cat}[E]_0[S]^h/\left({K_M}^h + [S]^h\right) \quad (2)$$

where $[S]$ is the substrate concentration, $[E]_0$ the enzyme concentration, and k_{cat} (catalytic constant), K_M (Michaelis constant), K_i (substrate inhibition constant) and h (Hill index) are the adjustable parameters [38].

4.7. Cell Internalization of Compound **23** *by Confocal Microscopy*

E. coli MG1655 cultures were grown to an OD_{600} of ~1.0 (1×10^9 cfu/mL) in Luria Broth (LB), washed, and suspended in 10 mM MES buffer pH 5.5 to a final OD_{600} = 1. To 100 µL of cell suspension (different dilutions tested, best results with a 1:2 dilution), BODIPY-septanoside **23** was added to a final 1 µM concentration with shaking. Cells were incubated for 1 h at 37 °C. Then, 1 µL FM 4–64 (1 µg/mL stock) and 1 µL DAPI (2 µg/mL stock) were added to the cells with shaking and incubated for 30 min at 37 °C for membrane and DNA staining, respectively. Cells were spun down by centrifugation for 60 s and resuspended in 1/10 their original volume. Three microliters of cell suspension were spotted onto 1.5% agarose pads containing solid LB medium. The cells were then imaged using a Nikon A1R spectral confocal microscope with a 60× oil immersion lens.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/ijms22094497/s1.

Author Contributions: A.R.P. and S.P., experimental and data analysis; and A.P. and M.W.P., conceptualization, supervision, and writing and acquisition of funding. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by NSF grant CHE-1506567 (to M.W.P.) and grant PID2019-104350RB-I00 from MICINN, Spain (to A.P.).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the corresponding authors upon reasonable request.

Acknowledgments: S.P. acknowledges a predoctoral contract from MICINN, Spain. We thank Alfredo Angeles-Boza and Samuel Juliano (UConn) for their support on confocal microscopy experiments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Pallister, E.; Gray, C.J.; Flitsch, S.L. Enzyme promiscuity of carbohydrate active enzymes and their applications in biocatalysis. *Curr. Opin. Struct. Biol.* **2020**, *65*, 184–192. [CrossRef]
2. Kytidou, K.; Artola, M.; Overkleeft, H.S.; Aerts, J.M.F.G. Plant Glycosides and Glycosidases: A Treasure-Trove for Therapeutics. *Front. Plant Sci.* **2020**, *11*. [CrossRef] [PubMed]
3. Gloster, T.M. Exploitation of carbohydrate processing enzymes in biocatalysis. *Curr. Opin. Chem. Biol.* **2020**, *55*, 180–188. [CrossRef]
4. Davies, G.J.; Planas, A.; Rovira, C. Conformational Analyses of the Reaction Coordinate of Glycosidases. *Acc. Chem. Res.* **2012**, *45*, 308–316. [CrossRef]
5. Sinnott, M.L. Catalytic Mechanisms of Enzymic Glycosyl Transfer. *Chem. Rev.* **1990**, *90*, 1171–1202. [CrossRef]
6. Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P.M.; Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **2014**, *42*, D490–D495. [CrossRef] [PubMed]
7. Dey, S.; Samanta, G.C.; Jayaraman, N. Advancements in synthetic and structural studies of septanoside sugars. In *Recent Trends in Carbohydrate Chemistry*; Rauter, A.P., Christensen, B., Somsak, L., Kosma, P., Adamo, R., Eds.; Elsevier Inc.: Amsterdam, The Netherlands, 2020; pp. 217–251.
8. Saha, J.; Peczuh, M.W. Synthesis and properties of septanose carbohydrates. In *Advances in Carbohydrate Chemistry and Biochemistry*; Academic Press: Cambridge, MA, USA, 2011; Volume 66, pp. 121–186.
9. Ben Nejma, A.; Nguir, A.; Ben Jannet, H.; Hamza, M.A.; Daïch, A.; Othman, M.; Lawson, A.M. New septanoside and 20-hydroxyecdysone septanoside derivative from Atriplex portulacoides roots with preliminary biological activities. *Bioorg. Med. Chem. Lett.* **2015**, *25*, 1665–1670. [CrossRef]
10. Le Merrer, Y.; Poitout, L.; Depezay, J.C.; Dosbaa, I.; Geoffroy, S.; Foglietti, M.J. Synthesis of azasugars as potent inhibitors of glycosidases. *Bioorg. Med. Chem.* **1997**, *5*, 519–533. [CrossRef]
11. Tauss, A.; Steiner, A.J.; Stütz, A.E.; Tarling, C.A.; Withers, S.G.; Wrodnigg, T.M. L-Idoseptanosides: Substrates of D-glucosidases? *Tetrahedron Asymmetry* **2006**, *17*, 234–239. [CrossRef]
12. Elshahawi, S.I.; Shaaban, K.A.; Kharel, M.K.; Thorson, J.S. A comprehensive review of glycosylated bacterial natural products. *Chem. Soc. Rev.* **2015**, *44*, 7591–7697. [CrossRef]
13. Guo, Z.; Tang, Y.; Tang, W.; Chen, Y. Heptose-containing bacterial natural products: Structures, bioactivities, and biosyntheses. *Nat. Prod. Rep.* **2021**. [CrossRef]
14. Tang, W.; Guo, Z.; Cao, Z.; Wang, M.; Li, P.; Meng, X.; Zhao, X.; Xie, Z.; Wang, W.; Zhou, A.; et al. D-Sedoheptulose-7-phosphate is a common precursor for the heptoses of septacidin and hygromycin B. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 2818–2823. [CrossRef]
15. Horwitz, J.P.; Chua, J.; Cubby, R.J.; Tomson, A.J.; Da Rooge, M.A.; Fisher, B.E.; Mauricio, J.; Klundt, I. Substrates for Cytochemical Demonstration of Enzyme Activity. I. Some Substituted 3-Indolyl-β-D-glycopyranosides. *J. Med. Chem.* **1964**, *7*, 574–575. [CrossRef] [PubMed]
16. Miller, J.H. *Experiments in Molecular Genetics*, 1st ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 1972.
17. Kiernan, J. Indigogenic substrates for detection and localization of enzymes. *Biotech. Histochem.* **2007**, *82*, 73–103. [CrossRef] [PubMed]
18. Böttcher, S.; Thiem, J. Indoxylic Acid Esters as Convenient Intermediates Towards Indoxyl Glycosides. *Eur. J. Org. Chem.* **2014**, *2014*, 564–574. [CrossRef]
19. Wolk, J.L.; Frimer, A.A. A simple, safe and efficient synthesis of Tyrian purple (6,6′-Dibromoindigo). *Molecules* **2010**, *15*, 5561–5580. [CrossRef]
20. Castro, S.; Duff, M.; Snyder, N.L.; Morton, M.; Kumar, C.V.; Peczuh, M.W. Recognition of septanose carbohydrates by concanavalin A. *Org. Biomol. Chem.* **2005**, *3*, 3869–3872. [CrossRef]
21. Duff, M.R.; Fyvie, W.S.; Markad, S.D.; Frankel, A.E.; Kumar, C.V.; Gascón, J.A.; Peczuh, M.W. Computational and experimental investigations of mono-septanoside binding by Concanavalin A: Correlation of ligand stereochemistry to enthalpies of binding. *Org. Biomol. Chem.* **2011**, *9*, 154–164. [CrossRef]
22. Sager, C.P.; Fiege, B.; Zihlmann, P.; Vannam, R.; Rabbani, S.; Jakob, R.P.; Preston, R.C.; Zalewski, A.; Maier, T.; Peczuh, M.W.; et al. The price of flexibility—a case study on septanoses as pyranose mimetics. *Chem. Sci.* **2018**, *9*, 646–654. [CrossRef]
23. Nagata, S.; Tomida, H.; Iwai-Hirose, H.; Tanaka, H.N.; Ando, H.; Imamura, A.; Ishida, H. Synthesis of a 1,2-: Cis -indoxyl galactoside as a chromogenic glycosidase substrate. *RSC Adv.* **2019**, *9*, 28241–28247. [CrossRef]

24. Böttcher, S.; Thiem, J. Synthesis of indoxyl-glycosides for detection of glycosidase activities. *J. Vis. Exp.* **2015**, *2015*, e52442. [CrossRef]
25. Böttcher, S.; Hederos, M.; Champion, E.; Dékány, G.; Thiem, J. Novel efficient routes to indoxyl glycosides for monitoring glycosidase activities. *Org. Lett.* **2013**, *15*, 3766–3769. [CrossRef]
26. Pote, A.R.; Vannam, R.; Richard, A.; Gascón, J.; Peczuh, M.W. Formation of and Glycosylation with Per-O-Acetyl Septanosyl Halides: Rationalizing Complex Reactivity En Route to p-Nitrophenyl Septanosides. *Eur. J. Org. Chem.* **2018**, *2018*, 1709–1719. [CrossRef]
27. de Melo, J.S.S.; Rondão, R.; Burrows, H.D.; Melo, M.J.; Navaratnam, S.; Edge, R.; Voss, G. Spectral and Photophysical Studies of Substituted Indigo Derivatives in Their Keto Forms. *ChemPhysChem* **2006**, *7*, 2303–2311. [CrossRef]
28. Rajesh, K.; Somasundaram, M.; Saiganesh, R.; Balasubramanian, K.K. Bromination of deactivated aromatics: A simple and efficient method. *J. Org. Chem.* **2007**, *72*, 5867–5869. [CrossRef] [PubMed]
29. Imming, P.; Imhof, I.; Zentgraf, M. An improved synthetic procedure for 6,6′-dibromoindigo (Tyrian Purple). *Synth. Commun.* **2001**, *31*, 3721–3727. [CrossRef]
30. Vallmitjana, M.; Ferrer-Navarro, M.; Planell, R.; Abel, M.; Ausín, C.; Querol, E.; Planas, A.; Pérez-Pons, J.A. Mechanism of the family 1 β-glucosidase from Streptomyces sp: Catalytic residues and kinetic studies. *Biochemistry* **2001**, *40*, 5975–5982. [CrossRef] [PubMed]
31. Stroobants, A.; Portetelle, D.; Vandenbol, M. New carbohydrate-active enzymes identified by screening two metagenomic libraries derived from the soil of a winter wheat field. *J. Appl. Microbiol.* **2014**, *117*, 1045–1055. [CrossRef] [PubMed]
32. Kowada, T.; Maeda, H.; Kikuchi, K. BODIPY-based probes for the fluorescence imaging of biomolecules in living cells. *Chem. Soc. Rev.* **2015**, *44*, 4953–4972. [CrossRef]
33. Uppal, T.; Bhupathiraju, N.V.S.D.K.; Vicente, M.G.H. Synthesis and cellular properties of Near-IR BODIPY-PEG and carbohydrate conjugates. *Tetrahedron* **2013**, *69*, 4687–4693. [CrossRef]
34. Aharoni, A.; Thieme, K.; Chiu, C.P.C.; Buchini, S.; Lairson, L.L.; Chen, H.; Strynadka, N.C.J.; Wakarchuk, W.W.; Withers, S.G. High-throughput screening methodology for the directed evolution of glycosyltransferases. *Nat. Methods* **2006**, *3*, 609–614. [CrossRef] [PubMed]
35. Liu, B.; Novikova, N.; Simpson, M.C.; Timmer, M.S.M.; Stocker, B.L.; Söhnel, T.; Ware, D.C.; Brothers, P.J. Lighting up sugars: Fluorescent BODIPY-: Gluco -furanose and -septanose conjugates linked by direct B-O-C bonds. *Org. Biomol. Chem.* **2016**, *14*, 5205–5209. [CrossRef] [PubMed]
36. Kolb, H.C.; Finn, M.G.; Sharpless, K.B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem. Int. Ed.* **2001**, *40*, 2004–2021. [CrossRef]
37. Green, M.R.; Hughes, H.; Sambrook, J.; MacCallum, P. *Molecular Cloning; A Laboratory Manual*, 4th ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 2012.
38. Segel, I.H. *Enzyme Kinetics: Behaviour and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems*; John Wiley & Sons Ltd.: New York, NY, USA, 1975.

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Chemical Biology

Carbohydrate de-*N*-acetylases acting on structural polysaccharides and glycoconjugates

Sergi Pascual and Antoni Planas

Abstract

Deacetylation of *N*-acetylhexosamine residues in structural polysaccharides and glycoconjugates is catalyzed by different families of carbohydrate esterases that, despite different structural folds, share a common metal-assisted acid/base mechanism with the metal cation coordinated with a conserved Asp-His-His triad. These enzymes serve diverse biological functions in the modification of cell-surface polysaccharides in bacteria and fungi as well as in the metabolism of hexosamines in the biosynthesis of cellular glycoconjugates. Focusing on carbohydrate de-*N*-acetylases, this article summarizes the background of the different families from a structural and functional viewpoint and covers advances in the characterization of novel enzymes over the last 2–3 years. Current research is addressed to the identification of new deacetylases and unravel their biological functions as they are candidate targets for the design of antimicrobials against pathogenic bacteria and fungi. Likewise, some families are also used as biocatalysts for the production of defined glycostructures with diverse applications.

Addresses

Laboratory of Biochemistry, Institut Químic de Sarrià, University Ramon Llull, 08017, Barcelona, Spain

Corresponding author: Planas, Antoni (antoni.planas@iqs.edu)

Current Opinion in Chemical Biology 2021, 61:9–18

This review comes from a themed issue on Biocatalysis and Biotransformation

Edited by Bernd Nidetzky and Tom Desmet

For a complete overview see the Issue and the Editorial

Available online 16 October 2020

https://doi.org/10.1016/j.cbpa.2020.09.003

Keywords

Carbohydrate esterases, de-*N*-acetylation, Metal-dependent hydrolases, Structural polysaccharides, Glycoconjugates, *N*-acetylglucosamine, Antimicrobial targets.

Introduction

Acetylation and deacetylation of biomolecules occurs on many different targets for a wide range of biological functions. The co-translational *N*-terminal protein acylation that impacts protein stability and localization and acetylation/deacetylation dynamics of histones that control their function are prime examples of structural and regulatory functions of adding or removing acetyl groups on polypeptides [1]. Likewise, deacetylation of *N*-acetylated and *O*-acetylated carbohydrates are involved in a plethora of biological functions: pathogenic microorganisms that deacetylate their surface glycans to evade immune responses or to support biofilm formation, deacetylation of glycoconjugates in signaling events, deacetylation of structural glycans in microbial cell wall morphogenesis, and sporulation or remodeling of plant cell wall polysaccharides are just some examples.

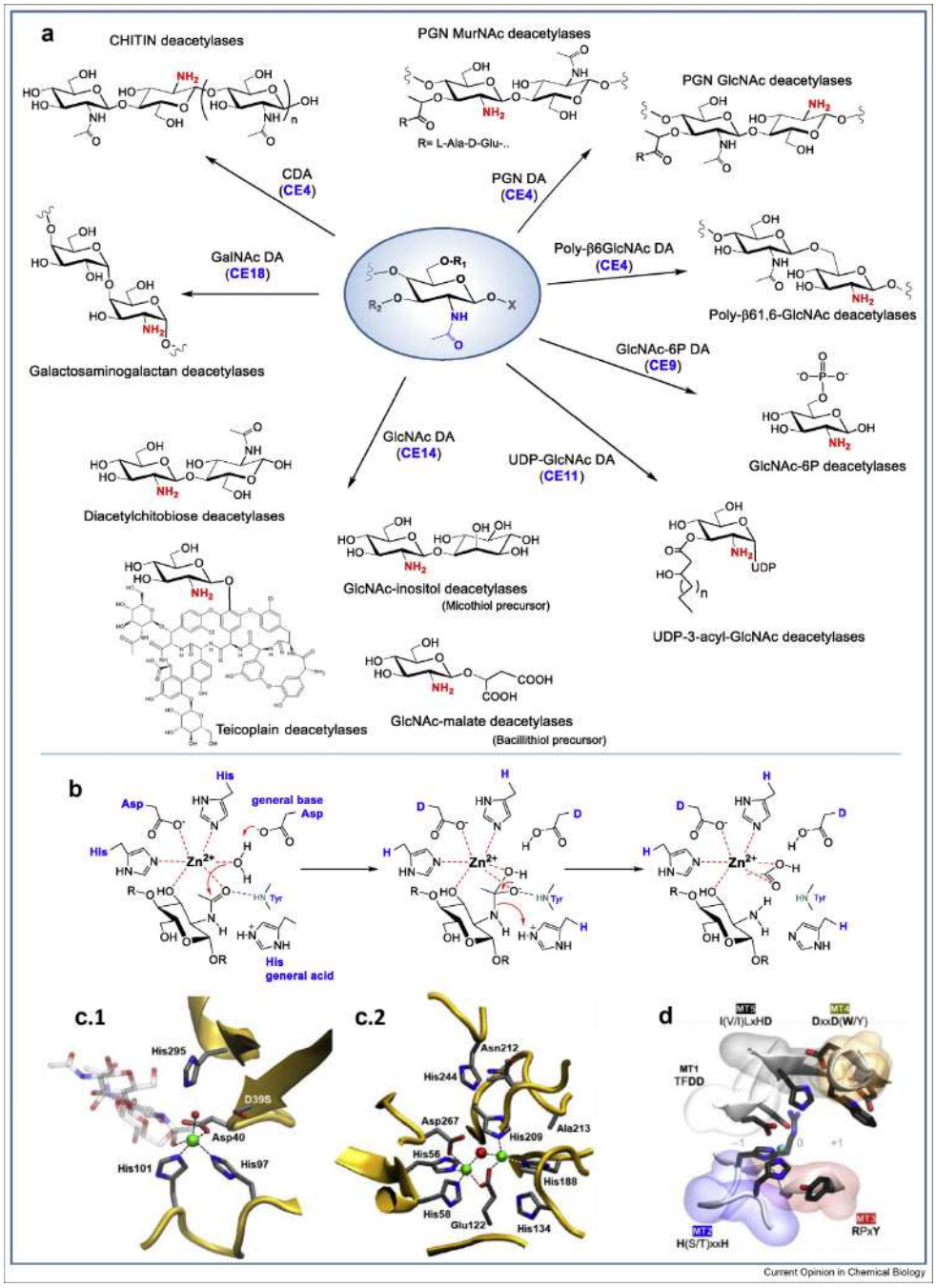
Carbohydrate deacetylases belong to the carbohydrate esterase (CE) class of enzymes in the Carbohydrate Active Enzymes classification (www.cazy.org [2]). An overview of the different CE families was recently reported in Ref. [3]. This article will focus on de-*N*-acetylases acting on different carbohydrate substrates, recent developments with regard to structural and functional properties, and carbohydrate deacetylases that may serve as therapeutic targets.

Carbohydrate esterase families, de-*N*-acetylation mechanisms, and structures

De-*N*-acetylases are grouped into families CE4, 9, 11, 14, and 18, plus few enzymes yet unclassified. De-*N*-acetylation of structural polysaccharides, chitin in fungi and insects, peptidoglycan in bacteria, and poly-β-1,6-*N*-acetylglucosamine in bacterial biofilms, is catalyzed by CE4 enzymes that share a $(\beta/\alpha)_7$-barrel structure, whereas deacetylation of small oligosaccharides and glycoconjugates is catalyzed by enzymes belonging to different CE families: diacetylchitobiose and GlcNAc-conjugates by CE14 enzymes with α/β-fold structure, N-acetylglucosamine-6-phosphate by CE9 enzymes with $(\beta/\alpha)_8$-barrel structure, and UDP-3-*O*-acyl-*N*-acetylglucosamine by CE11 enzymes with two-layer-sandwich fold (Figure 1a). Just recently, a new family CE18 has been created after characterization of a *N*-acetylgalactosamine deacetylase acting on bacterial exopolysaccharides [4].

Despite different folds, carbohydrate de-*N*-acetylases share a common metal-assisted acid/base mechanism. It was first proposed for the peptidoglycan GlcNAc

Figure 1


a
CHITIN deacetylases
PGN MurNAc deacetylases
R= L-Ala-D-Glu-..
PGN GlcNAc deacetylases
CDA
(CE4)
PGN DA
(CE4)
GalNAc DA
(CE18)
Poly-β6GlcNAc DA
(CE4)
Poly-β61,6-GlcNAc deacetylases
Galactosaminogalactan deacetylases
GlcNAc-6P DA
(CE9)
GlcNAc-6P deacetylases
GlcNAc DA
(CE14)
UDP-GlcNAc DA
(CE11)
Diacetylchitobiose deacetylases
GlcNAc-inositol deacetylases
(Micothiol precursor)
UDP-3-acyl-GlcNAc deacetylases
Teicoplain deacetylases
GlcNAc-malate deacetylases
(Bacillithiol precursor)
b
His
Asp
general base
Asp
His
general acid
c.1
His295
D39S
Asp40
His101
His97
c.2
Asn212
His244
Asp267
His209
Ala213
His56
His188
His58
Glu122
His134
d
MT3
I(V/I)LxHD
MT4
DxxD(W/Y)
MT1
TFDD
MT2
H(S/T)xxH
MT5
RPxY
Current Opinion in Chemical Biology

deacetylase *Sp*PgdA when its X-ray structure was solved in 2005 [5]. The consensus mechanism is shown in Figure 1b. The metal cation (commonly $Zn^{2þ}$, $Co^{2þ}$, $Fe^{2þ}$, $Mn^{2þ}$, or $Cd^{2þ}$) is coordinated by the conserved motif Asp-His-His. The first 3D structure of a chitin deacetylase (CE4, *Vc*CDA) in complex with substrates showed that a sugar hydroxyl group also participates in metal coordination [6]. Most CE families with de-*N*-acetylase activity contain a single metal cation, except family CE9 that contains both mononuclear and binuclear metal binding sites [7] (Figure 1c).

Deacetylases acting on structural polysaccharides: chitin, peptidoglycan, and poly-b1,6-GlcNAc deacetylases (CE4)

The CE4 family is composed of not only chitin (CDA, EC3.5.1.41) and chitooligosaccharide deacetylases (COD, EC3.5.1.-), peptidoglycan *N*-acetylglucosamine deacetylases (PGN GlcNAc DA, EC3.5.1.104), peptidoglycan *N*-acetylmuramic acid deacetylases (PGN MurNAc DA, EC3.5.1.-), and poly-b-1,6-*N*-acetylglucosamine deacetylases (PNAG DA, EC3.5.1.-) but also some acetylxylan esterases that are de-*O*-acetylases (AXE, EC3.1.1.72).

CE4 enzymes share a conserved region known as NodB homology domain (for NodB COD involved in Nod factors biosynthesis [8]) of ≥150 aa long and defined by five conserved motifs, which were proposed after the 3D structure resolution of *Sp*PgdA in 2005 [5] and *Cl*CDA in 2006 [9]. These motifs form the active site and, as new 3D structures are released, their description is refined based on more extensive sequence and structural alignments (Figure 1d) [10].

Chitin deacetylases

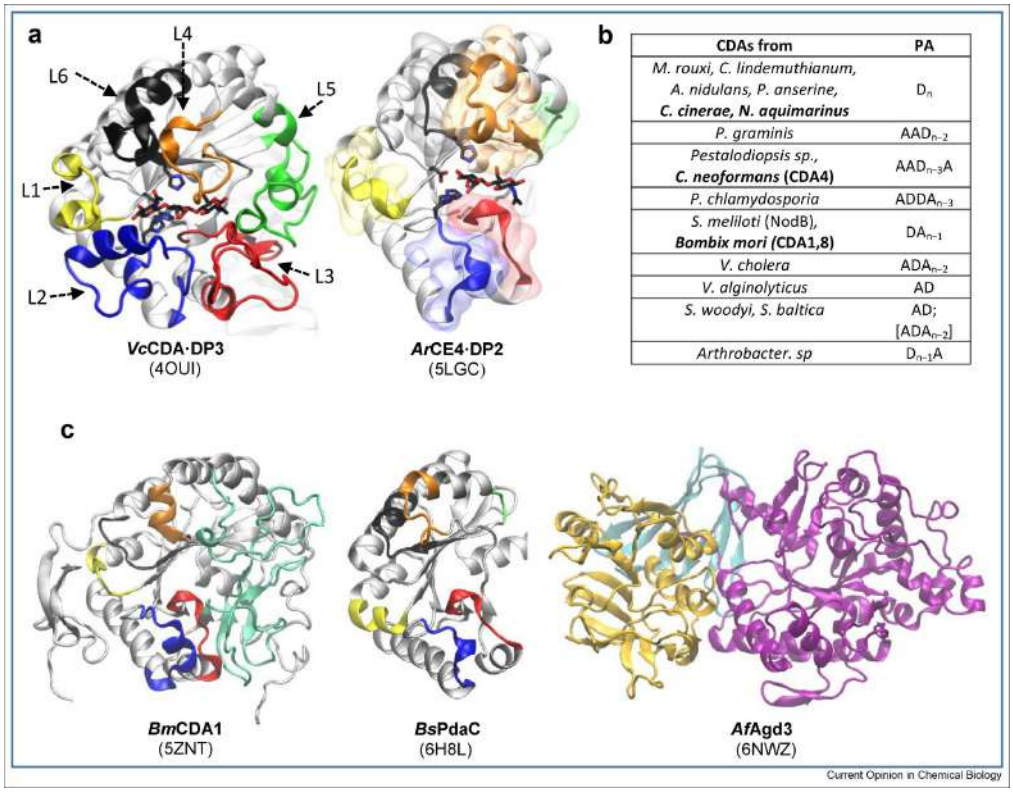
Many CDAs and CODs have been characterized in archaea, marine bacteria, fungi, and insects (recent reviews [11,12]) and localized in different cellular compartments. Periplasmic fungal CDAs are generally tightly coupled to a chitin synthase to rapidly deacetylate newly synthesized chitin to chitosan. Extracellular fungal CDAs are secreted to alter the physicochemical properties of the cell wall, either for protection against exogenous chitinases and evasion of host's immune defenses or to initiate sporulation or autolysis. In bacteria, CDAs are either intracellular, as those involved in Nod factors biosynthesis in *Rhizobium* species, or extracellular, as those involved in the catabolism of chitin in marine bacteria [11]. In addition to numerous studies addressed to decipher the role CDAs play in the biology of their natural organisms, there is a current interest in the discovery and biochemical characterization of novel CDAs to use them as biocatalysts for the production of partially deacetylated chitooligosaccharides (paCOS) as bioactive molecules in different application fields, or to inhibit them since they are potential targets against pathogenic microorganisms [13].

CDAs exhibit diverse deacetylation patterns resulting from different action mechanisms (multiple-attack, multiple- or single-chain modes of action) and different pattern recognition of their substrates. A fundamental question is how these enzymes define the pattern of deacetylation. The X-ray structure of *Vc*CDA$substrate complexes (PDB 4NZ1, 4OUI) highlighted the role of a series of loops that shape the binding cleft in substrate binding and led to the proposal of the "subsite capping model" (loop 1 to 6, Figure 2a) [6]. The deacetylation patterns of characterized CDAs until 2018 were reviewed in Ref. [11]. Since then, new CDAs are being identified to unravel their biological functions and specificity on COS substrates, thus expanding the toolbox of biocatalysts for the preparation of paCOS with defined patterns of deacetylation (Figure 2b). Most CDAs also act in reverse mode following the same acetylation regioselectivity of $(GlcN)_n$ substrates than in the de-*N*-acetylation reaction. Significantly, the combination of different CDAs has allowed the enzymatic production of all partially acetylated chitosan tetramers [14].

Recently identified fungal CDAs include *Coprinopsis cinerea* Cda1-3 [15,16] and *Aspergillus fumigatus* Cod4 [17]. *Cryptococcus neoformans* contains four CDAs; three of them are GPI-anchored to the membrane acting on cell wall chitin during chitosan biosynthesis, but the fourth (*Cn*CDA4) is a secreted enzyme with exceptional specificity for a GlcN residue at subsite −1. It has been proposed that it may further deacetylate exposed cell wall chitosans or released oligomers by human chitinases, thus making the fungus less susceptible to host immunosurveillance [18]. CDA1-3 from the hypervirulent *Cryptococcus gattii* have shown that *C. gattii* has evolved an alternate regulation of chitosan biosynthesis during mammalian infection compared with *C. neoformans* [19]. CDAs have been identified for the first time in diatoms, *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*, which might be acquired by horizontal gene transfer from bacteria and fungi, respectively [20]. Several bacterial CDAs have been recently identified from *Actinobacter schindler* [21],

Carbohydrate de-N-acetylase families, mechanism and active sites. (a) Carbohydrate de-*N*-acetylases showing the substrate specificities of each CE family. NH_2- in blue is the deacetylated−NAc moiety in the corresponding product. (b) Metal-assisted acid/base mechanism of carbohydrate de-*N*-acetylases. Conserved Asp-His-His metal-binding triad, general acid (conserved His) and general base (conserved Asp). (c) Metal binding sites in de-*N*-acetylases, C.1) mononuclear metal coordination (*Vc*CDA, PDB 4OUI, mutant D39S at the general base) and C.2) binuclear metal coordination (*Ms*NagA, PDB 6FV3) found in some CE9 enzymes. (d) Conserved motifs MT1−MT5 defining the active site of CE4 enzymes. Motif 1 (TFDD) includes the general base aspartate (first D) and the metal-binding aspartate (second D) that, together with two His of motif 2 (H(S/T)xxH), form the characteristic Asp-His-His triad of CE4 enzymes.

Figure 1



3D structures of representative and novel chitin deacetylases (CDAs). (a) X-ray structures of *Vibrio cholera* CDA (*Vc*CDA, long loops, deep and narrow cleft) and *Arthrobacter* CDA (*Ar*CE4, short loops, open and shallow cleft) in complex with chitooligosaccharide substrates. Loops 1 to 6 (color-labelled) illustrate their function according to the "subsite capping model" [6,11]. (b) Reported CDAs with known deacetylation pattern on chitooligosaccharides. PA: pattern of acetylation of end products from COS. A: GlcNAc; D: GlcN. In bold, characterized enzymes since 2018. (c) Relevant new 3D structures of CE enzymes since 2018 (left) insect CDA, loops 1 to 6 characteristic of CE4 enzymes labelled as in A), new loop insertion labelled in magenta (middle) *Bs*PdaC has dual MurNAc DA activity on peptidoglycan and GlcNAc DA activity on chitooligosaccharides (right) *Af*Agd3, GalNAc exopolysaccharide deacetylase, N-terminus domain (yellow) CBM87, C-terminus (pale green) founding member of family CE18.

Bacillus licheniformis [22] and *Rhodococcus equi* [23] (not yet characterized for their deacetylation pattern), whereas the *Nitratireductor aquimarinus* CDA was shown to act on COS with DP $\geq$ 4 and fully deacetylate DP4 [24]. Insect CDAs is an emerging field of study to unravel the biology of cuticle formation and peritrophic membrane modifications [25] and represent promising targets for insecticide development. Knock-out experiments have identified specific functions of CDA-encoding genes [26e28]. Different CDAs have been expressed and characterized from *Bombyx mori* [29,30], reporting the first 3D structures of insect CDAs (*Bm*CDA1, 5ZNT; *Bm*CDA8, 5ZNS) [29]. Remarkably, they have two unique loop regions that contribute to the distinct architecture with a much longer and wider substrate-binding cleft than the microbial CDAs (Figure 2c).

Peptidoglycan deacetylases

Remodeling of bacterial cell wall peptidoglycan (PGN) has critical functions in PGN maturation, elongation, septation, turnover, and recycling. Modifications of PGN represent an important strategy for pathogenic bacteria to evade innate immunity of the host and control autolysins [31]. Common modifications of the glycan backbone are de-*N*-acetylation of GlcNAc and/or MurNAc residues, which alter the properties of the cell wall and play an important role in pathogenesis. These and other modifications to the PGN glycan protect it from lysis by both exogenous host-defense agents, such

as the lysozymes of host innate immune systems, and the endogenous autolysins involved in PGN metabolism [32,33].

PGN GlcNAc and MurNAc DAs have mutually exclusive specificities, but currently no sequence or structural signatures can be assigned to each enzyme class to predict specificity and function. PGN GlcNAc DAs (PgdAs) are better characterized and mostly reported in Gram-positive bacteria. *Sp*PgdA (2C1G) is the prototype enzyme of the family [5], which is also active on COS. As shown by many knock-out studies, *pgdA*-defective mutants are more sensitive to lysozyme and less virulent. Recent advances in the functional characterization of PgdAs include *Lactococcus lactis* [34] and *Listeria monocytogenes* [35]. After *Sp*PgdA, other X-ray structures have been released, the *S. mutants* PgdA (2W3Z, 2008), which was found to be inactive on PGN, and its natural substrate is yet unidentified [36], the *Eubacterium rectale* Pgd (5JMU, 2016, not biochemically characterized) and the *Bacillus cereus* PgdA (4L1G, 2017), which has shown that the proline residue in the conserved motif 3 undergoes an a-carbon hydroxylation that promotes active-site maturation [37]. Whether this post-translational modification is general among CE4 enzymes remains to be analyzed. PGN DAs from *B. cereus* and *Bacillus anthracis* are extensively studied [38] and targets for the design of inhibitors as antimicrobials [39].

Less studied are PGN MurNAc DAs (PdaAs). Only two X-ray structures have been solved, those of *Bacillus subtilis* PdaA (1W17) and *B. anthracis* (2J13), but only *Bs*Pda has been biochemically characterized [40]. *Bs*PdaA is the current model for a MurNAc deacetylase involved in spore cortex PGN formation during sporulation, where MurNAc residues are converted to muramic d-lactam. *Bs*PdaA deacetylates the MurNAc residues after stem peptide removal by L-alanine amidase CwlD [41]. Other putative MurNAc deacetylases identified by studies with deficient mutants include *Bs*PdaB [42] and the two recently identified putative MurNAc DAs, PdaA1 and PdaA2 from *Clostridium difficile* [43]. *Bs*PdaC is a novel type of MurNAc DA discovered in 2012 [44], and its specificity and 3D structure (6H8L, Figure 2c) were recently characterized [45]. *Bs*PdaC is unique in that it is a MurNAc deacetylase acting on intact PGN, it is not involved in sporulation, and strikingly, it is also active on COS, an activity that was thought to be restricted to GlcNAc DAs. *Bs*PdaC maintains the conserved metal-binding Asp-His-His triad as opposed to canonical MurNAc DAs that have an Asn substitution not involved in metal coordination. It has been proposed to be the first representative of a new subclass of PGN deacetylase with dual specificity but yet unknown function [45].

Poly-b-1,6-*N*-acetylglucosamine deacetylases

PNAG is a major biofilm component of many pathogenic bacteria [46]. Enzymes involved in biofilm formation are thus targets for the design of antibiofilm agents and antimicrobials [47,48]. PNAG and its partially de-*N*-acetylated derivative were originally isolated from the biofilms of Gram-positive *Staphylococcus epidermis* and found to be also widespread in Gram-negative bacteria. PNAG biosynthesis is mediated by the *pga*ABCD operon in *Escherichia coli* [49] and other Gram-negative bacteria (or the homologous gene clusters *ica*ADBC in *Staphylococcus* or *hms*HFRS in *Yersinia* species), where PgaC and PgaD form the synthase complex, PgaA is an outer-membrane porin that mediates PNAG export, and PgaB contains a CE4 deacetylase domain [50]. The first PgaB enzyme with solved 3D structure was the *E. coli* PgaB (3VUS) that deacetylates up to 20% GlcNAc residues of PNAG and facilitates its transport to the extracellular matrix. In other bacteria (i.e. *Bordetella bronchiseptica*) deacetylation does not appear to be required for transport, but it is required for robust three-dimensional biofilm formation [51]. Three other X-ray structures have been reported, *Ammonifex degensii* (*Ad*IcaB, 4WCJ), *B. bronchiseptica* (*Bb*PgaB, formerly BpsB, 5BU6), and *Aggregatibacter actinomycetemcomitans* (*Aa*PgaB, 4U10) [52] deacetylases. Whereas *Ad*IcaB is a monodomain protein, the others contain a second C-t domain that is critical for PNAG export but had unclear function. It has been recently shown that the C-terminal domain of *Bb*PgaB and *Ec*PgaB functions as glycoside hydrolases acting on partially deacetylated PNAG [53]. The solved 3D structure of the hydrolase domain (6AU1 and 4F9D, respectively) assigned a new GH153 family in the CAZy database containing poly-b-1,6-D-glucosamine hydrolases most of them in multidomain PgaB-like proteins.

The CE4 domain of PNAG deacetylases is a circular permutation of the canonical CE4 sequence of chitin and peptidoglycan deacetylases with some variations in the positioning of active site residues [10]. Although the conserved motifs MT1-5 are a signature of the CE4 family, PNAG DAs reveal different spatial arrangements, more significantly the MT4 (DxxD (W/Y), which has a water molecule in place of the conserved Asp residue that interacts (and presumably activates) the catalytic acid His in other CE4 enzymes [10]. It has been hypothesized that the inability of this water molecule to activate the catalytic acid is consistent with the lower specific activity of these enzymes compared to the other CE4 subfamilies [54].

GlcNAc deacetylases on small oligosaccharides and glycoconjugates (CE14)

The LmbE protein family (Pfam02585) is composed of metallohydrolases catalyzing the deacetylation of

GlcNAc of small oligosaccharides and glycoconjugates. This family is defined by a common PIG-L domain that stands for one of the first characterized members, the *N*-acetyl-D-glucosaminylphosphatidylinositol deacetylase (GlcNAc-PI or PIG-L) involved in the biosynthesis of GPI membrane anchor precursors [55]. CAZy classifies some LmbE members in family CE14 (antibiotic and GlcNAc-PI de-*N*-acetylases are not yet classified in CAZy).

Archaeal diacetylchitobiose DAs are involved in the chitinolitic machinery in archeons that differs from the bacterial and eukaryotic pathways. The *Pyrococcus furiosus* (3WL4) and *Pyrococcus horikoshii* (3WE7) enzymes deacetylate the GlcNAc unit at the nonreducing end [56]. This specificity differs from COD enzymes in marine bacteria (CE4 enzymes), which deacetylate diacetylchitobiose on the reducing end [57]. An unrelated GlcNAc deacetylase has been recently identified in *Sulfolobus solfataricus*, active on single GlcNAc and its aryl-derivatives, showing deacetylation of both the a- and b-anomers. Its biological function is unknown but hypothesized that may act in concerted action with a yet unidentified a-N-acetylglucosaminidase [58]. A bacterial CE14 deacetylase from *B. cereus* (*Bc*ZBP, BC1534, 2IXD) also deacetylates $(GlcNAc)_{1\text{-}3}$ but has a broad substrate specificity [59,60].

Another specificity in CE14 enzymes is found in *Mycobacterium tuberculosis* MshB, a N-acetyl-1-D-inosityl-2-amino-2-deoxy-D-glucopyranoside deacetylase (GlcNAc-inositol DA, 1Q74) that is involved in the mycothiol (MSH) biosynthetic pathway [61]. A recently crystalized enzyme similar to MshB is that of *B. subtilis* BshB (2PT2), a deacetylase involved in bacillithiol (BSH) biosynthesis [62]. In bacteria that lack the tripeptide glutathione (GSH), bacillithiol in *Bacillus* and related Firmicutes, and mycothiol in Actinobacteria are the major low-molecular-weight thiol antioxidants that play key role in detoxification of electrophiles and an important role in resistance to antibiotics, thus the enzymes involved in their biosynthetic pathways are antimicrobial targets [63,64]. In addition to their function as primary reducing agents, MSH and BSH are also substrates for transferase enzymes in the detoxification of electrophilic species: the micothiol-conjugate amidase (Mca), which allows GlcN-Ins to be recycled back to the MSH biosynthesis pathway, and bacillithiol-conjugate amidase (Bca), which recycles GlcN-malate for BSH biosynthesis [63,65].

Other LmbE members include GlcNAc deacetylases involved in the biosynthesis of the antibiotic lipoglycopeptides teicoplanin and A40926. During biosynthesis after glycosyltransfer to the peptide aglycone, the GlcNAc-pseudopeptide is deacetylated before long chain hydrocarbon attachment (recently reviewed in Ref. [66]). The crystal structures of teicoplanin deacetylase Ost2 (2XAD) and A40926 deacetylase (3DFI) show a unique, structurally mobile capping lid that likely serves as a determinant of substrate specificity. Remarkably, manipulation of the biosynthetic pathway by regioselective engineered enzymes, a Orf2 deacetylase and a Orf1 acyltransferase, produced new analogues with 2e3 orders of magnitude better antimicrobial activity against resistant *Staphilococcus aureus* than teicoplanin and vancomycin [67].

GlcNAc-6-phosphate deacetylases (CE9)

N-acetylglucosamine-6-phosphate deacetylases (GlcNAc-6-P DA, EC 3.5.1.25) belong to family CE9 and are mononuclear or binuclear metal-dependent amidases with variable oligomeric organizations. This activity produces GlcN-6-P that is the precursor of peptidoglycan and teichoic acid biosynthesis. The first 3D structure was release in 2002 (*Thermotoga maritima,* 1O12) and the first characterized enzyme member was NagA from *B. subtilis* (2VHL) [68]. It contains a binuclear metal binding site with high similarity to the urease superfamily. Other structures were solved (*E. coli* NagA, 2P50 and the *Vibrio cholera* homolog, 3EGJ), showing a mononuclear metal-binding site architecture. Both metal architectures seem to be common, since new recently characterized GlcNAc-6-P DAs, *Mycobacterium smegmatis* (*Ms*NagA) (6FV3) [7] and *Pasteurella multocida* (6JKU in 2020, [69]) contain binuclear and mononuclear Zn^{2+} binding sites, respectively.

UDP-3-*O*-acyl-GlcNAc deacetylases (LpxC) (CE11)

LpxC (EC 3.5.1.108) catalyzes the first committed step in Lipid A biosynthesis, deacetylating UDP-(3-*O*-(*R*-3-hydroxymyristoyl))-*N*-acetylglucosamine. The first 3D structure was solved in 2003 for the *Aquifex aeolicus* enzyme (1P42) [70]. Later, the enzymes from *E. coli*, *Pseudomonas aeruginosa*, and *Yersinia enterocolitica* have been solved (http://www.cazy.org/CE11_structure.html). Since the enzyme is essential for Lipid A biosynthesis, it is a validated target for the design of inhibitors as antimicrobials (LpxC as a drug target reviewed in Ref. [71]). More recently, a number of medicinal chemistry investigations have been reported and numerous X-ray structures of enzymeeinhibitor complexes have been deposited in the PDB. Most inhibitors share a Zn^{2+}-chelating hydroxamate moiety and a lipophilic chain binding to the hydrophobic tunnel of the enzyme's binding site. The concept of "cryptic inhibitor envelope," in which dynamic ligand conformations delineate an inhibitor envelop, led to the development of potent inhibitors [72]. Recent highlights include proline-derived 3,4-dihydroxypyrrolidine hydroxamates [73], optimization of the first LpxC inhibitor (Achaogen's ACHN-975) that reached human clinical testing [74], Fe^{2+}-activated prodrug design to cage known hydroxamic acid LpxC inhibitors against

P. aeruginosa infections [75], triazole ring introduction into reported hydroxamic inhibitors [76], and a scaffold hopping approach and hit-to-lead optimization that resulted in the discovery of novel chemical series from which pyrrolo-imidazolones were the most promising in a murine infection model but yet with poor pharmacokinetic profiles [77]. No successful antibiotics targeting LpxC are yet available, and numerous programs are ongoing.

GalNAc polysaccharide deacetylases (CE18)

Galactoaminogalactan (GAG), an α1,4-linked linear polysaccharide of Gal and GalNAc units, is essential for *A. fumigatus* and *A. oryzae* biofilm formation and a key virulence factor. Similarly to biofilm PNAG exopolysaccharide (section 3.3), GAG is synthesized by a gene cluster that contains a polysaccharide deacetylase (Agd3) whose deletion impairs biofilm formation and renders less virulent strains [78,79]. Recently, *Af*Agd3 was characterized and its 3D structure determined [4]. The C-terminal domain is a metal-dependent deacetylase with low homology to CE4 enzymes, and it is the founding member of a new CE18 family (GalNAc DAs), whereas the N-t domain comprises a new carbohydrate binding domain CBM87. Bioinformatics and phylogenetic analysis found that Agd3-like enzymes are present in both fungal and bacterial kingdoms. Significantly, they have four conserved motifs CM1-4 that resemble MT1-5 in CE4 enzymes.

In conclusion, the discovery and characterization of de-*N*-acetylases either acting on structural polysaccharides or glycoconjugates in the hexosamine metabolism is an active field with significant potential in biomedicine and biotechnology since they are targets for the design of antimicrobials, and potential biocatalysts for the production of defined glycostructures.

Funding source

Supported by grant PID2019-104350RB-I00 from Ministry of Science and Innovation (MICINN), Spain. S.P. acknowledges a predoctoral contract from MICINN.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. Bürger M, Chory J: Structural and chemical biology of deacetylases for carbohydrates, proteins, small molecules and histones. *Commun Biol* 2018, 1:1–11.

2. Lombard V, Golaconda-Ramulu H, Drula E, Coutinho PM, Henrissat B: The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 2014, 42:490–495.

3. Nakamura AM, Nascimento AS, Polikarpov I: Structural diversity of carbohydrate esterases. *Biotechnol Res Innov* 2017, 1:35–51.

4. ** Bamford NC, Le Mauff F, Van Loon JC, Ostapska H, Snarr BD, Zhang Y, Kitova EN, Klassen JS, Codée JDC, Sheppard DC, *et al.*: Structural and biochemical characterization of the exopolysaccharide deacetylase Agd3 required for Aspergillus fumigatus biofilm formation. *Nat Commun* 2020, 11: 2450.
The article describes the characterization and 3D structure determination by X-ray crystallography of a galactosaminogalactan deacetylase in bacterial biofilm formation. It contains a novel N-t deacetylase domain that is the founding member a new CAZy family CE18 and a C-t domain with a new carbohydrate-binding module CBM87.

5. * Blair DE, Schuttelkopf AW, MacRae JI, van Aalten DMF: Structure and metal-dependent mechanism of peptidoglycan deacetylase, a streptococcal virulence factor. *Proc Natl Acad Sci* 2005, 102:15429–15434.
Reference paper on the structural fold and mechanism of family CE4 as metal-dependent deacetylases.

6. * Andrés E, Albesa-Jové D, Biarnés X, Moerschbacher BM, Guerin ME, Planas A: Structural basis of chitin oligosaccharide deacetylation. *Angew Chem Int Ed* 2014, 53:6882–6887.
Paper reporting the first 3D structure of competent enzyme–substrate complexes of a CE4 enzyme elucidating the coordination of the substrate with the metal cation and proposing the "subsite capping model" as a determinant of the deacetylation pattern by chitin deacetylases.

7. * Ahangar MS, Furze CM, Guy CS, Cooper C, Maskew KS, Graham B, Cameron AD, Fullam XE: Structural and functional determination of homologs of the Mycobacterium tuberculosis N-acetylglucosamine-6-phosphate deacetylase (NagA). *J Biol Chem* 2018, 293:9770–9783.
3D structure of NagA from *M. smegmatis* with a binuclear metal binding site characteristic but restricted to CE9 de-*N*-acetylases.

8. John M, Rohrig H, Schmidt J, Wieneke U, Schell J: Rhizobium NodB protein involved in nodulation signal synthesis is a chitooligosaccharide deacetylase. *Proc Natl Acad Sci* 1993, 90:625–629.

9. Blair DE, Hekmat O, Schüttelkopf AW, Shrestha B, Tokuyasu K, Withers SG, van Aalten DM: Structure and mechanism of chitin deacetylase from the fungal pathogen Colletotrichium lindemuthianum. *Biochemistry* 2006, 45:9416–9426.

10. Aragunde-Pazos H, Biarnés X, Planas A: Substrate recognition and specificity of chitin deacetylases and related family 4 carbohydrate esterases. *Int J Mol Sci* 2018, 19:412.

11. * Grifoll-Romero L, Pascual S, Aragunde H, Biarnés X, Planas A: Chitin deacetylases: structures, specificities, and biotech applications. *Polymers* 2018, 10:352.
Comprehensive review paper on chitin deacetylases compiling and discussing characterized CDAs until 2018 with regard to their deacetylation pattern on chitooligosaccharides and the elaboration of the "subsite capping model" on this CE4 family.

12. Schmitz Auza, Koberidze Rasche, Fischer Bortesi: Conversion of chitin to defined chitosan oligomers: current status and future prospects. *Mar Drugs* 2019, 17:452.

13. * Naqvi S, Moerschbacher BM: The cell factory approach toward biotechnological production of high-value chitosan oligomers and their derivatives: an update. *Crit Rev Biotechnol* 2017, 37:11–25.
Review article on production strategies of chitosan oligosaccharides emphasizing the need of structurally defined partially deacetylated oligomers to unravel their biological functions and develop new applications.

14. ** Hembach L, Cord-Landwehr S, Moerschbacher BM: Enzymatic production of all fourteen partially acetylated chitosan tetramers using different chitin deacetylases acting in forward or reverse mode. *Sci Rep* 2017, 7:17692.
This article confirms that most CDAs are able to catalyze the reverse reaction by acetylating glucosamine oligomers with the same regioselectivity than the natural deacetylation reaction on COS. Remarkably, by combining different deacetylases with diverse specificities the full

set of partially deacetylated chitosan tetrasaccharides are produced and purified as single products.

15. Wang Y, Niu X, Guo X, Yu H, Liu Z, Zhang Z, Yuan S: Heterologous expression, characterization and possible functions of the chitin deacetylases, Cda1 and Cda2, from mushroom Coprinopsis cinerea. *Glycobiology* 2018, 28:318–332.

16. Bai Y, Wang Y, Liu X, Zhao J, Kang LQ, Liu Z, Yuan S: Heterologous expression and characterization of a novel chitin deacetylase, CDA3, from the mushroom Coprinopsis cinerea. *Int J Biol Macromol* 2020, 150:536–545.

17. Xie M, Zhao X, Yang L, Jin C: Chitin deacetylases Cod4 and Cod7 are involved in polar growth of Aspergillus fumigatus. *Microbiologyopen* 2020, 9:1–14.

18. ** Hembach L, Bonin M, Gorzelanny C, Moerschbacher BM: Unique subsite specificity and potential natural function of a chitosan deacetylase from the human pathogen Cryptococcus neoformans. *Proc Natl Acad Sci U S A* 2020, 117:3551–3559.
The article reports the characterization of the secreted CDA4 from this pathogenic fungus which is more active on partially deacetylated chitosan oligomers and involved in virulence. It is proposed that further deacetylation of cell wall chitosan or releases oligomers by human chitinases is a strategy to make the fungus less susceptible to host immunity.

19. Lam WC, Upadhya R, Specht CA, Ragsdale AE, Hole CR, Levitz SM, Lodge JK: Chitosan biosynthesis and virulence in the human fungal pathogen Cryptococcus gattii. *mSphere* 2019, 4. e00644-19.

20. Shao Z, Thomas Y, Hembach L, Xing X, Duan D, Moerschbacher BM, Bulone V, Tirichine L, Bowler C: Comparative characterization of putative chitin deacetylases from Phaeodactylum tricornutum and Thalassiosira pseudonana highlights the potential for distinct chitin-based metabolic processes in diatoms. *New Phytol* 2019, 221:1890–1905.

21. Ye W, Ma H, Liu L, Yu J, Lai J, Fang Y, Fan Y: Biocatalyzed route for the preparation of surface-deacetylated chitin nanofibers. *Green Chem* 2019, 21:3143–3151.

22. Bhat P, Pawaskar GM, Raval R, Cord-Landwehr S, Moerschbacher B, Raval K: Expression of Bacillus licheniformis chitin deacetylase in E. coli pLysS: sustainable production, purification and characterisation. *Int J Biol Macromol* 2019, 131:1008–1013.

23. Ma Q, Gao X, Bi X, Tu L, Xia M, Shen Y, Wang M: Isolation, characterisation, and genome sequencing of Rhodococcus equi: a novel strain producing chitin deacetylase. *Sci Rep* 2020, 10:4329.

24. Chai J, Hang J, Zhang C, Yang J, Wang S, Liu S, Fang Y: Purification and characterization of chitin deacetylase active on insoluble chitin from Nitratireductor aquimarinus MCDA3-3. *Int J Biol Macromol* 2020, 152:922–929.

25. Liu X, Cooper AMW, Zhang J, Zhu KY: Biosynthesis, modifications and degradation of chitin in the formation and turnover of peritrophic matrix in insects. *J Insect Physiol* 2019, 114:109–115.

26. Yu RR, Liu WM, Zhao XM, Zhang M, Li DQ, Zuber R, Ma EB, Zhu KY, Moussian B, Zhang JZ: LmCDA1 organizes the cuticle by chitin deacetylation in Locusta migratoria. *Insect Mol Biol* 2019, 28:301–312.

27. Zhang M, Ji Y, Zhang X, Ma P, Wang Y, Moussian B, Zhang J: The putative chitin deacetylases Serpentine and Vermiform have non-redundant functions during Drosophila wing development. *Insect Biochem Mol Biol* 2019, 110:128–135.

28. Liu Y, Yang J, Yao L, Li S, Chen Y, Yang H, Fan D: Chitin deacetylase: a potential target for Mythimna separata (Walker) control. *Arch Insect Biochem Physiol* 2020, https://doi.org/10.1002/arch.21666.

29. ** Liu L, Zhou Y, Qu M, Qiu Y, Guo X, Zhang Y, Liu T, Yang J, Yang Q: Structural and biochemical insights into the catalytic mechanisms of two insect chitin deacetylases of the carbohydrate esterase 4 family. *J Biol Chem* 2019, 294:5774–5783.
This article describes the first X-ray structures of insect CDAs, family CE4 deacetylases but with two unique loop insertions that define a much longer and wider substrate binding cleft than microbial CDAs. These new structures significantly contribute to further structure–function studies and to evaluate these CDAs as targets for insecticide development.

30. Liu L, Qu M, Liu T, Chen Q, Guo X, Yang J, Yang Q: Biochemical characterization of three midgut chitin deacetylases of the Lepidopteran insect Bombyx mori. *J Insect Physiol* 2019, 113:42–48.

31. Wolf AJ, Underhill DM: Peptidoglycan recognition by the innate immune system. *Nat Rev Immunol* 2018, 18:243–254.

32. Davies C, Dillard JP, Cava F, Yadav AK, Espaillat A: Bacterial strategies to preserve cell wall integrity against environmental threats. *Front Microbiol* 2018, 9:2064.

33. Clarke AJ: Chemical biology of peptidoglycan acetylation and deacetylation. *Bioorg Chem* 2014, 54:44–50.

34. Cao L, Liang D, Hao P, Song Q, Xue E, Caiyin Q, Cheng Z, Qiao J: The increase of O-acetylation and N-deacetylation in cell wall promotes acid resistance and nisin production through improving cell wall integrity in Lactococcus lactis. *J Ind Microbiol Biotechnol* 2018, 45:813–825.

35. Rismondo J, Wamp S, Aldridge C, Vollmer W, Halbedel S: Stimulation of PgdA-dependent peptidoglycan N-deacetylation by GpsB-PBP A1 in Listeria monocytogenes. *Mol Microbiol* 2018, 107:472–487.

36. Deng MD, Urch JE, Ten Cate JM, Rao VA, Van Aalten DMF, Crielaard W: Streptococcus mutans SMU.623c codes for a functional, metal-dependent polysaccharide deacetylase that modulates interactions with salivary agglutinin. *J Bacteriol* 2009, 91:394–402.

37. ** Fadouloglou VE, Balomenou S, Aivaliotis M, Kotsifaki D, Arnaouteli S, Tomatsidou A, Efstathiou G, Kountourakis N, Miliara S, Griniezaki M, *et al.*: Unusual α-carbon hydroxylation of proline promotes active-site maturation. *J Am Chem Soc* 2017, 139:5330–5337.
It is observed that a conserved proline residue in the active site of *Bc*PgdA undergoes an α-carbon hydroxylation that promotes active-site maturation and enhances activity. Possibly neglected in other CE4 enzymes, whether it is a common post-translational modification in other PGN deacetylases and related CE4 enzymes remains to be analyzed.

38. * Andreou A, Giastas P, Christoforides E, Eliopoulos EE: Structural and evolutionary insights within the polysaccharide deacetylase gene family of bacillus anthracis and bacillus cereus. *Genes* 2018, 9:386.
Article that analyzes the diversity of peptidoglycan deacetylases in closely related pathogenic *B.anthracis* and *B. cereus* and also provides a review entry to a number of structural and biochemical studies of polysaccharide deacetylases in these organisms.

39. * Balomenou S, Koutsioulis D, Tomatsidou A, Tzanodaskalaki M, Petratos K, Bouriotis V: Polysaccharide deacetylases serve as new targets for the design of inhibitors against Bacillus anthracis and Bacillus cereus. *Bioorg Med Chem* 2018, 26:3845–3851.
Related to the previous paper (38), peptidoglycan deacetylases in these pathogenic organisms are evaluated as targets for the design of inhibitors. This is the beginning of future medicinal chemistry studies for the development of effective drug candidates.

40. Blair DE, Van Aalten DMF: Structures of Bacillus subtilis PdaA, a family 4 carbohydrate esterase, and a complex with N-acetyl-glucosamine. *FEBS Lett* 2004, 570:13–19.

41. Fukushima T, Kitajima T, Sekiguchi J: A polysaccharide deacetylase homologue, PdaA, in Bacillus subtilis acts as an N-acetylmuramic acid deacetylase in vitro. *J Bacteriol* 2005, 187:1287–1292.

42. Fukushima T, Tanabe T, Yamamoto H, Hosoya S, Sato T, Yoshikawa H, Sekiguchi J: Characterization of a polysaccharide deacetylase gene homologue (pdaB) on sporulation of Bacillus subtilis. *J Biochem* 2004, 136:283–291.

43. Coullon H, Rifflet A, Wheeler R, Janoir C, Boneca IG, Candela T: N-Deacetylases required for muramic-lactam production are involved in Clostridium difficile sporulation, germination, and heat resistance. *J Biol Chem* 2018, 293:18040–18054.

44. Kobayashi K, Sudiarta IP, Kodama T, Fukushima T, Ara K, Ozaki K, Sekiguchi J: Identification and characterization of a novel polysaccharide deacetylase C (PdaC) from Bacillus subtilis. *J Biol Chem* 2012, 287:9765–9776.

45. ** Grifoll-Romero L, Sainz-Polo MA, Albesa-Jové D, Guerin ME, Biarnés X, Planas A: Structure-function relationships underlying the dual N-acetylmuramic and N-acetylglucosamine specificities of the bacterial peptidoglycan deacetylase PdaC. *J Biol Chem* 2019, 294:19066–19080.
This article reports the 3D structure and mode of action of *Bs*PdaC, a unique PGN MurNAc deacetylase not involved in sporulation that it is active on intact peptidoglycan but also has GlcNAc DA activity on chitooligosaccharides. It is proposed as the first member of new subclass of peptidoglycan deacetylases with yet unknown functions.

46. Limoli DH, Jones CJ, Wozniak DJ: Bacterial extracellular polysaccharides in biofilm formation and function. *Microbiol Spectr* 2015, 3. MB-0011-2014.

47. * Karygianni L, Ren Z, Koo H, Thurnheer T: Biofilm matrixome: extracellular components in structured microbial communities. *Trends Microbiol* 2020, 28.
Comprehensive review on the composition and functions of extracellular polymeric substances in biofilms with updated bibliography and emphasizing that targeting biofilm-specific conditions could lead to precise and effective antibiofilm approaches.

48. Melander RJ, Basak AK, Melander C: Natural products as inspiration for the development of bacterial antibiofilm agents. *Nat Prod Rep* 2020, https://doi.org/10.1039/D0NP00022A.

49. Itoh Y, Rice JD, Goller C, Pannuri A, Taylor J, Meisner J, Beveridge TJ, Preston JF, Romeo T: Roles of pgaABCD genes in synthesis, modification, and export of the Escherichia coli biofilm adhesin poly-b-1,6-*N*-acetyl-D-glucosamine. *J Bacteriol* 2008, 190:3670–3680.

50. Low KE, Howell PL: Gram-negative synthase-dependent exopolysaccharide biosynthetic machines. *Curr Opin Struct Biol* 2018, 53:32–44.

51. Little DJ, Milek S, Bamford NC, Ganguly T, Difrancesco BR, Nitz M, Deora R, Howell PL: The protein BpsB is a poly-b-1,6-*N*-acetyl-D-glucosamine deacetylase required for biofilm formation in Bordetella bronchiseptica. *J Biol Chem* 2015, 290:22827–22840.

52. Parthiban C, Varudharasu D, Shanmugam M, Gopal P, Ragunath C, Thomas L, Nitz M, Ramasubbu N: Structural and functional analysis of de-*N*-acetylase PgaB from periodontopathogen Aggregatibacter actinomycetemcomitans. *Mol Oral Microbiol* 2017, 32:324–340.

53. * Little DJ, Pfoh R, Le Mauff F, Bamford NC, Notte C, Baker P, Guragain M, Robinson H, Pier GB, Nitz M, *et al.*: PgaB orthologues contain a glycoside hydrolase domain that cleaves deacetylated poly-b(1,6)-N-acetylglucosamine and can disrupt bacterial biofilms. *PLoS Pathog* 2018, 14, e1006998.
It was known that PNAG deacetylases have a C-terminus domain important for PNAG export but with unclear function. This article reports that this domain of PgaB enzymes is a glycosylhydrolase and the solved 3D structures assign a new GH153 family in the CAZy database classification.

54. Little DJ, Poloczek J, Whitney JC, Robinson H, Nitz M, Howell PL: The structure- and metal-dependent activity of Escherichia coli PgaB provides insight into the partial de-*N*-acetylation of poly-b-1,6-*N*-acetyl- D-glucosamine. *J Biol Chem* 2012, 287:31126–31137.

55. Viars S, Valentine J, Hernick M: Structure and function of the LmbE-like superfamily. *Biomolecules* 2014, 4:527–545.

56. Mine S, Niiyama M, Hashimoto W, Ikegami T, Koma D, Ohmoto T, Fukuda Y, Inoue T, Abe Y, Ueda T, *et al.*: Expression from engineered Escherichia coli chromosome and crystallographic study of archaeal N,N-diacetylchitobiose deacetylase. *FEBS J* 2014, 281:2584–2596.

57. Li X, Wang LX, Wang X, Roseman S: The chitin catabolic cascade in the marine bacterium Vibrio cholerae: characterization of a unique chitin oligosaccharide deacetylase. *Glycobiology* 2007, 17:1377–1387.

58. Iacono R, Strazzulli A, Maurelli L, Curci N, Casillo A, Corsaro MM, Moracci M, Cobucci-Ponzano B: GlcNAc De-*N*-Acetylase from the hyperthermophilic archaeon sulfolobus solfataricus. *Appl Environ Microbiol* 2019, 85. e01879-18.

59. Fadouloglou VE, Deli A, Glykos NM, Psylinakis E, Bouriotis V, Kokkinidis M: Crystal structure of the BcZBP, a zinc-binding protein from Bacillus cereus. *FEBS J* 2007, 274:3044–3054.

60. Deli A, Koutsioulis D, Fadouloglou VE, Spiliotopoulou P, Balomenou S, Arnaouteli S, Tzanodaskalaki M, Mavromatis K, Kokkinidis M, Bouriotis V: LmbE proteins from Bacillus cereus are de-*N*-acetylases with broad substrate specificity and are highly similar to proteins in Bacillus anthracis. *FEBS J* 2010, 277:2740–2753.

61. Maynes JT, Garen C, Cherney MM, Newton G, Arad D, Av-Gay Y, Fahey RC, James MNG: The crystal structure of 1-D-myo-Inosityl 2-Acetamido-2-deoxy-a -D-glucopyranoside deacetylase (MshB) from Mycobacterium tuberculosis reveals a zinc hydrolase with a lactate dehydrogenase fold. *J Biol Chem* 2003, 278:47166–47170.

62. * Woodward RL, Castleman MM, Meloche CE, Karpen ME, Carlson CG, Yobi WH, Jepsen JC, Lewis BW, Zarnosky BN, Cook PD: X-ray crystallographic structure of BshB, the zinc-dependent deacetylase involved in bacillithiol biosynthesis. *Protein Sci* 2020, 29:1035–1039.
Article reporting the 3D structure of a deacetylase BshB that participates in bacillitiol biosynthesis with high similarity to *Mycobacterium* MshB involved in mycothiol biosynthesis. Enzymes involved in the biosynthesis of these low-molecular-weight thiol antioxidants are antimicrobial targets.

63. Chandrangsu P, Loi V, Antelmann H, Helmann JD: The role of bacillithiol in gram-positive Firmicutes. *Antioxidants Redox Signal* 2018, 28:445–462.

64. Reyes AM, Pedre B, De Armas MI, Tossounian M-AA, Radi R, Messens J, Trujillo M: Chemistry and redox biology of mycothiol. *Antioxidants Redox Signal* 2018, 28:487–504.

65. Steffek M, Newton GL, Av-Gay Y, Fahey RC: Characterization of Mycobacterium tuberculosis mycothiol S-conjugate amidase. *Biochemistry* 2003, 42:12067–12076.

66. * Yushchuk O, Ostash B, Truman AW, Marinelli F, Fedorenko V: Teicoplanin biosynthesis: unraveling the interplay of structural, regulatory, and resistance genes. *Appl Microbiol Biotechnol* 2020, 104:3279–3291.
Comprehensive review on antibiotic lipoglycopeptides that are frontline therapies for treating severe infections caused by multidrug-resistant Gram-positive pathogens. An integrated model of teicoplanin biosynthesis is provided.

67. ** Huang CM, Lyu SY, Lin KH, Chen CL, Chen MH, Shih HW, Hsu NS, Lo IW, Wang YL, Li YS, *et al.*: Teicoplanin reprogrammed with the N-Acyl-Glucosamine pharmacophore at the penultimate residue of aglycone acquires broad-spectrum antimicrobial activities effectively killing gram-positive and -negative pathogens. *ACS Infect Dis* 2019, 5:430–442.
Nice example of protein engineering to modify the selectivity of a deacylase and a transferase to redirect the N-acylglucosamine pharmacophore to a different position of the cyclic peptide aglycon resulting in an analog with markedly enhanced antimicrobial activity against resistant *Staphilococcus aureus*.

68. Vincent F, Yates D, Garman E, Davies GJ, Brannigan JA: The three-dimensional structure of the N-acetylglucosamine-6-phosphate deacetylase, NagA, from Bacillus subtilis: a member of the urease superfamily. *J Biol Chem* 2004, 279:2809–2816.

69. Manjunath L, Coombes D, Davies J, Dhurandhar M, Tiwari VR, Dobson RCJ, Sowdhamini R, Ramaswamy S, Bose S: Quaternary variations in the structural assembly of *N*-acetylglucosamine-6-phosphate deacetylase from *Pasteurella multocida* Protein. *Struct Funct Bioinform* 2020, https://doi.org/10.1002/prot.25996.

70. Whittington DA, Rusche KM, Shin H, Fierke CA, Christianson DW: Crystal structure of LpxC, a zinc-dependent deacetylase essential for endotoxin biosynthesis. *Proc Natl Acad Sci U S A* 2003, 100:8146–8150.

71. Erwin AL: Antibacterial drug discovery targeting the lipopolysaccharide biosynthetic enzyme LpxC. *Cold Spring Harb Perspect Med* 2016, 6:a025304.

72. Lee CJ, Liang X, Wu Q, Najeeb J, Zhao J, Gopalaswamy R, Titecat M, Sebbane F, Lemaitre N, Toone EJ, *et al.*: Drug design from the cryptic inhibitor envelope. *Nat Commun* 2016, 7: 10638.

73. Kalinin DV, Agoglitta O, Van de Vyver H, Melesina J, Wagner S, Riemann B, Schäfers M, Sippl W, Löffler B, Holl R: Proline-based hydroxamates targeting the zinc-dependent deacetylase LpxC: synthesis, antibacterial properties, and docking studies. *Bioorg Med Chem* 2019, 27:1997–2018.

74. Cohen F, Aggen JB, Andrews LD, Assar Z, Boggs J, Choi T, Dozzo P, Easterday AN, Haglund CM, Hildebrandt DJ, *et al.*: Optimization of LpxC inhibitors for antibacterial activity and cardiovascular safety. *ChemMedChem* 2019, 14:1560–1572.

75. Blank BR, Talukder P, Muir RK, Green ER, Skaar EP, Renslo AR: Targeting mobilization of ferrous iron in Pseudomonas aeruginosa infection with an iron(II)-Caged LpxC inhibitor. *ACS Infect Dis* 2019, 5:1366–1375.

76. Hoff K, Mielniczuk S, Agoglitta O, Teresa Iorio M, Caldara M, Bülbül EF, Melesina J, Sippl W, Holl R: Synthesis and biological evaluation of triazolyl-substituted benzyloxyacetohydroxamic acids as LpxC inhibitors. *Bioorg Med Chem* 2020, https://doi.org/10.1016/j.bmc.2020.115529.

77. * Surivet JP, Panchaud P, Specklin JL, Diethelm S, Blumstein AC, Gauvin JC, Jacob L, Masse F, Mathieu G, Mirre A, *et al.*: Discovery of novel inhibitors of LpxC displaying potent in vitro activity against gram-negative bacteria. *J Med Chem* 2020, 63: 66–87.
Medicinal chemistry work in which a scaffold hopping approach on a known family of methylsulfone hydroxamate LpxC inhibitors has identified several hit series eliciting potent antibacterial activities against *Enterobacteriaceae* and *Pseudomonas aeruginosa*.

78. Lee MJ, Geller AM, Bamford NC, Liu H, Gravelat FN, Snarr BD, Le Mauff F, Chabot J, Ralph B, Ostapska H, *et al.*: Deacetylation of fungal exopolysaccharide mediates adhesion and biofilm formation. *mBio* 2016, 7. e00252-16.

79. Miyazawa K, Yoshimi A, Sano M, Tabata F, Sugahara A, Kasahara S, Koizumi A, Yano S, Nakajima T, Abe K: Both galactosaminogalactan and α-1,3-glucan contribute to aggregation of Aspergillus oryzae hyphae in liquid culture. *Front Microbiol* 2019, 10.