Synthesis of Sonic Hedgehog protein and development of peptide binders using phage display libraries

Judith Palà Pujadas

ADVERTIMENT. La consulta d’aquesta tesi queda condicionada a l’acceptació de les següents condicions d'ús: La difusió d’aquesta tesi per mitjà del servei TDX (**www.tdx.cat**) i a través del Dipòsit Digital de la UB (**diposit.ub.edu**) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d’investigació i docència. No s’autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d’un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s’autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (**www.tdx.cat**) y a través del Repositorio Digital de la UB (**diposit.ub.edu**) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you’re accepting the following use conditions: Spreading this thesis by the TDX (**www.tdx.cat**) service and by the UB Digital Repository (**diposit.ub.edu**) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it’s obliged to indicate the name of the author.

Programa de Doctorat de Química Orgànica

Synthesis of Sonic Hedgehog protein and development of peptide binders using phage display libraries

Judith Palà Pujadas

Tesi doctoral dirigida per:

Dr. Juan B. Blanco Canosa
QAC-CSIC
Departamento de Química Biológica
Modelización Molecular

Prof. Fernando Albericio Palomera
Universitat de Barcelona

Barcelona, 2018

ACKNOWLEDGMENTS

I després de quatre anys ha arribat el moment que semblava tan llunyà quan vaig començar a fer els primers passos per un laboratori i a tastar això de fer ciència. Recordo quan buscava un grup on fer el doctorat, i finalment vaig venir-me a entrevistar amb el Juan. Una de les coses que més em va agradar va ser la seva sinceritat, l'energia, empenta i mil idees que tenia en ment pel projecte.

D'aquell dia ja fa uns anys, i he crescut molt des de llavors, tan personal com professionalment. Sembla ahir quan feia servir per primera vegada un HPLC, i és quan la gent et ve a preguntar o consultar com funciona un equip, que t'adones que ja ets gran, i que portes aquí força temps. Un temps en el que han passat moltes històries, en el que hi ha hagut moments de tot, en el que he tingut l'oportunitat de conèixer i compartir moments amb persones entranyables. Han sigut uns anys que quedaran sota el paraigua de "quan estava fent la tesis" i serà la manera com s'explicaran les batalletes d'aquesta època. I durant aquesta pel·lícula titulada "Com començar una tesis doctoral, sobreviure durant i no morir intentant escriure-la" hi ha actuat molts actors, tots ells amb un paper que ha aportat el seu granet de sorra al guió. Així doncs, m'agradaria dedicar aquesta tesi a totes aquestes persones que han tingut el seu moment de protagonisme durant aquesta etapa.

En primer lloc vull agrair als meus directors, al **Juan** i al **Fernando**, la confiança que van dipositar en mi, i haver-me donat el suport necessari durant aquests anys per tirar endavant el projecte. En especial al **Juan**, por todo el conocimiento que me has transmitido durante estos años, el apoyo que me has dado y la paciencia que has tenido conmigo. Por haber estado codo con codo en este proyecto, por tu tiempo y por animarme a emprender nuevas aventuras en las Américas. Sé que no he sido una estudiante ejemplar, que por ser la primera te ha tocado una con algo de carácter y que no lleva bien las críticas, pero bueno con el tiempo creo que nos hemos amoldado. Y ahora que ya nos conocíamos, va y se acaba el doctorado! Pues nada ahora me toca volar. Espero haber estado a la altura, y que yo también te haya enseñado un poco aunque es difícil porque tú sabes bastante de todo. En definitiva, un placer haber sido tu first PhD Student.

A tots els meus companys de grup, que no de lab perquè hem estat sempre desperdigats. Als que hi eren quan vaig començar a donar els primers tombs per un lab de pèptids: a la **Marta** i a la **Myriam**, per ajudar-me amb tots els problemes que tenia amb els HPLCs, al **Jan**, per les seves excèntriques ocurrències, al **Michael**, who taught us wee words in Scotish and was the definitely the most passionate guy about Barça, a la **Laia**, pel seu somriure contagiós i els seus consells que sempre acabaven amb un “però si tu ets molt jove”, al **Jesús**, por apoyarme durante mis momentos de bajón cuando me trasladé al lab 100, a **Kamil**, breve compañero de vitrina con quien era imposible no reír con alguna de sus mil bromes, a **Anaïs**, mi compañera de vitrina más duradera, por nuestras charlas durante horas y la conclusión de que los hombres son así, y a **Iván**, por recordarme siempre lo torpe que soy y dejar constancia con tus fotos.

A tots aquells que encara hem aguantat després de tots els problemes habidos y por haber: a **l’Alex**, un jugador de golf que tiene esto de la química como hobby, a **l’Helena**, pel teu suport durant tots aquests anys, i les estones posant-nos al dia dels cotilleos. Als dos us desitjo moltíssima sort en la nova etapa de PhDs. Al **Gerardo**, algo bueno que ha tenido tantos años de tesi es que al final nos hemos llevado bien y todo, i a la **Núria**, per ajudar-me amb els westerns, i a distància amb les ELISA.

A la meva família adoptiva del lab 300, per aguantar les meves queixes i esternuts constants. En primer lloc a **l’Ernest**, per acollir-me al seu laboratori, i a la **Eva**, perquè sense ella la paperassa seria encara més difícil. A la **Júlia**, companya d’al·lèrgies i de semiprep, la **Cris D.**, sempre disposada a donar un cop de mà ja sigui amb els equips, l’excel o la burocràcia de la tesi, a la **Cris Fuster**, amb qui comparteixo un amor platònic anomenat GELS, al **Salva**, un dels meus ecos quan esternudo, pel seu positivisme i haver-me carregat a Porto quan estava ben tullida, potser aquí també podria incloure-hi el **Pol** abans que m’inundés la vitrina, al **Pep**, per ser la mosca més pesadeta del lab, a qui molestaràs ara?, a la **Macarena**, por tu ayuda infinita siempre que la he necesitado, a **l’Adam**, un còmplice en el crim del salmó fumat, al **Martí**, per fer més entretinguts els llargs dies d’escriptura al lab, a la Monica, por tu expertise en CD, a **Tony**, if I learned something is that when you bet is to win, your bottle of wine is coming soon, promise, a **Federica**, por las tardes de merienda y las imágenes de la portada, i a la resta de gent, al **Fabio**, la **Sandra**, la **Maria**, la **Txell**, la **Daniela i** la **Cris G**. que m’han arrencat un somriure inclús els dies més difícils.

Als de UQC, companys de seminaris, a la **Myriam** pel seu positivisme, i al **Dani**, sempre amb una resposta per la química que no surt i molts ànims per quan segueix sense funcionar, a **Patri**, por nuestras charlas durante mis mil horas purificando.

Als companys de promoció de l’IRB, pels dinars compartits durant aquests anys: al **Craig**, **l’Ernest**, la **Gemma**, la **Laura**, i **l’Helena R**., i **l’Anna**, perquè ens hem unit molt aquests últims anys, i perquè ets qui millor m’entén quan dic que vull tornar a Canadà, potser perquè fer de bosseta de te ens queda pendent?

També als meus recent estrenats companys de pis: a la **Bea** i al **Sergio**, porque hacéis que me apetezca llegar a casa aunque sea para ver una peli mala o celebrar fin de año tres días antes.

A **l'Ari**, per les nostres xerrades de mil temes, i perquè mirant-me als ulls saps llegir-me. A les meves farmacèutiques: la **Eider** i la **Mireia**, companyes de nits de NHH, perquè em lieu però també em compreneu, a la **Núria**, la meva groupie número 1, a qui confio tant de mi, abonada als capítols setmanals de les meves aventures, a la **Iris**, tota una doctora, per tot el que hem compartit, aquí i a l'altra banda de l'oceà, i el que ens queda allà on estiguem.

També a tot l'equip de la Farmàcia Gaudí: al **Marcel**, per no cansar-se de les meves excedències quan he anat a fer les Amèriques, a la **Bea**, per tots els matins de dissabte i creure més en mi del que jo mateixa crec moltes vegades, a la **Vicky**, pel seu positivisme contagiós i les seves teories de peixos i tiburons, a la **Cris**, pels soparets i moments compartits.

Als meus pilars, la **Roser**, que tot i estar lluny sé que sempre hi puc comptar, a **l'Aleix**, perquè amb poques paraules diu molt, a la **Montse**, amb qui m'encanta perdrem per l'est o l'oest. Al **Carles**, pels seus consells i contagiar-me ambició. A **l'Alex**, per les birres de tarda que acaben de nit. També a tots aquells que han estat presents de manera important durant aquest temps, en especial a **l'Antoni**, que va patir els meus alts i baixos de tota aquesta aventura .

I also wanted to thank all the people I met in Canada. Especially to **Dev**, for giving me the opportunity to work at his lab. To **Gang**, for all his knowledge and teaching when I was naïve in phage display, to **Christine**, for her help getting through the paperwork, to **Shane** for amusing the lab with his music, to **Penny**, for her understanding about being a chemistry PhD Student, for her helpfulness and friendship, to **Leoni**, for all the chats about life, to **Tunga**, for all lunch times shared even under the rain, to **Valencio**, for his laughter and marriage questions, to **Gianlucca**, **Jan** and **Isabel**, for welcoming me to their trip adventures. To **Joan** and **Monica**, for their BBQs, to **Mireia**, for the first advises when I landed in Toronto, and all the months shared.

Also to other people outside the lab, to **Rébecca**, my long distance best friend, for her endless help in settling my second time, and thanks in advance for helping me with my PR. To **Afshan**, someone who really felt me, to **Staza**, hopefully we'll have more adventures together. To all the people from Steps, who was my family in Toronto. To **Almu**, por los momentos haciendo las gorditas y las noches de bailoteo. A **Cristina**, por compartir tanto, por emocionarnos juntas. And finally to **Ilnam**, for making every moment rememberable.

Vull dedicar aquesta tesis a la meva **família**. Als meus **tiets** i **cosins**, en especial als meus padrins, el **Cinto** i la **Isabel**, per estar sempre al meu costat, a les meves **tietes guays** per ensenyar-me a escriure abans de caminar quasi, a **l'Albert**, amb qui sempre puc comptar, i a **l'Uri**, per les nostres xerrades.

Als meus avis, per donar-me sempre suport en les meves decisions i amor incondicional. A la **iaia**, la generositat personificada, la que sempre em diu que sóc massa impacient i intranquil·la, però que tot i això m'estima inclús quan marxo lluny. A la **iaia Maria**, perquè allà on siguis espero que estiguis orgullosa de mi. I en especial a tu **avi**, que vas deixar-nos a mig camí d'aquest projecte, per tot el que em vas ensenyar, per tots els records que ens vas deixar.

Al meu **germà**, per compartir tants moments amb mi, per aguantar-me quan he sigut una mica mandona, per dir-me les coses tal com són, i perquè mai em dius que no quan realment et necessito.

I finalment, vull acabar dedicant aquesta tesis als **meus pares**, un exemple de superació, per donar-m'ho tot, per lluitar per nosaltres, per deixar-me llibertat per decidir, i recolzar-me en les meves decisions. A tu mama, per la teva paciència infinita amb aquesta filla teva que és una mica "sargento", perquè sempre m'has ensenyat la importància de ser una dona independent i per animar-me a fer realitat els meus somnis. I a tu papa, per mimar-me i consentir-me en la justa mesura, per tot el que m'has ensenyat, i perquè tot i les discrepàncies en política i altres temes, m'has demostrat que sempre estàs al meu costat. Gràcies a tots dos per fer-me créixer en uns valors que m'han permès arribar fins aquí.

There are no shortcuts
to any place worth going.

Beverly Sills

CONTENTS

ABBREVIATIONS I
INTRODUCTION 1
REFERENCES 7
CHAPTER 1: Synthesis of MeDbz linker derivatives 11
INTRODUCTION 13
1. Chemical Ligation 16
2. Native Chemical Ligation 17
2.1. Principle of NCL 17
2.2. Limitations of NCL 18
3. Synthesis of C-terminal thioester peptides 20
3.1. Direct synthesis of C-terminal thioester on resin using Fmoc-SPPS 20
3.2. Thioesterification of fully protected peptide acids 20
3.3. Conversion of resin-bound peptide allyl esters into peptide thioesters 21
3.4. BAL strategy 22
3.5. Safety catch linker using Kenner’s sulfonamide linker 22
3.6. SPPS of peptide thioesters at aryl hydrazide linker 23
3.7. SPPS of thioesters through O -> S and N -> S acyl shift 24
3.8. Peptide α-N-acyl-benzimidazolinones as precursors for peptide α-thioesters .. 25
3.9. Peptide O-aminoanilides as cryptothioester 26
3.10. SEA approach as crypto-thioesters 27
3.11. SEAlide as cryptothioesters 28
4. Kinetic controlled ligation 28
5. Ligations inspired by NCL 30
5.1. Conformational assisted ligation 30
5.2. Expressed Protein Ligation 30
OBJECTIVES 33
RESULTS AND DISCUSSION 37
1. Activation of MeDbz peptides in solution 39
1.1. Effect of the pH on the hydrolysis of the pNO$_2$-Phoc-MeDbz peptides 41
1.2. Effect of the acidic character of the substituted phenol leaving group on the hydrolysis of the aryloxycarbonyl-MeDbz peptides 42
1.2.1. mNO2-Phoc-MeDbz 43
1.2.2. Phoc-MeDbz 44

1.2.3. pF-Phoc-MeDbz 44

1.2.4. pSO_2Me-Phoc-MeDbz 45

1.2.5. pCN-Phoc-MeDbz 46

2. NCL assay using model *p*CN-Phoc-MeDbz peptides 46

3. Racemization studies 48

4. KCL using *p*CN-Phoc-MeDbz peptides 49

5. Stability of *p*CN-Phoc-MeDbz peptides under common protein synthesis reactions. 51

5.1. Acm removal 51

5.2. Desulfurization 52

5.3. Thz conversion to Cys 54

CONCLUSIONS 57

REFERENCES 61

CHAPTER 2:Synthesis of Sonic Hedgehog 69

INTRODUCTION 71

1. Hh signaling pathway 73

2. Biosynthesis of Sonic Hedgehog 76

3. Structure of Shh 77

3.1. Hydrophobic modifications in Shh 78

3.1.1. Modifications at the N-terminus 78

3.1.2. Modifications at the C-terminus 79

4. Functions of Hh and related diseases 79

5. Role of Hh in cancer 81

6. Inhibitors and agonists of Hh 82

6.1. Smo inhibitors 84

6.1.1. Cyclopamine and derivatives 84

6.1.2. Other small molecules 84

6.2. Shh inhibitors 85

6.2.1. Monoclonal antibody 5E1 85

6.2.2. Robotnikinin 85

6.3. GLI inhibitors 85

7. Importance of Hh and ShhN 86

OBJECTIVES 87

RESULTS AND DISCUSSION 91

1. Synthesis of natural ShhN: Palm-ShhN-cholesterol 93

1.1. Retrosynthesis of ShhN: strategy A 93
1.2. Strategy A towards the Palm-ShhN-cholesterol 94
1.3. Synthesis of the peptide fragments 95
1.4. Retrosynthesis of srategy B towards the synthesis of Palm-ShhN-cholesterol 100
1.5. Strategy B toward natural Palm-ShhN-cholesterol 100
1.6. Synthesis of the peptide fragments 101
1.7. KCL and NCL of the peptide fragments 104
1.7.1. Synthesis of the N-terminal block: Palm-Ala^1-Arg^{78}-pCN-Phoc-MeDbz 105
1.7.2. Synthesis of the C-terminal block: Cys^{79}-Gly^{174}-cholesterol 106
1.8. Strategy C towards the synthesis of Palm-ShhN-cholesterol 108
1.9. Synthesis of the peptide fragments 109
2. Synthesis of the Palm-ShhN-biotin 111
2.1. Restrosynthesis of the synthetic analog Palm-Shh-biotin 112
2.2. Synthetic strategy towards the synthetic analog Palm-ShhN-biotin 112
2.3. Synthesis of the peptide fragments 113
2.4. . Synthesis of the C-terminal block: Cys^{79}-Lys^{175}-PEG$_2$-biotin 114
2.5. Assembling of the C-terminal and N-terminal parts 119
2.6. Folding 122
2.6.1. . Folding of Palm-ShhN-biotin by dialysis 122
2.6.2. Protein Gel Electrophoresis and silver staining 124
2.6.3. Quick dilution strategies 124
2.6.4. Circular Dichroism 125
2.6.5. ELISA 126
3. Synthesis of the IleIle-ShhN-biotin 127
3.1. Retrosynthetic analysis of IleIle-ShhN-biotin 127
3.2. Synthetic strategy towards IleIle-ShhN-biotin 128
3.3. Synthesis of the N-terminal block: Ile^1-Arg^{79}-*p*CN-Phoc-MeDbz 129
3.3.6. Folding of the IleIle-ShhN-biotin 133
CONCLUSIONS 135
CHAPTER 3: Phage Display against Sonic Hedgehog 147
INTRODUCTION 163
1. Methodology 151
2. Filamentous bacteriophage biology 152
3. Phage display in protein pVIII 154

4. Phage display in protein pIII 155

5. Phagemid vectors and helper phage 155

  5.1. Phagemid vectors 155

  5.2. Helper phage 156

6. Design and construction of libraries for phage display 157

  6.1. Random peptide libraries 157

  6.2. Rational peptide libraries 159

7. Affinity maturation 159

8. Other display systems 159

  8.1. Bacterial display 160

  8.2. Yeast display 160

  8.3. mRNA and ribosome display 160

OBJECTIVES 163

RESULTS AND DISCUSSION 167

1. Library selection to perform phage display against ShhN 169

2. Phage display against commercial ShhN 171

  2.1. Biopanning against commercial ShhN 172

  2.2. Analysis of unique clones obtained 174

    2.2.1. Analysis of unique clones obtained from library SCF32 for commercial ShhN 174

    2.2.2. Analysis of the unique clones obtained from library SCF42 for commercial ShhN 175

3. Phage display against N-terminal biotinylated ShhN 176

  3.1. Biopanning against N-terminal biotinylated ShhN 177

    3.1.1. Analysis of the unique clones obtained from library SCF32 for N-terminal biotinylated ShhN 178

4. Phage display against synthetic Palm-ShhN-Biotin analog 179

  4.1. Biopanning against Palm-ShhN-Biotin 180

    4.1.1. Analysis of the unique clones obtained from library SCF32 for Palm-ShhN-Biotin 181

  4.2. Cross reactivity ELISA 182

5. Phage display against IleIle-ShhN-Biotin analog 182

  5.1. Biopanning against IleIle-ShhN-Biotin 183

6. Competitive elisa 184

6.1. Competitive ELISA with binders from phage display against Palm-ShhN-Biotin 184
7. Affinity maturation for Palm-ShhN-Biotin 185
7.1. Affinity maturation with binders from SCF32 obtained for Palm-ShhN-Biotin. 186
7.1.1. Amplification of the gene containing the unique clones 187
7.1.2. Digestion and ligation of the amplified inserts into pIII paghemid vector 188
7.2. Design of the library for affinity maturation 188
CONCLUSIONS 191
REFERENCES 195
MATERIALS AND METHODS 201
1. Reagents and solvents 203
2. Materials 204
3. Instrumentation 204
3.1. General instrumentation 204
3.2. RP-HPLC. 205
3.2.1. Analytitcal RP-HPLC 205
3.2.2. Semi-preparative RP-HPLC 205
3.3. Mass spectrometry 205
3.4. MALDI-TOF 205
3.5. Nuclear magnetic resonance 206
3.6. Microwave synthesis. 206
3.7. Circular dichroism 206
4. Buffers 206
5. Peptide synthesis and characterization 207
5.1. General considerations 207
5.2. Colorimetric tests 208
5.3. Conditioning of the resin 209
5.4. Loading of the first aminoacid 209
5.5. Fmoc removal 209
5.6. Quantification of resin loading capacity 209
5.7. Chain elongation in manual synthesis 210
5.8. Coupling of the linker Fmoc-MeDbz 211
5.9. Coupling of the first aminoacid after the MeDbz linker 211
5.10. Chain elongation in automated microwave synthesis 211

5.11. Cleavage and deprotection of side chains 212
5.12. Purification 212
5.13. Peptide characterization 213
6. Synthesis of *p*CN-Phoc-MeDbz peptides and derivatives 213
6.1. Synthesis of Fmoc-MeDbz linker 213
6.2. General synthesis of phenylchloroformates 213
6.3. General procedure for the synthesis of *p/m*X-Phoc-MeDbz peptides 214
6.4. General procedure for the synthesis of MeNbz peptides 214
6.5. General activation of *p*X-Phoc-MeDbz peptides in solution 215
7. NCL and KCL 215
7.1. General protocol for NCL: 215
7.2. General procedure for the synthesis of MESNa thioester peptides from *p*CN-Phoc-MeDbz 215
7.3. Racemization experiments: 216
7.4. KCL experiment between MeNbz, *p*CN-Phoc-MeDbz and cysteinyl peptide: ... 216
8. Common reactions on chemical protein synthesis 216
8.1. General procedure for Acm removal: 216
8.2. General procedure for radical-based desulfurization: 216
8.3. General procedure for thiazolidine removal: 217
9. Synthesis of Palm-ShhN-biotin 217
9.1. Peptide fragment syntheses 217
9.2. Fragment ligations 224
9.3. KCL of other strategies 229
10. Synthesis of IleIle-ShhN-biotin 232
10.1. Peptide fragment synthesis 232
10.2. Ligation of the fragments 233
11. Folding of ShhN analogs 235
11.1. Electrophoresis 235
11.2. Dialysis 237
11.3. FAST dilution 237
11.4. ELISA 238
12. Phage display 238
12.1. Library preparation for phage selection rounds 238
12.2. Protein reconstitution or refolding 239

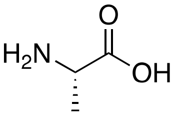
12.3. Biopanning selections ... 239
12.4. Selection of colonies ... 241
12.5. ELISA ... 241
12.6. PCR ... 242
12.7. Competitive ELISA ... 243
12.8. 12.8. Affinity maturation. Switch from pVIII to pIII phagemid vector ... 245
  12.8.1. pIII Phagemid vector miniprep ... 245
  12.8.2. Plasmid digestion from selected clones from pVIII vectors. ... 246
  12.8.3. Ligation ... 247
12.9. Cross-reactivity ELISA of Palm-Shh-biotin binders against recombinant ShhN 248
Product characterization ... 251
ANNEXES ... i

ABBREVIATIONS

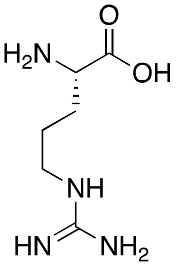
2-CTC	2-chlorotrytil chloride resin
4-DMAP	4-(dimethylamino)pyridine
AA	amino acid
Ab	antibody
Abs	absorbance
ACN	acetonitrile
Ag	antigen
BCC	Basal cell carcinoma
Boc	*tert*-butyloxycarbonyl
CD	circular dichroism
CPS	chemical protein synthesis
CSC	Cancer stem cell
DCM	dichloromethane
Dhh	Desert Hedgehog
DIC	diisopropylcarbodiimide
DIEA	*N,N*-diisopropylethylamine
DMF	dimethylformamide
DMF	*N, N*-dimethylformamide
DMSO	dimethylsulfoxide
EPL	Expressed protein ligation
Equiv	equivalent
ESI	electrospray ionization
Et$_2$O	diethyl-ether
Fmoc	9-fluorenymethyloxycarbonyl
Gli	Glioma related transcription factors.
HATU	1-[Bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-*b*]pyridinium 3-oxid hexafluorophosphate
HBTU	*N,N,N′,N′*-Tetramethyl-*O*-(1*H*-benzotriazol-1-yl)uronium hexafluorophosphate
Hh	Hedhegog pathway
HHAT	Hedgehog Acyl Transferase
IC$_{50}$	half maximal inhibitory concentration
Ihh	Indian Hedgehog
IP2	inositol triphosphate
IPTG	Isopropyl β-D-1-thiogalactopyranoside
KCL	Kinteic controlled ligation
LB	lysogeny broth
m/z	mass over charge
MaB	monoclonal antibody
MeDbz	o-aminomethylanilide linker
MeNbz	N-acyl-N′-methylacylurea
MESNa	2-Mercaptoethanesulfonic acid sodium salt
MPOH	4-mercaptophenol
MS	mass spectrometry
MW	molecular weight

NaOH	sodium hydroxide
NCL	native chemical ligation
NMR	nuclear magnetic resonance
Palm	Palmitic acid
pCN-Phoc-MeDbz	p-cyanophenyloxycarbonyl-MeDbz
PDA	photodiode array
PDB	protein data bank
pNO2-Phoc-MeDbz	p-nitrophenyloxycarbonyl-MeDbz
Ptc	Patched
rT	room temperature
SCF	scaffold
ShhN	Sonic Hedgehog
Smo	Smoothened
SUMO	small ubiquitin-like modifier
*t*Bu	*tert*-butyl
TCEP	(tris(2-carboxyethyl)phosphine)
TFA	trifluoroacetic acid
TIS	Triisopropylsilane
Trt	trityl
UPLC	ultra-high performance liquid chromatography
UV	ultraviolet
VA-044	VA-044 2,2'-Azobis[2-(2-imidazolin-2-yl)propane] Dihydrochloride

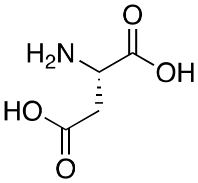
Amino acids



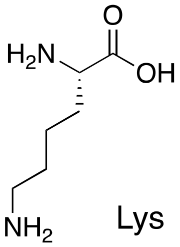
Ala



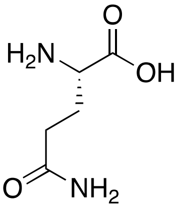
Arg



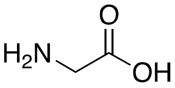
Asp



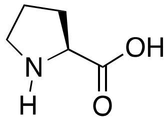
Lys



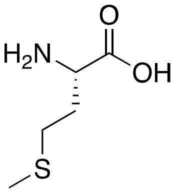
Gln



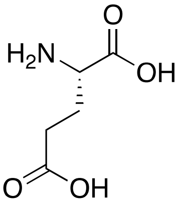
Gly



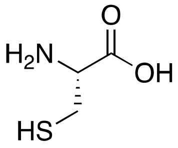
Pro



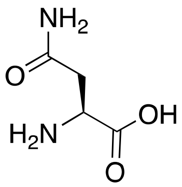
Met



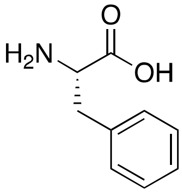
Glu



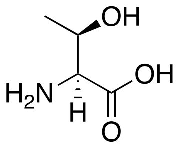
Cys



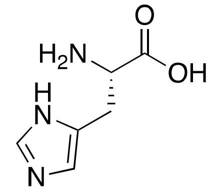
Asn



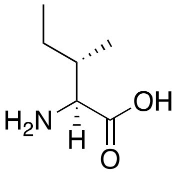
Phe



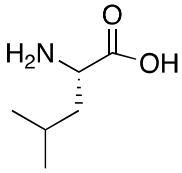
Thr



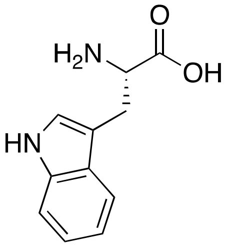
His



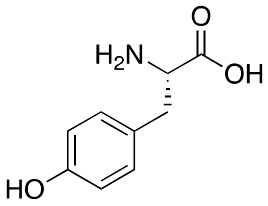
Ile



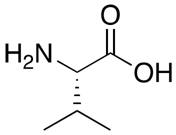
Leu



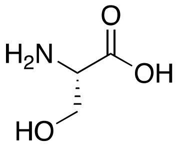
Trp



Tyr

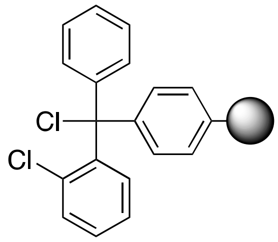


Val

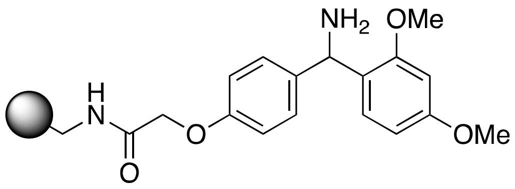


Ser

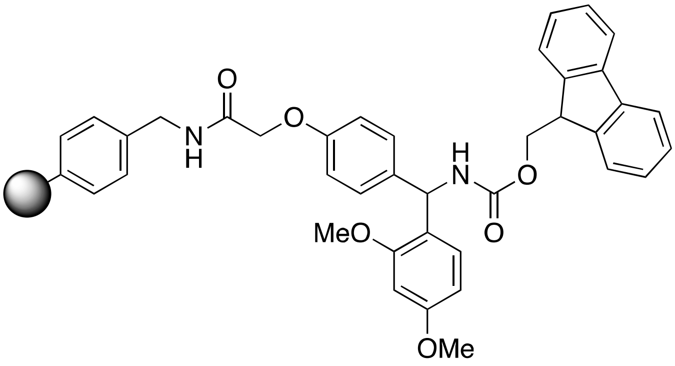
Resins



2-chlorotrytil chloride resin

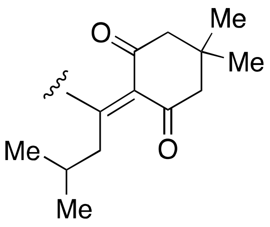

NH2
OMe
H
N
O
O
OMe

Rink Amide ChemMatrix

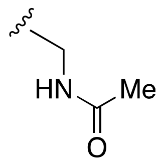

O
O
N
H
H
N
O
O
MeO
OMe

Fmoc-Rink Amide Aminomethyl resin

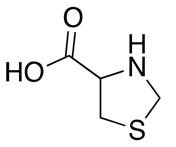
Special protecting groups


O
Me
Me
O
Me
Me

ivDde

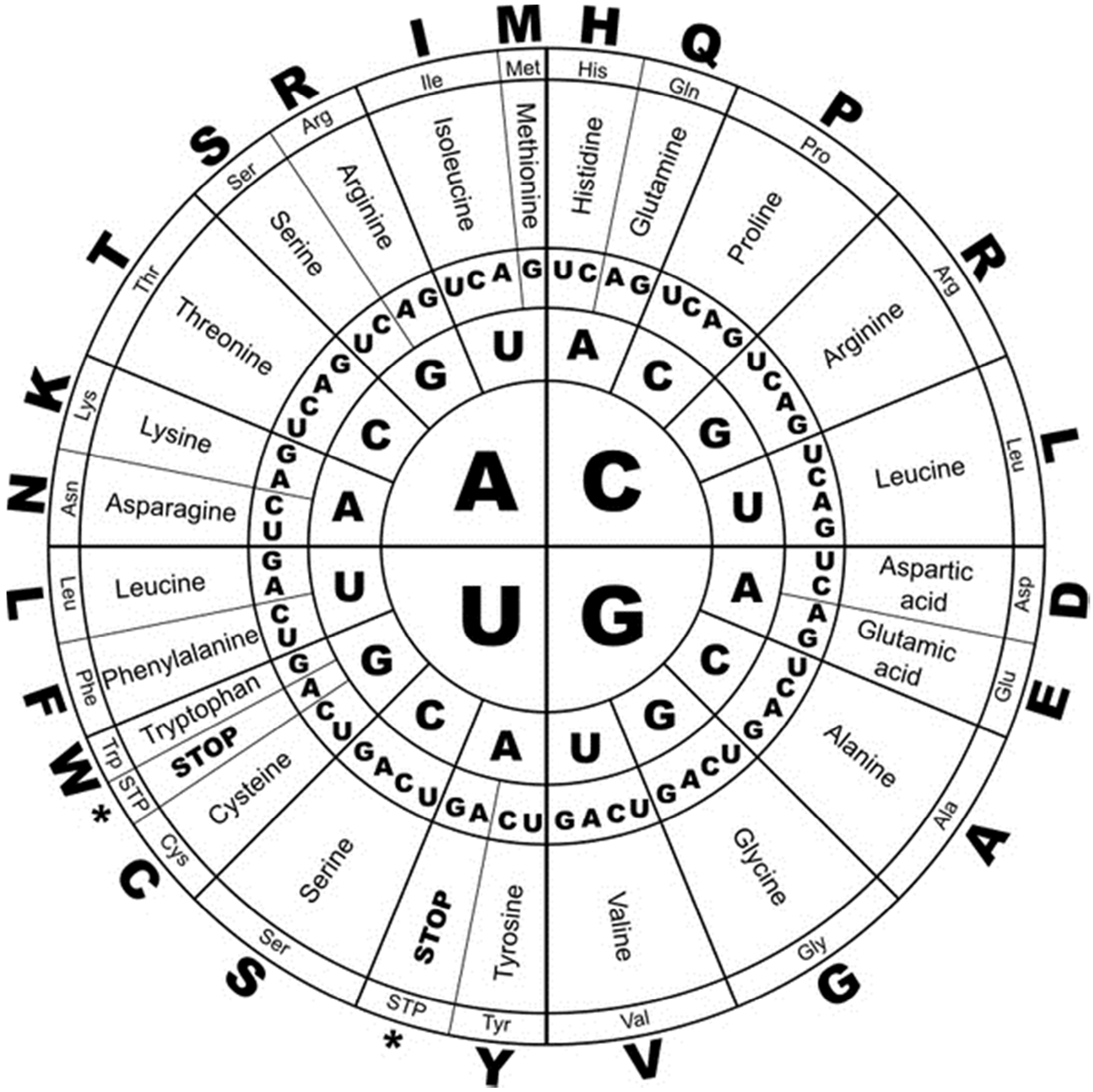


Acm



Thiazolidine

Genetic code


A
C
U
G
U
C
A
G
UCAG
Methionine
Met
M
Isoleucine
Ile
I
Arginine
Arg
R
Serine
Ser
S
Threonine
Thr
T
Lysine
Lys
K
Asparagine
Asn
N
Leucine
Leu
L
Phenylalanine
Phe
F
Tryptophan
Trp
W
STOP
STP
*
Cysteine
Cys
C
Serine
Ser
S
STOP
STP
*
Tyrosine
Tyr
Y
Valine
Val
V
Glycine
Gly
G
Alanine
Ala
A
Glutamic acid
Glu
E
Aspartic acid
Asp
D
Leucine
Leu
L
Arginine
Arg
R
Proline
Pro
P
Glutamine
Gln
Q
Histidine
His
H

INTRODUCTION

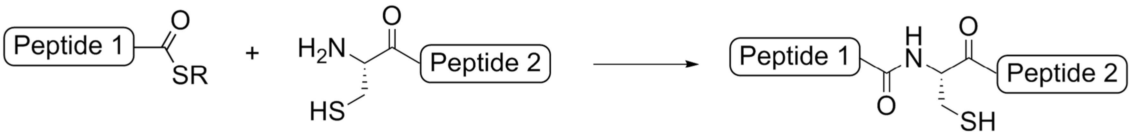
After the Human Genome Project was completed in 2003 claiming the successful sequencing of the whole DNA genome, the world entered into the proteomic era, in which proteins would star as the main actors.

The sequencing of the human genome resulted in 16,537 predicted proteins, which are responsible for the complexity of the organism.[1] If anyone can imagine the proteins as flexible and dynamic systems, which in addition have many intermolecular interactions among them, this complexity increases exponentially. Besides, what make the proteins the real protagonists in living organisms are the biological functions that sustain the existent equilibrium within a complex network and separates any individual from healthy to sick.

The importance acquired by proteins since the beginning of the proteomic era moved the scientific community towards the research of new tools to enable the study of these biological macromolecules and their interactions. This proteomic blooming experienced a great growth in the last years and fostered technologies such as recombinant DNA methods and engineering of new cell types designed to produce recombinant proteins. Despite the great achievements and progresses in this field, some proteins remain still challenging to be produced by these methods. Some of them result in low expression yields, while others become really difficult or even impossible to be obtained from living systems, like proteins with post-translational modifications, membrane proteins or proteins that become toxic at expression levels. Moreover and moving a step further, the use of cells to express proteins just permits the introduction of natural aminoacids.[2] Very recently, different methods using orthogonal *t*RNA allowed also the introduction of synthetic residues in proteins using bacteria cells.[3] The unmet production of these types of proteins called for the need of new and robust technologies that would allow the access to these macromolecules and overcome the limitations of the current methodologies. In this regard, the obtainment of proteins by chemical synthesis gives a unique opportunity to achieve the production of these entities.

Chemical protein synthesis (CPS) represents an alternative to the production of proteins by recombinant methods. It allows a full control of the protein sequence, and also the introduction of unnatural amino acids or residues for site-specific protein modification. Accordingly, synthesis of post-translational modified proteins such as glyco, phospho or lipoproteins is also feasible.

Since the middle of the last century, the scientific community witnessed the evolution of the peptide chemistry field, from the synthesis of the first peptides[4] to the development of solid phase peptide synthesis (SPPS)[5] and the first chemical synthesis of an active protein by chemical ligations in solution.[6] These major discoveries confirmed the power and usefulness of synthetic chemistry to achieve the preparation of peptides and proteins. Moreover, these remarkable highlights stablished the bases for the development of new reactions, such as native chemical ligation (NCL). Since it was reported by Dawson and Kent in 1994,[7] the importance of NCL was reaffirmed by hundreds of proteins synthetized using this methodology. It was also the source of inspiration for other reactions, like expressed protein ligation (EPL). The requirement of C-terminal peptide thioesters for NCL represented a limitation due to its challenging synthesis by Fmoc-SPPS. Nevertheless, the great importance of NCL in the protein chemistry field fostered the search of different strategies to efficiently synthetize C-terminal thioester peptides by Fmoc-SPPS. Within all procedures described until the moment, the most used strategies are the *N*-acylurea approach[8] and the hydrazide methodology.[9] The *N*-acylurea approach, reported by Blanco-Canosa and Dawson in 2008 is based on the use of linker 3,4-diaminobenzoic acid (Dbz) as a thioester precursor.



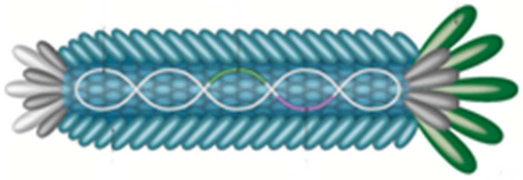
Native Chemical Ligation

In order to span the scope of NCL for the synthesis of large proteins, Kent *et al*. introduced the concept of kinetic controlled ligation (KCL).[10] Differently from the standard NCL, KCL allows the ligation of peptides in the N to C direction. Thereby, the combination of both approaches facilitates the chemical synthesis of larger proteins by a more convergent approach.

Taking advantage of the opportunities offered by KCL and NCL, in the recent years have been reported numerous efforts focused to develop methodologies for the synthesis of peptides able to undergo KCL. The main caveat of these peptides is the controlled activation of the C-terminal amino acid to give NCL at the desired time. Until the moment, the most efficient approach are the alkyl/aryl thioesters and the hydrazide methodology. In the present thesis one of the main objectives is the development of a methodology to prepare thioesters in solution based on the N-acylurea approach.

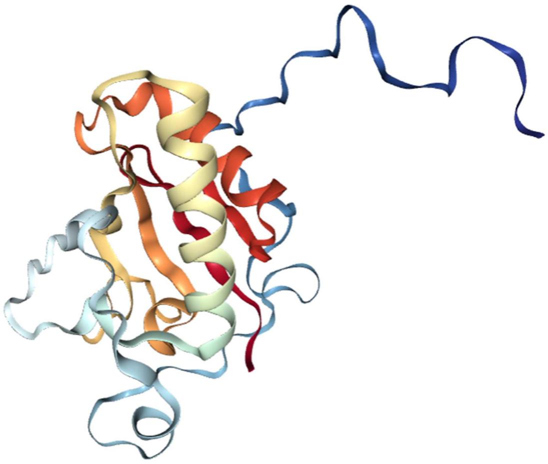
All the advances reported in the production of proteins, either by chemical or recombinant methods, enabled a detailed determination of their structure and also their implication in disease, a key aspect for further development of possible treatments. Currently, proteins play a major role in important diseases such as cancer and are the object of study in several research areas. Therefore, it is highly valuable the acces to sufficient amounts of the protein of interest. At this point is where the CPS gets importance.

Inhibition of protein function by small molecules, peptides or antibodies is a widely used strategy to tackle the incorrect activity of a protein involved in a particular disease. Perhaps the most clever and powerful technique for the development of ligands against a concrete protein is phage display, described in 1985 by Smith.[11] This technology was inspired by the evolution of the natural immune system, and it is basically based on the selection of the ligands that present higher affinity for the target of interest. Since it was reported, another peptide or protein displaying systems (RNA display, yeast display, bacterial cell display) have been described. In addition, the most recent and promising advances in the area combine these DNA or RNA encoded peptide libraries with the application of simple chemistry reactions in order to display more complex peptides, which translates in the obtainment of very specific and tight binders for a concrete target.[12,13]



Filamentous bacteriophage

The present thesis has as principal goal the challenging synthesis of the Sonic Hedgehog protein (ShhN) by applying the methodology developed throughout. This post-translational modified protein results interesting because of its lipidic nature and the biological functions that regulates. During its biosynthesis, ShhN undergoes two post-translational modifications consisting on the attachment of two lipid residues, a palmitic acid at the N-terminus and a cholesterol at the C-terminus, to finally achieve the active form.[14] Due to these modifications, ShhN is difficult to prepare by recombinant methods, and thereby chemical synthesis becomes an alternative. ShhN is the main ligand of the Hedgehog (Hh) signaling pathway. It is involved in organogenesis, tissue patterning or neurodevelopment during embryogenesis. In adults, regulates cell growth and fate, maintenance of stem cells and tissue repairing. Aberrant signaling in Hh pathway have been related to different diseases including cancer.[14] The importance of this pathway in cancer led to the development of inhibitors targeting components downstream on the signaling cascade. Although the known implication of ShhN ligand in aberrant activation of Hh pathway, until the moment, just one small molecule and a monoclonal antibody have been reported to bind this protein.[15,16] There is a need to develop new binders for ShhN. Thereby, we envisaged that the chemical synthesis of ShhN would not just represent a synthetic challenge, but may also help to study or develop ligands that could inhibit its signaling.



Sonic Hedgehog

Having in mind the idea to develop new binders for ShhN, phage display seems a very suitable and interesting technology to find new peptides that could inhibit the aberrant activation of the Hh pathway caused by ShhN. Actually, since the chemical synthesis of ShhN may stablish a feasible strategy towards its preparation, the synthesis of the D-protein could also be possible in a future. At this point, the application of mirror phage display will allow to find D-ligands. The main advantage of the D-peptides is their low immunogenecity and increased resistance to proteases, and therefore their increased clearance time.

The combination of protein chemistry with DNA engineered technologies such as phage display represents a unique approach towards the development of tight and specific binders that later can become candidates for the treatment of diseases in where proteins are missregulated.

REFERENCES

(1) Paik, Y. K.; Overall, C. M.; Deutsch, E. W.; Hancock, W. S.; Omenn, G. S. Progress in the Chromosome-Centric Human Proteome Project as Highlighted in the Annual Special Issue IV. *J. Proteome Res.* **2016**, *15* (11), 3945–3950.

(2) Dawson, P. E.; Kent, S. B. H. S Ynthesis of Native Proteins By C Hemical Ligation. *Annu. Rev. Biochem.* **2000**, *69*, 923–960.

(3) Wang, Q.; Parrish, A. R.; Wang, L. Expanding the Genetic Code for Biological Studies. *Chem. Biol.* **2009**, *16* (3), 323–336.

(4) du Vigneaud, V.; Resslbr, C.; Swan, J. M.; Roberts, C. W.; Katsoyannis, P. G. The Synthesis of Oxytocin. *J. Am. Chem. Soc.* **1954**, *76* (12), 3115–3121.

(5) Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **1963**, *85* (14), 2149–2154.

(6) Schnölzer, M.; Kent, S. B. H. Constructing Proteins by Dovetailing Unprotected Synthetic Peptides: Backbone-Engineered HIV Protease. *Science (80).* **1992**, *256* (5054), 221–225.

(7) Dawson, P. E.; Muir, T. W.; Kent, S. B. H. Synthesis of Proteins by Native Chemical Ligation. *Science (80-.).* **1994**, *266* (925), 776–779.

(8) Blanco-Canosa, J. B.; Dawson, P. E. An Efficient Fmoc-SPPS Approach for the Generation of Thioester Peptide Precursors for Use in Native Chemical Ligation. *Angew. Chemie - Int. Ed.* **2008**, *47* (36), 6851–6855.

(9) Camarero, J. A.; Hackel, B. J.; De Yoreo, J. J.; Mitchell, A. R. Fmoc-Based Synthesis of Peptide α-Thioesters Using an Aryl Hydrazine Support. *J. Org. Chem.* **2004**, *69* (12), 4145–4151.

(10) Bang, D.; Pentelute, B. L.; Kent, S. B. H. Kinetically Controlled Ligation for the Convergent Chemical Synthesis of Proteins. *Angew. Chemie - Int. Ed.* **2006**, *45* (24), 3985–3988.

(11) Smith, G. Filamentous Fusion Phage: Novel Expression Vectors That Display Cloned Antigens on the Virion Surface. *Science (80-.).* **1985**, *228* (4705), 1315–1317.

(12) Yamagishi, Y.; Shoji, I.; Miyagawa, S.; Kawakami, T.; Katoh, T.; Goto, Y.; Suga, H. Article Natural Product-Like Macrocyclic N -Methyl-Peptide Inhibitors against a Ubiquitin Ligase Uncovered from a Ribosome-Expressed De Novo Library. *Chem. Biol.* **2011**, *18* (12), 1562–1570.

(13) Ng, S.; Derda, R. Phage-Displayed Macrocyclic Glycopeptide Libraries. *Org. Biomol. Chem.* **2016**, *14* (24), 5539–5545.

(14) Varjosalo, M.; Taipale, J. Hedgehog: Functions and Mechanisms. *Genes Dev* **2008**, *18*, 2454–2472.

(15) Maun, H. R.; Wen, X.; Lingel, A.; De Sauvage, F. J.; Lazarus, R. A.; Scales, S. J.; Hymowitz, S. G. Hedgehog Pathway Antagonist 5E1 Binds Hedgehog at the Pseudo-Active Site. *J.*

Biol. Chem. **2010**, *285* (34), 26570–26580.

(16) Stanton, B. Z.; Peng, L. F.; Maloof, N.; Nakai, K.; Wang, X.; Duffner, J. L.; Taveras, K. M.; Hyman, J. M.; Lee, S. W.; Koehler, A. N.; Chen, J. K.; Fox, J. L.; Mandinova, A.; Schreiber, S. L. A Small Molecule That Binds Hedgehog and Blocks Its Signaling in Human Cells. *Nat. Chem. Biol.* **2009**, *5* (3), 154–156.

CHAPTER 1:

Synthesis of MeDbz linker derivatives

INTRODUCTION

The chemical synthesis of proteins was first envisaged by Fischer in the early 1900, a century before the entrance into proteomics.[1] Despite the efforts carried out by different scientist on that time, it was not until years later in 1954 when Vincent du Vigneaud reported the first chemical synthesis of oxytocin, a 9-mer protein.[2] This achievement encouraged organic chemists for the synthesis of other and larger proteins, most of them focusing their researches on the synthesis of insulin.[3] A remarkable advance in peptide synthesis was reported in 1963 by Merrifield,[4] who described the synthesis of polypeptides on a solid support. Unfortunately, due to aggregation and side reactions, the synthesis of pure peptide with high purity on solid support was limited to peptides no longer than 30 aminoacids. Even until the moment, despite the use of automated microwave synthesizers, the synthesis of pure peptides longer than 50 aminoacids become troublesome.

The development of other synthetic strategies to achieve the chemical synthesis of proteins or longer peptides was still required. First approaches envisaged the synthesis of proteins as the assembling of fully protected peptides (**Figure *1***),[5] but the poor solubility of protected peptides and the risk of epimerization on the C-terminal aminoacid limited the preparation of proteins using this route.

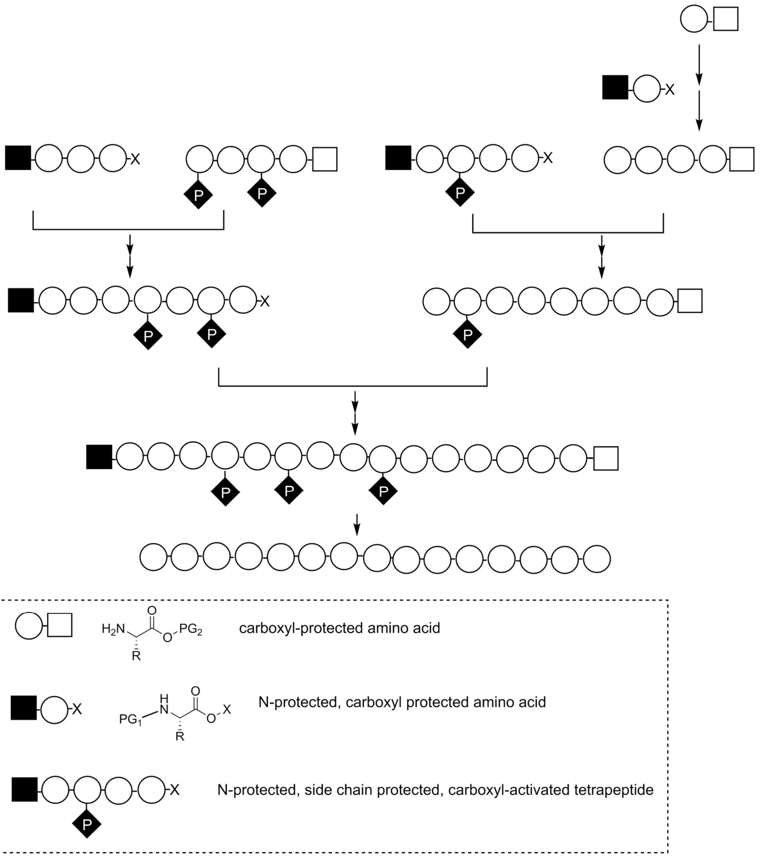

X
P
H₂N
O
O
PG₂
R
carboxyl-protected amino acid
PG₁
H
N
O
O
X
R
N-protected, carboxyl protected amino acid
N-protected, side chain protected, carboxyl-activated tetrapeptide

Figure 1. Peptide fragment condensation.[6]

Other strategies that would not require the fully protection of peptides would represent a remarkable advantage in protein preparation. Moreover, there was also a need to develop an approach that would use milder reaction conditions and thereby, avoid to epimerization problems.

1. Chemical Ligation

In 1992 when Kent reported a new approach for the synthesis of proteins, called Chemical Ligation. The key feature of this strategy was a nucleophilic chemoselective reaction between a thiocarboxylate present at the tail of one peptide and an alkyl bromide at the terminus of the other fragment (**Figure *2***). The resulting assembled polypeptide contained a thioester bond instead of a natural amide bond.[7] The high selectivity of this reaction made it possible to work with unprotected peptides, what dramatically improved their solubility and therefore overcame one of the main drawbacks described until the moment. Using this approach Kent and colleagues reported the synthesis of the "backbone-engineered" protein HIV-1.[7] This synthetic protein displayed the same activity as the natural one, thus assessing that the presence of a thioester bond substituting a natural polypeptide bond did not interfere dramatically on the protein properties.

This chemoselective reaction was used for Kent and Milton to synthetize the mirror HIV-1 bearing D amino acids.[8] This achievement stated the chemical ligation as a powerful methodology to access "backbone-engineered" but also allows the synthesis of mirror proteins and enzymes that proved to be fully active.

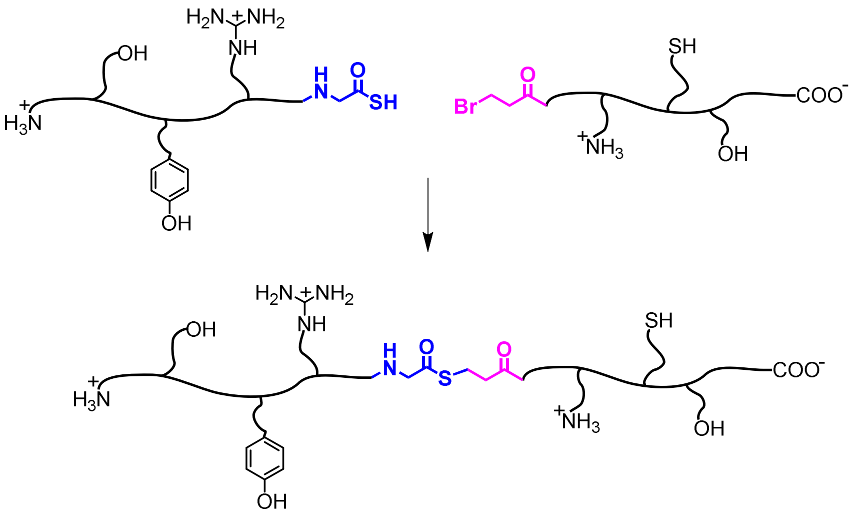


Figure 2. Chemoselective ligation described by Kent et al. [7]

2. Native Chemical Ligation

Nevertheless, it was questioned whether these “backbone-engineered” proteins could be used as models to study protein properties and interactions, since strictly speaking, the presence of thioester bonds may alter protein function or interactions in some cases. So, despite the fact that the synthesis of HIV-1 was a great success in protein chemistry field, there was still the need for an approach that would allow the chemical synthesis of natural proteins.

In 1994 Dawson and Kent reported the concept of Native Chemical Ligation (NCL), described as the chemoselective reaction between two unprotected peptide fragments to obtain a final polypeptide containing a natural amide bond.

2.1. Principle of NCL

NCL was an extension of chemical ligation and take advantage of the known *S* to *N* acyl shift. So, in fact, this chemoselective reaction takes place between a thioester moiety tailored at the C-terminus of one peptide and the thiol group from the N-terminal cysteine of the other peptide fragment. The reaction takes place in two steps, being the first a thiol-thioester exchange between the C-terminal thioester and the thiol group from the cysteine. Next, there is a *S* to *N* rearrangement between the amino group from the cysteine and the thioester formed, resulting in the formation of an amide bond (**Figure *3***). The first step of NCL, the thiol-thioester exchange is reversible, whereas the subsequent rearrangement is irreversible. This last step involving the amide bond formation shifts the equilibrium of the thiol-thioester exchange towards the next irreversible step, promoting the complete ligation between the fragments.[9]

Like chemical ligation, NCL occurs between unprotected peptides under mild reaction conditions in aqueous buffer. The denaturing 6 M guanidine hydrochloride, 0.2 M phosphate buffer helps to solubilize the peptide fragments, thus avoiding solubility problems and limitations due to peptide length. Another significant advantage of using NCL is that racemization is negligible, being less than 1%.[10]

Unlike chemical ligation, in NCL the use of an aryl thiol additive is required in order to increase the ligation rate of the ligation by converting the alkylthioester into a more reactive aryl thioester.

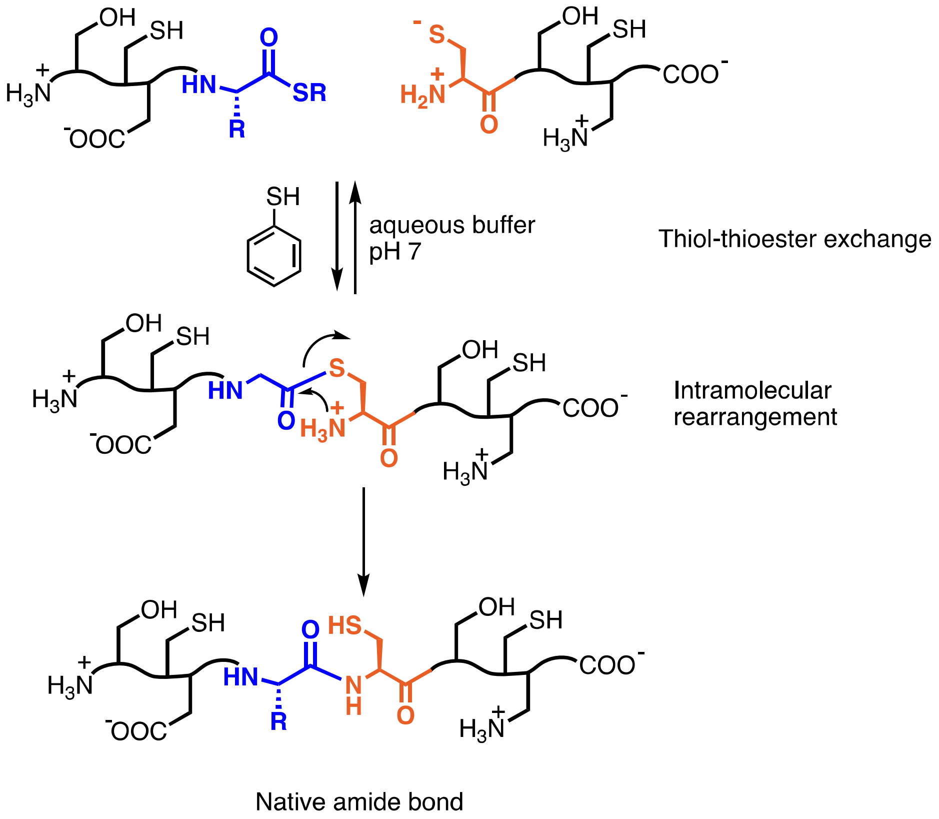


Figure 3. Scheme of NCL mechanism.[6]

2.2. Limitations of NCL

NCL is to be a very useful strategy for the synthesis of proteins, but also face some limitations. For instance, the requirement of a cysteine at an optimal location within the protein for NCL or the preparation of C-terminal thioester peptides becoming the most challenging features of this new approach.

Cysteine is a low frequency residue in proteins (1.48 %), and around the 8% of human proteins do not even contain this aminoacid. Moreover, the presence of a cysteine residue in a protein is not always in a suitable position for NCL regarding fragment length or aminoacid junction. Contrary, alanine is much more frequent in proteins (8.42 %) therefore increasing the possibilities of finding it in an optimal location in a protein sequence.[11] Taking advantage of these considerations and in order to extent the use of NCL reaction to proteins lacking cysteine residues, Dawson *et al.* described the use of desulfurization conditions on the final ligated polypeptide to convert the cysteine introduced for NCL into the natural alanine. This approach broadened the use of NCL into alanine site [12]. Following this strategy, several groups translated the idea into other aminoacids, thus adding a mercapto group and removing it after ligation allowed NCL to occur independently on the original residue present at the N-terminus [13,14] (***Figure 4***).

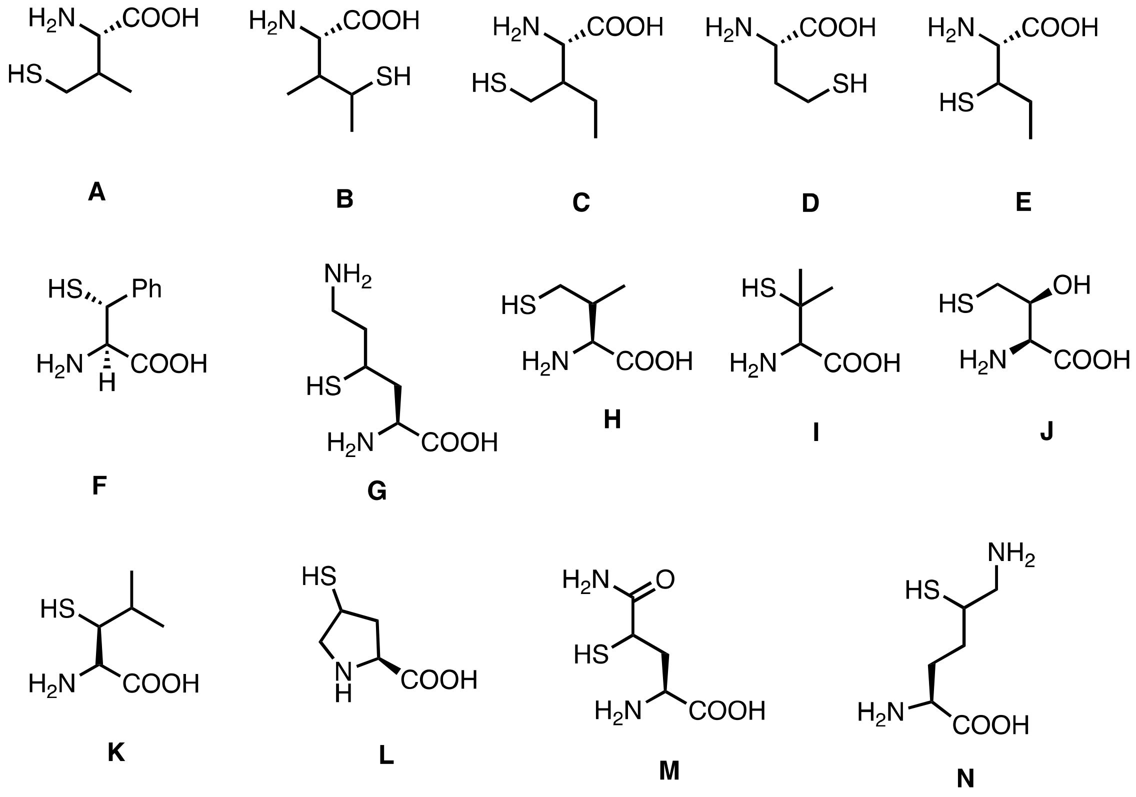


Figure 4. Example of non-natural amino acids that can be used as potential ligation site in NCL. After desulfurization they will give A, valine; B and C Isoleucine; D, aminobutyric acid; E, leucine; F, phenylalanine; G, lysine; H, valine; I, leucine; J, threonine; K, valine; L, proline; M, asparagine; N, lysine.[12,14]

Another approach to become independent of the presence of cysteines in the protein was introduced by Hondal et al.[15] who described the use of selenocystine in NCL. They demonstrated that selenocysteine could be successfully used in NCL giving the desired ligated product. Actually, they demonstrated that selenocysteine reacts faster than cysteine, and therefore selective ligations would be feasible. Besides, deselenization conditions to generate an alanine at the ligation point had been reported. These conditions allow the selective removal of the seleno atom in the presence of thiol groups, thus avoiding extra protecting steps on the cysteine side chain.[16]

3. Synthesis of C-terminal thioester peptides

The preparation of C-terminal thioester peptides is easily accessible by Boc SPPS. Nevertheless, the repeated use of strong acid treatment and the hard cleavage conditions impairs the synthesis of glyco or phosphopeptides using Boc chemistry. In addition, automated synthesizers had been optimized for Fmoc-SPPS, because the reagents used in this chemistry are milder than the strong and corrosive TFA required for Boc-SPPS.

Due to these limitations, the synthesis of C-terminal thioester peptides by Fmoc-SPPS came up as an alternative. However, the fact that thioesters are sensitive to basic conditions used for Fmoc removal made their synthesis challenging.

Fortunately, the efforts made in this field until the moment, allow nowadays the preparation of C-terminal thioester peptides by Fmoc-SPPS. Several strategies, described below, were reported to directly prepare peptide thioester or precursors using Fmoc chemistry.

3.1. Direct synthesis of C-terminal thioester on resin using Fmoc-SPPS

The first strategies to prepare thioesters by Fmoc chemistry focused on the development of special Fmoc removal cocktails that would minimize the aminolysis of the thioester. For example, Li *et al.* developed a mixture of 2% hexamethylene imine, 2% HOBt and 25% 1-methylpyrrolidone (NMP) in dimethylsulfoxide (DMSO).[17] This cocktail mixture was successfully used for the synthesis of a glycopeptide by Hojo *et al.*[18] Other mixtures to remove Fmoc group using less nucleophilic basic conditions were reported.[19] However, in all these strategies aminolysis of the thioester occurred during the first deprotection steps, yielding with the loss of the peptide.

3.2. Thioesterification of fully protected peptide acids

Another approach towards the synthesis of C-terminal thioester peptides was the activation of the peptide acid once elongated and cleaved from the resin. Activation step was carried out by activating reagents, however, the risk of epimerization during this step was one of the main drawbacks of this methodology. Also, the need of fully protected peptides during activation step arose with solubility problems. However, some examples of proteins synthetized using this approach were reported (

Figure **5** and **Figure *6***), for instance the synthesis of glycopeptide cluster as a part of a vaccine for the treatment of the ovarian carcinoma.[20] Other examples using PyBOP and DIEPA as activators were reported for the synthesis of the pro-neuropeptide Y[21] and in the synthesis of the cyclic peptide CVO2.[22]

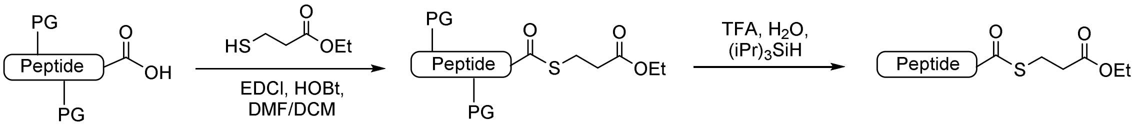


Figure 5. Activation of acid peptide using EDCl and HOBt. Adapted from[20] PG= protecting group.

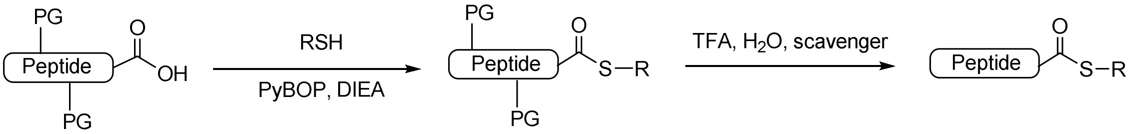


Figure 6. Activation of acid peptide using PyBOP and DIEA. Adapted from[21,22]. PG= protecting group; R= alkyl or aryl moiety.

3.3. Conversion of resin-bound peptide allyl esters into peptide thioesters

To overcome the need of working with protected peptides in solution, several strategies were evolved to carry out the thioesterification steps on resin.

For instance, a peptide is anchored to the resin via a side chain while maintaining protected the α-carboxilic acid. After peptide elongation, the acid is deprotected and thioestererification is carried out (**Figure *7***). This strategy requires high activation conditions, which lead to epimerization.[23] Despite this disadvantage, the synthesis of some proteins using this methodology had been described, like the synthesis of EPO. [24]

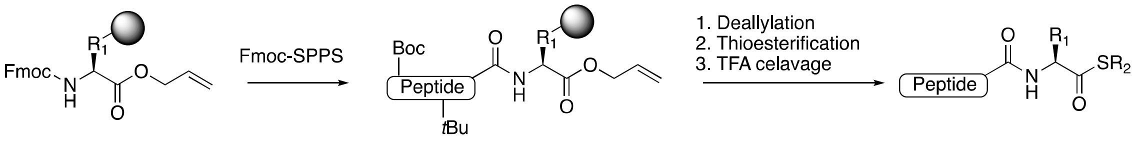

Fmoc
R1
Fmoc-SPPS
Boc
Peptide
tBu
1. Deallylation
2. Thioesterification
3. TFA celavage
SR2

Figure 7. On-resin thioesterification of side chain anchored allyl ester peptides. [23] R_1 = amino acid side chain; R_2 = aryl or alkyl moiety.

3.4. BAL strategy

Unlike the strategy described before, in the BAL approach the peptide is anchored to BAL linker by the amino group of the α-amino acid instead of the side chain. This difference overcome the sequence limitation allowing any residue at the C-terminal position.[25] Like the previous strategy using allyl ester peptides, the carboxylic group remains protected during peptide elongation (**Figure *8***) and thioesterification requiring high activation conditions take place after selective deprotection of the C-terminal acid. Apart from the risk of epimerization, this approach also faces the challenging peptide elongation on secondary amines.

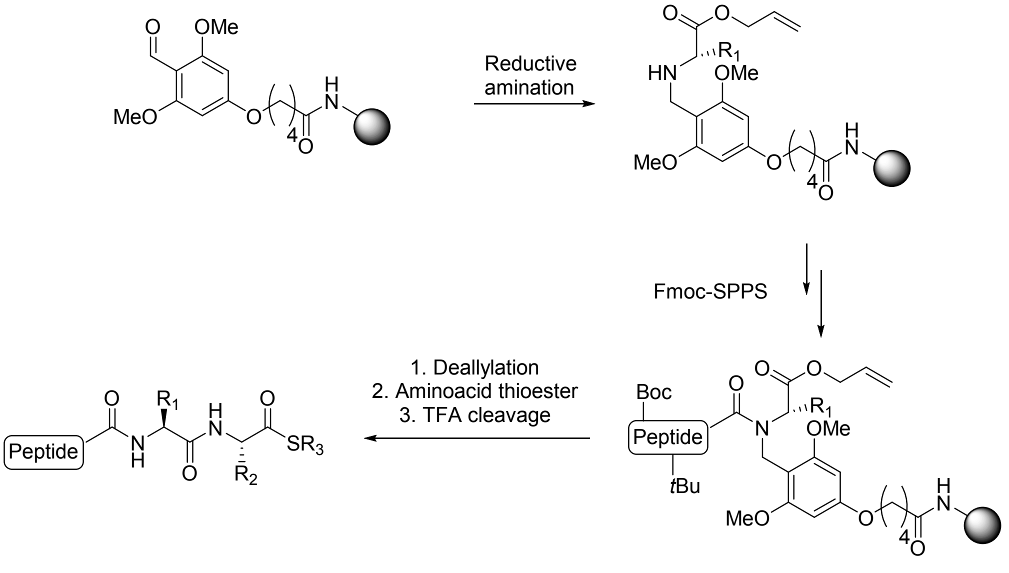

OMe
MeO
Reductive amination
R1
Fmoc-SPPS
1. Deallylation
2. Aminoacid thioester
3. TFA cleavage
Boc
Peptide
tBu
SR3
R2

Figure 8. BAL strategy for the obtainment of C-terminal thioesters.[23] R_1 = amino acid side chain; R_2 = amino acid side chain; R_3 = alky or aryl moiety.

3.5. Safety catch linker using Kenner's sulfonamide linker

One of the most used approaches for the synthesis of thioesters is the Kenner's sulfonamide safety catch linker. This strategy relies on the use of Kenner's linker which, after peptide elongation, is activated through alkylation and after thiol nucleophilic attack renders the fully protected thioester peptide[5] (**Figure *9***). Although this methodology was widely used, it was observed epimerization and side reactions during linker activation [26]. In the literature several examples using this approach to obtain C-terminal thioesters for the synthesis of proteins have been reported such as the synthesis diptericin, a 82-mer glycoprotein [27]. This synthetic strategy has become one of the most used methodologies for peptide thioester synthesis until very recently.

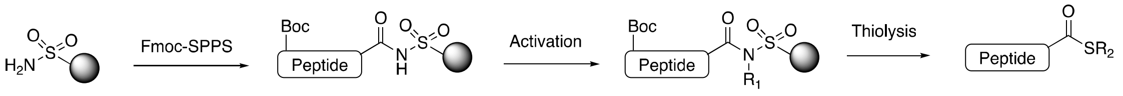


Figure 9. Safety catch linker for obtainment of C-terminal thioesters.[23] R_1 = Methyl from alkylation with TMS-CHN_2 or ICH_2CN. R_2 = aryl or alkyl moieties.

3.6. SPPS of peptide thioesters at aryl hydrazide linker

Similar to Kenner's safety catch linker, the aryl hydrazide linker is also suitable for the synthesis of C-terminal peptide thioesters by Fmoc/^{t}Bu chemistry, giving the protected peptide-thioesters upon thiolysis of the resin. Differently to the sulfonamide linker, in this case the activation of the hydrazine linker occurs by oxidation which may become troublesome in the presence of susceptible residues like methionine[28] (**Figure *10***). Some proteins, like the Src homology 3 (SH3) protein domain, were successfully synthetized using this linker approach.[28]

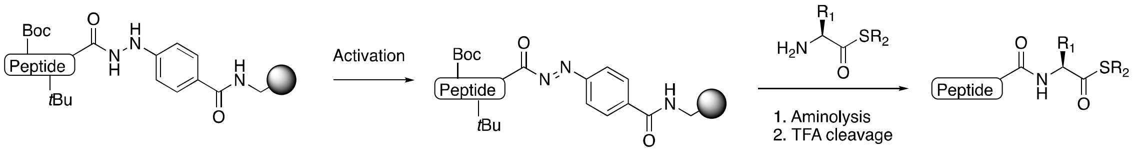


Figure 10. Hydrazine linker used for the synthesis of C-terminal peptide thioesters.[23] R_1 = amino acid side chain; R_2 = aryl or alkyl moiety.

Recently, a novel approach which renders the protected peptide thioester after on-resin thioesterification has been reported by Elashal *et al.* In their work, inspired from Jensen's approach,[29] the thioester is obtained from a cyclic urethane formed in a serine residue on resin.[30] Thioesterification occurs on resin, just cleaving the corresponding protected peptide, thus avoiding further purification steps (**Figure *11***).

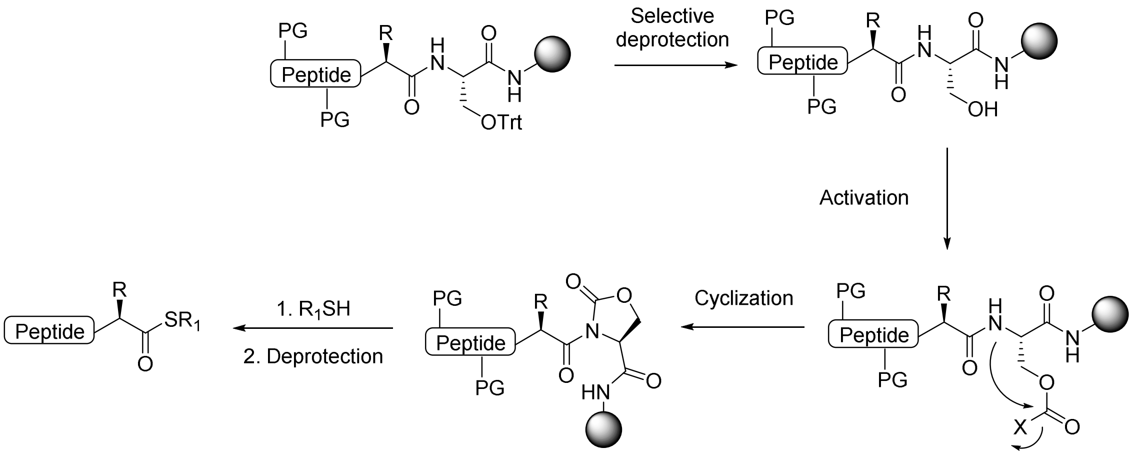

Selective
deprotection
Activation
Cyclization
1. R_1SH
2. Deprotection
Peptide
PG
R
OTrt
OH
X
SR_1

Figure 11. Activation of serine side chain to render a cyclic uretane susceptible for thioesterification on resin. Adapted from [30]. R = amino acid side chain; R_1 = alkyl or aryl moiety; X = electrophile.

3.7. SPPS of thioesters through O -> S and N -> S acyl shift

Among the below reported approaches, the access to C-terminal thioesters by intramolecular *O* or *N* to *S* acyl shift became one the most promising. The basis of these strategies is the intramolecular shift that takes place between a thiol group present and an amide or ester bond. This intramolecular rearrangement would render the final thioester peptide.

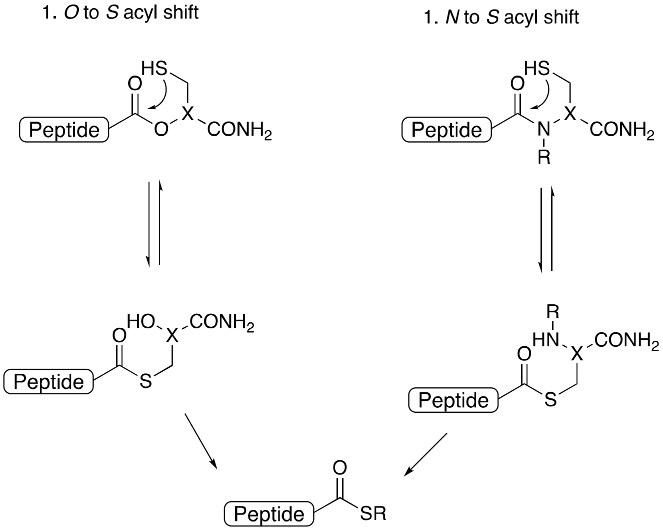

1. *O* to *S* acyl shift
1. *N* to *S* acyl shift
Peptide
HS
X
$CONH_2$
HO
HN
R
S
SR

Figure 12. Principle of N and O to S acyl shift for preparation of C-terminal peptide thioesters.[23]

This idea was adopted for several groups, resulting in the emergence of a wide variety of methodologies to afford the desired thioesters[23] (**Figure *12***). Since 2013 the strategies undergoing *N* to *S* acyl shift have overwhelmed *O* to *S* acyl shift.[11] The most novel example of this methodology have been recently reported by Terrier *et al.* for the synthesis of Big Defensin 1 protein[31] In their work, they describe the use of a new device based on a *N*-(2-hydroxybenzyl) cysteine which under NCL conditions undergoes a rapid *N* to *S* shift yielding the corresponding thioester.

3.8. Peptide α-N-acyl-benzimidazolinones as precursors for peptide α-thioesters

In 2008 Blanco-Canosa and Dawson reported a novel approach in which the 3,4-diaminobenzoic acid linker was used as a precursor for peptide thioester synthesis.[32] In their work, the 3,4-diaminobenzoix acid was first attached to the resin, and used as a linker. Then, the first residue was coupled on one of the amines using controlled coupling conditions to avoid double incorporation in both of the amines. After peptide elongation, the free amine of the linker was acylated using *p*-nitro-phenylchloroformate, which underwent intramolecular cyclization to afford the corresponding *N*-acyl-benzimidazolinone (Nbz) peptide. Once the peptide is cleaved from the resin, and also fully deprotected, the Nbz group is susceptible to thiolysis, either directly in NCL or in the presence of thiols (**Figure *13***). This new methodology overcame epimerization observed in other approaches since it does not require high activation of the residue at the C-terminus. This approach was widely used for the synthesis of several proteins. One of the last examples reported was the synthesis of Src SH2 domain, by Keitou *et. al*.[33] Also the synthesis of the Human Growth Hormone, a 191-mer protein was achieved adopting this strategy.[34]

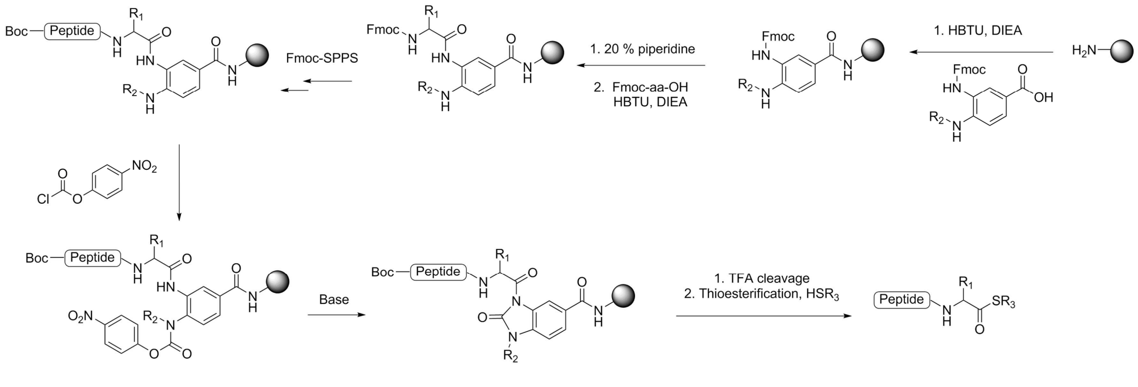


Figure 13. Synthesis of C-terminal peptide thioesters using the Nbz approach. R_2 = H, Me (second generation *N*-acylurea linker); R_3 = aryl or alkyl moiety. Adapted from [32].

Although this methodology was successfully used for several syntheses, it was also noticed that its use in high glycine rich sequences resulted in branched peptides due to the coupling on the *O*-aminoanilide amine of the linker. In order to overcome this disadvantage, later in 2015, Blanco-Canosa *et al.* described a new derivative from the 3,4-diaminobenzoic acid, in which one amine was methylated, making both of the amines different in terms of reactivity [35](**Figure 13**). The new linker displayed major robustness during peptide elongation even under hard coupling conditions.

3.9. Peptide O-aminoanilides as cryptothioester

Based on the use of the 3,4-diaminobenzoic acid to perform ligations in the N to C direction, Liu and coworkers proposed a new strategy in which the use of $NaNO_2$ gives an activated benzotriazole peptide in solution, which can undergo directly NCL.[36] The peptides are inert prior activation, which motivated to name them as crypto-thioester peptides. The activation using $NaNO_2$ takes places in ice bath at -10 °C for 20 minutes at pH 3, and a posterior addition of 4-mercaptophenylacetic acid (MPAA) adjusting pH to 6.9 initiates of the NCL reaction (**Figure *14***). The epimerization observed in this method is in agreement with the reported for the Nbz method, being less than 1 %.

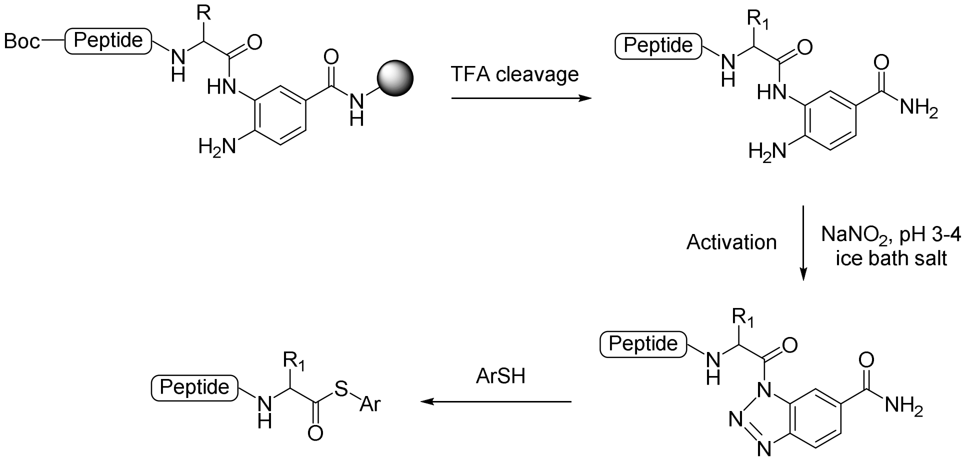


Figure 14. Preparation of C-terminal peptide thioesters from benzotriazole approach. Adapted from [36].

The fact that this methodology allows the activation of the peptide in solution represents an advantage to the Nbz methodology, in which the peptides are already susceptible to thioesterification without the need of prior activation.

Among all the strategies and approaches reported for the synthesis of C-terminal thioesters, the most successful one is the hydrazide ligation reported first by Liu *et al.*[37] This approach involves the chemoselective reaction between a C-terminal hydrazide peptide and the N-terminal cysteine from another fragment. The great advantage of this methodology is the easy preparation of hydrazides either by Boc or Fmoc/^{t}Bu SPPS; moreover, hydrazide peptides could be also obtained from bacteria. This approach is composed by two steps: the first one consisting on the oxidation of the hydrazide peptide into an azide upon the addition of $NaNO_2$ at -10ºC and pH 3; the second part consists on the addition of the thiol and pH adjustment to 7 which would initiate the ligation. Similar to NCL, racemization was observed in less than 1%.[37] An advantage to NCL is that hydrazide ligation allows the assembling of peptides on the N to C direction.

3.10. SEA approach as crypto-thioesters

Other strategies describing the development of new crypto-thioesters to allow assembling of peptides in the N to C direction have been reported in the last years. Ollivier and co-workers reported the use of bis(2-sulfanylethyl)amino group named SEA group and its selectively activation upon ligation conditions.[38] The SEA^{off} is inactive due to the disulfide bridge between the thiol residues, but in the presence of reducting agents such as TCEP, the disulfide bridge is reduced (SEA^{on}) and the amide undergo intramolecular rearrangement rendering a C-terminal thioester. The resulting peptide can undergo further ligations after conversion into the corresponding thioester (**Figure *15***). This approach was used for the synthesis of the K1 domain of Hepatocyte Growth Factor.

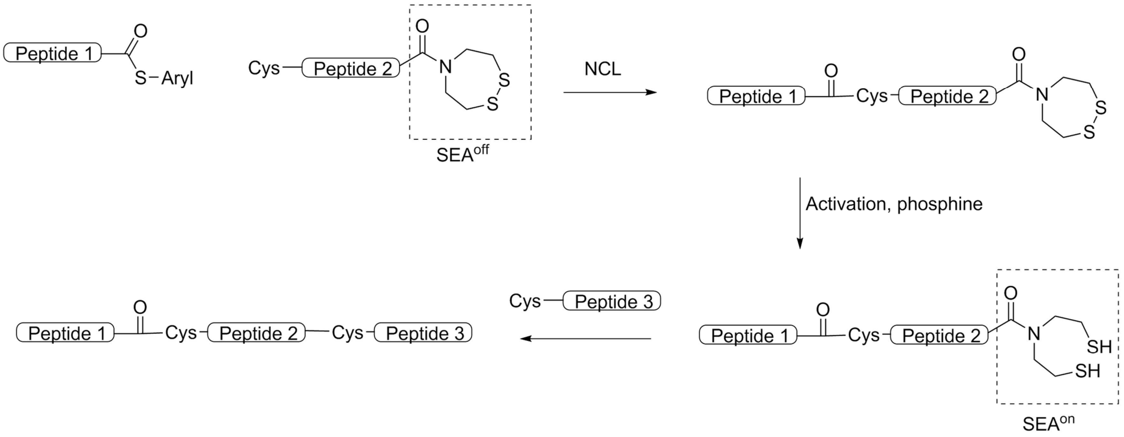


Figure 15. Peptide assembly on the N to C direction using the SEA group. Adapted from [38].

3.11. SEAlide as cryptothioesters

Another crypto-thioester relying on the use of *N*-sulfanylethylanilide (named SEAlide), was reported by Otaka *et al.* in 2013.[39] In this novel methodology (which mechanism remains unknown) the selective activation of SEAlide into a thioester is carried out upon the addition of phosphate salt (**Figure *16***).

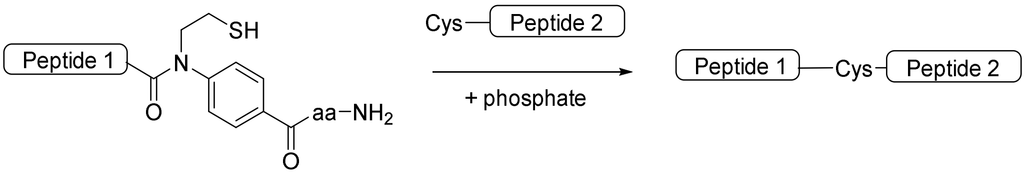


Figure 16. N-sulfanylethylanilide as cryptothioester in presence of phosphate salts give NCL product.

Bearing the idea of crypto-thioester peptides, also known as masked or latent thioesters, part of the work conducted in the present thesis was focused on the generation of Nbz peptides in solution, and therefore making them suitable for the use in synthesis to the N to C direction.

The use of new methodologies for NCL, like Nbz, N to O or N to S acyl shift or hydrazide ligations, have increased since 2010 in detriment of conventional thioesters. Actually in 2016 these novel approaches represented the 54% of the publications regarding chemical synthesis of proteins. Among them, it seems that hydrazide ligation is taking the major role (30%), while the Nbz ranks the second position being reported in the 13% of the publications.[11]

4. Kinetic controlled ligation

NCL allowed the access to several synthetic proteins; however, most of the synthesis reported had been achieved by a single NCL between two peptide fragments and therefore the proteins synthetized were 80-mer on average. Larger proteins required a different synthetic approach than sequential ligations on the C to N direction and temporal protection of cysteine. To facilitate convergent synthesis and avoid the cysteine protection, it was developed KCL. The idea of a convergent approach called Kinetic Controlled Ligation (KCL), was reported by Kent and Pentelute in 2006.[40]

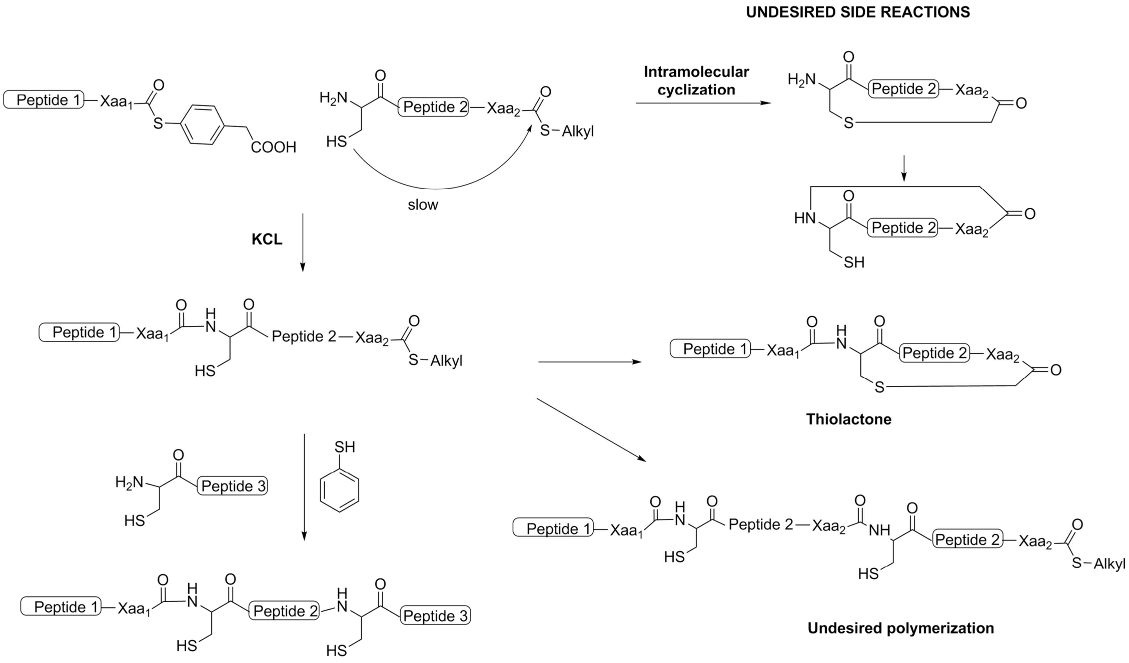
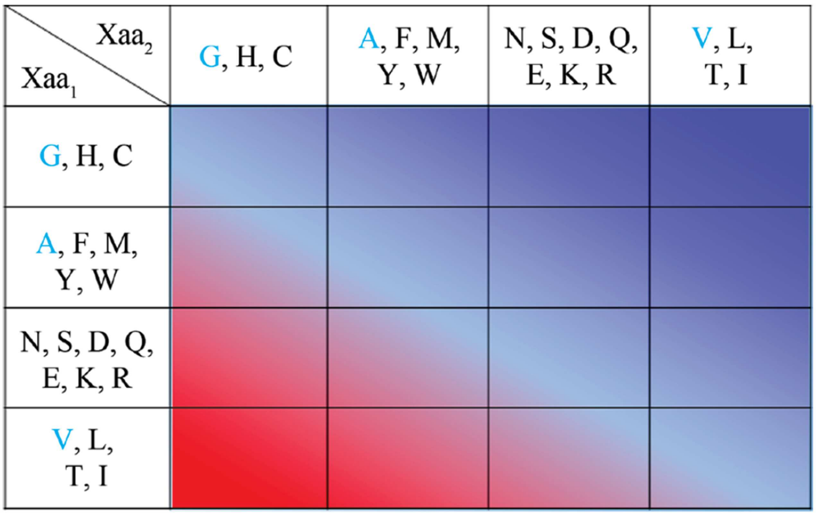


Figure 17. Kinetic Controlled Ligation scheme. Adapted from [41].

Sequential NCL ligations require a peptide with dual reactivity, bearing a cysteine at the N-terminal extreme and a thioester at the C-terminus, which needs total control on the reactivity. In their work, Kent and Pentelute first hypothesized that alkyl and arylthioesters differed in reactivity, expecting the arylthioester to undergo faster ligation than the alkyl. This would also avoid intramolecular cyclization of the cysteine peptide and oligomerization. To confirm this theory, the first KCL ligation between an arylthioester peptide and a cysteine peptide with C-terminal alkylthioester was performed rendering in majority the ligated product (**Figure *17***). This approach was confirmed by the convergent synthesis of protein crambin.[40].

In a more detailed study into KCL mechanisms, Lee *et al.* pointed out that also the amino acids present at the C-terminus of both peptides (Xaa_1 and Xaa_2 in figure 14) played an important role in the ligation [41]. After a wide screening of different residues occupying the C-terminal positions of both peptides they concluded that the kinetics of KCL depends mainly on the C-terminal amino acid (Xaa_1 in **Figure *14***) of the arylthioester peptide (**Figure *18***). Thereby, glycine and alanine offered faster ligations than branched peptides. However, the amino acid present at the C-terminus of the cysteine peptide (Xaa_2 from **Figure *14***) is also implied in the kinetics of the ligation. In this position, high hindered residues prevent the intramolecular cyclization whereas alanine or glycine are highly prone to undergo this side reaction, thus leading to an unproductive fragment.[41]



Xaa_2 / Xaa_1	G, H, C	A, F, M, Y, W	N, S, D, Q, E, K, R	V, L, T, I
G, H, C	🟦	🟦	🟦	🟦
A, F, M, Y, W	🟥	🟦	🟦	🟦
N, S, D, Q, E, K, R	🟥	🟥	🟦	🟦
V, L, T, I	🟥	🟥	🟥	🟦

Figure 18. KCL reactions combining different amino acids at the Xaa_1 and Xaa_2 positions. Blue color shows efficient combination of Xaa_1 and Xaa_2 residues to obtain the desired KCL product. Red color represent Xaa_1 and Xaa_2 combinations that displayed poor KCL efficiency[41]

Since KCL was described, the synthesis of proteins by a convergent approach increased, allowing access to larger chemical synthetized proteins. In addition, the methodologies based on the use of thioester precursors or crypto-esters and their activation in solution became an indispensable tool for these syntheses.

5. Ligations inspired by NCL

5.1. Conformational assisted ligation

After NCL was reported, several groups focus their efforts on the research of other types of ligations to access synthetic proteins. For instance, conformational assisted ligation, reported by Beligere and Dawson showed the efficient ligation between two fragments due to their proximity on the folded state; moreover, they demonstrated that the presence of cysteine was not essential for the ligation to take place.[42]

5.2. Expressed Protein Ligation

A further extension of NCL is Expressed Protein Ligation (EPL). Reported first in 1998 by Muir *et al.*, [43] it consists on the ligation of a recombinant peptide or protein with a synthetic or expressed polypeptide. Normally the synthetic peptide harbors the N-terminal cysteine, whereas the recombinant protein contains the required C-terminal thioester, which is prepared via intein splicing.[44] EPL had become a useful tool for the introduction of post-translational modifications into proteins, labeling and also has applications in protein NMR studies.[45]

A variation from EPL includes the introduction of enzymes, such as subtiligase[46] or sortase,[47] that mediate the ligation of the recombinant protein to synthetic peptides containing a particular aminoacid sequence.

OBJECTIVES

The NCL described by Dawson and Kent required the synthesis of C-terminal thioester peptides which was challenging by Fmoc-SPPS. Right now, the different approaches reported allow the efficient preparation of peptide α-thioesters either by Fmoc or Boc SPPS.

In our group we use the *N*-acylurea approach developed by Blanco-Canosa and Dawson to prepare peptide thioester precursors for NCL. Nevertheless, this methodology can not be used in KCL. We envisaged that new derivatives from the already reported *N*-acylurea methodology could be used for KCL and ligations from the N to C direction.

The main objectives of the present chapter are:

1. The development of a methodology to activate new MeDbz derivatives in solution and afford the corresponding C-terminal peptide thioester.

2. The use of these new MeDbz derivatives for NCL *in situ.*

3. To explore the utility of the MeDbz derivatives synthetized as masked thioesters in KCL.

RESULTS AND DISCUSSION

In our group the synthesis of C-terminal thioester peptides is carried out using the *N*-acylurea approach described by Blanco-Canosa and Dawson.[32] However, as it was pointed out in the introduction, this methodology just allows the synthesis of thioester precursors for its use in NCL, therefore only the assembling of peptides in a sequential C to N direction.

In the last years, there was an increasing interest about the synthesis of proteins by KCL and thereby, a blooming of methodologies to prepare crypto-thioesters that could be used in N to C ligations.

Bearing these ideas in mind, we aimed to develop a new strategy derived from the *N*-acylurea approach that would allow its use in KCL. The main idea was to develop a methodology that would give Nbz peptides in solution, and therefore, they could be generated *in situ* for NCL.

1. Activation of MeDbz peptides in solution

Our goal was the activation of the ***o*-aminomethylanilide linker** (MeDbz) in solution and *in situ* conversion into the ***N*-acyl-*N'*-methylacylurea** (MeNbz) corresponding peptide. In order to do that, it was envisaged the cyclization on solution of ***p*-nitrophenyloxycarbonyl-MeDbz** peptides, abbreviated as ***p*$\mathbf{NO_2}$-Phoc-MeDbz**, which are precursors of the MeNbz peptides. The cyclization on resin occurs under basic conditions with a solution of 0.5 M diisopropylethylamine (DIEA) in DMF. Based on this precedent, we searched for similar conditions to promote the cyclization of the MeDbz linker in aqueous buffer (**Figure *1***).

To confirm this hypothesis, a model peptide LYSAA-MeDbz was first synthetized using the MeDbz linker. After peptide elongation, the linker was acylated with *p*-nitrophenylchloroformate rendering the *p*NO_2-Phoc-MeDbz peptide on resin. Then, the peptide was cleaved using standard cleavage conditions (95% TFA, 2.5% TIS, 2.5% H_2O) to finally obtain the deprotected *p*NO_2-Phoc-MeDbz peptide.

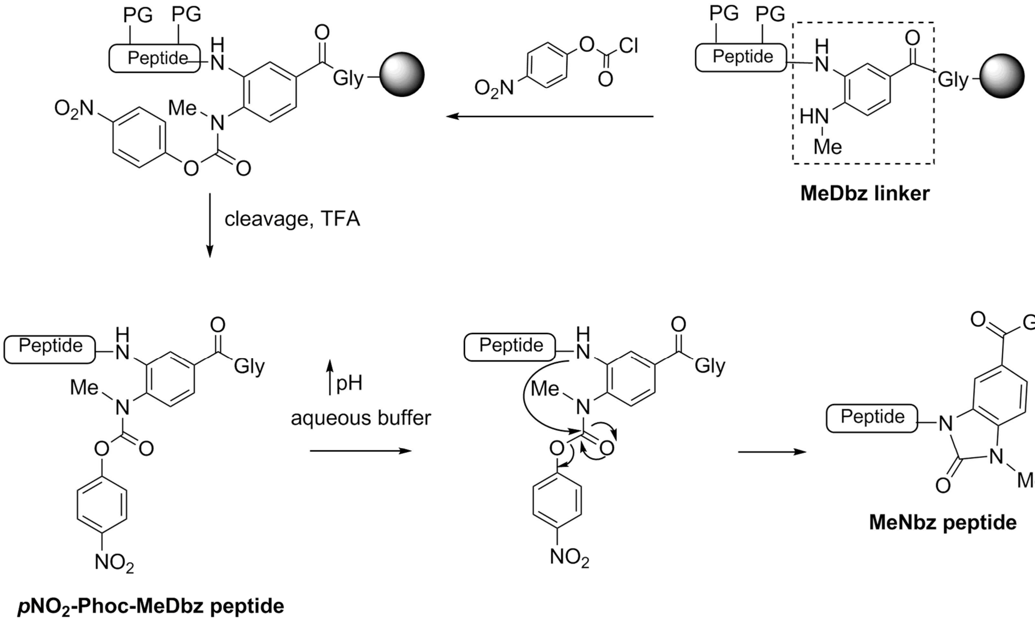


Figure 1. Scheme of hypothesized activation of $p\mathrm{NO_2}$-Phoc-MeDbz peptides in solution to afford MeNbz peptides.

Next, the $p\mathrm{NO_2}$-Phoc-MeDbz peptide was dissolved at a concentration of 2 mM in the **NCL buffer:** 6 M guanidinium hydrochloride, 0.2 M phosphate and 20 mM tris-carboxyethylphosphine (TCEP), containing 100 mM 4-mercaptophenol (MPOH). The pH of the buffer was carefully raised to 7.5 using a $\mathrm{NaOH_{(aq)}}$ solution (1 M) and the reaction was monitored by HPLC. It was anticipated that the thioesterification would be rapid once the MeNbz was formed, and in principle detect only the corresponding thioester peptide. As expected, the thioester peptide was observed after 3 hours of reaction. Unfortunately, apart from the desired product, significant hydrolysis of the carbamate moiety was also observed. This hydrolysis, which renders the **MeDbz peptide** that does not cyclize into the Nbz peptide, was inefficient for further ligation (**Figure *2***).

After a careful analysis of the results, two possibilities were considered for this hydrolysis: the basic pH of the buffer or the the high acid character of the *p*-nitrophenol leaving group which would overactivate the carbamate moiety. In order to identify the main reason for the hydrolysis, we carried out different experiments.

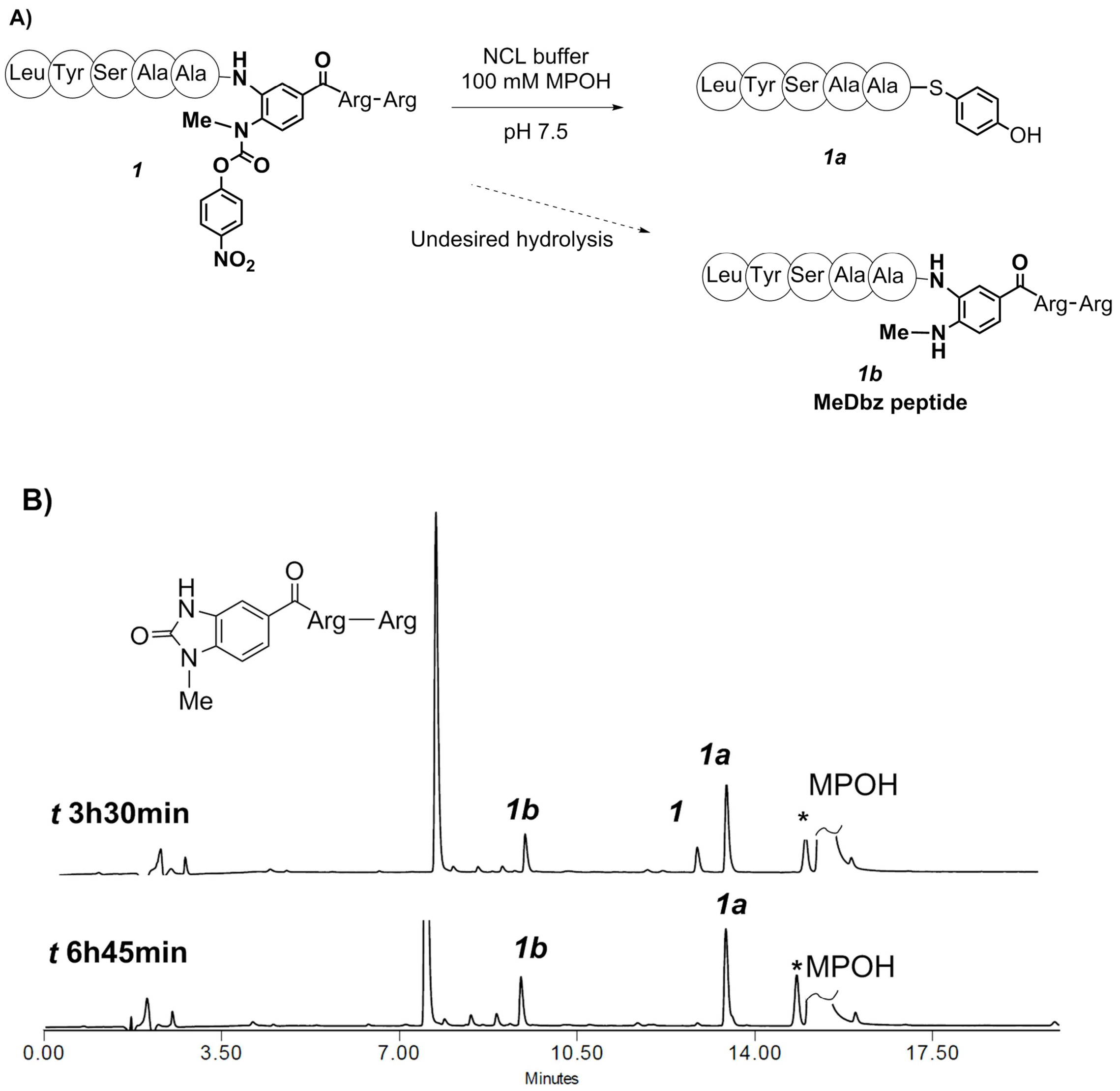

A)
Leu Tyr Ser Ala Ala
H N
O
Arg-Arg
Me N
O O
NO_2
1
NCL buffer
100 mM MPOH
pH 7.5
Undesired hydrolysis
Leu Tyr Ser Ala Ala S
OH
1a
Me N H
1b
MeDbz peptide
B)
O N H N Me
Arg—Arg
t 3h30min
1b
1
1a
MPOH
*
t 6h45min
1b
1a
*MPOH
0.00
3.50
7.00
10.50
14.00
17.50
Minutes

Figure 2. A) Scheme of the activation reaction of a model $p\mathrm{NO_2}$-Phoc-MeDbz peptide in NCL buffer containing 100 mM MPOH at pH 7.5. B) HPLC at different reaction times. (*) = p-nitrophenol.

1.1. Effect of the pH on the hydrolysis of the $p\mathrm{NO_2}$-Phoc-MeDbz peptides

To determine the effect of the pH, a model $p\mathrm{NO_2}$-Phoc-MeDbz peptide was dissolved into NCL buffer containing MPOH, and the pH was maintained neutral to slightly acidic (6.80). The reaction was monitored by HPLC during 6 hours. In addition to the peptide thioester and the starting $p\mathrm{NO_2}$-Phoc-MeDbz peptide, significant hydrolysis on the carbamate moiety was also observed. Actually, although the hydrolysis of the carbamate moiety was independent of the pH, the amount of starting peptide was still significant, which indicated that, the cyclization to MeNbz peptide was slow under this pH conditions and a slightly basic pH was required to accelerate the linker cyclization (**Figure *3***). This experiment confirmed that lower pH did not avoid hydrolysis, and thereby, the hydrolysis is caused by other mechanisms.

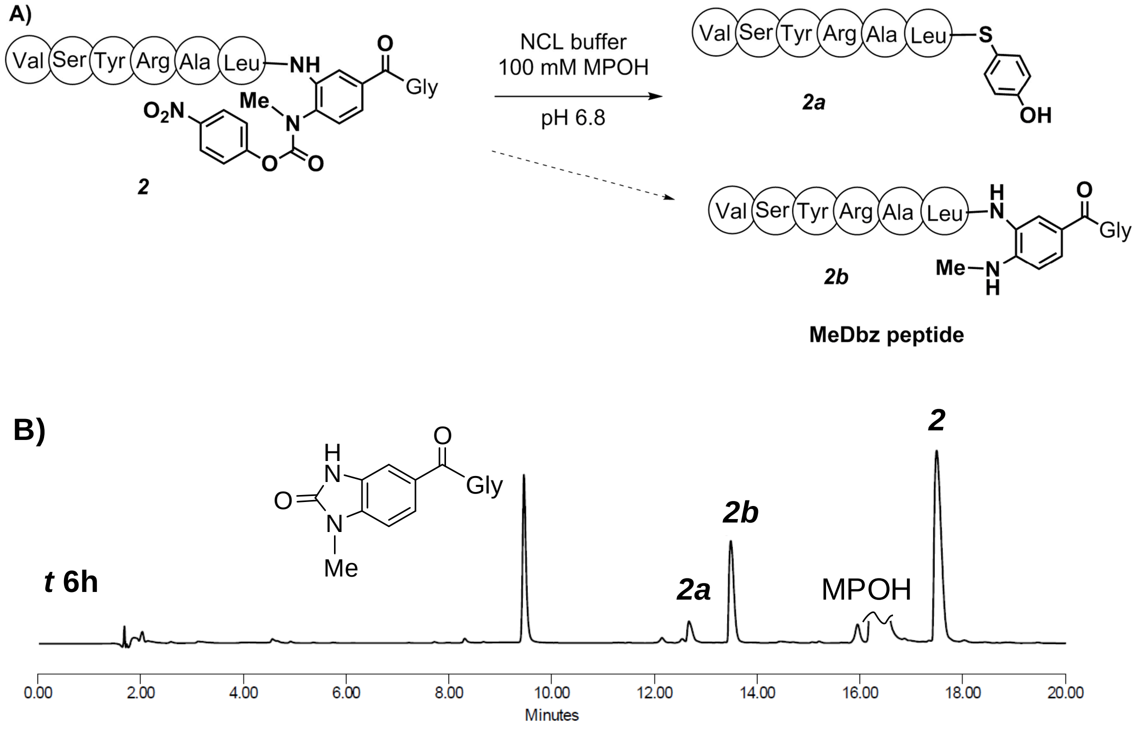


Figure 3. A) Scheme of the activation reaction of a model pNO_2-Phoc-MeDbz peptide dissolved in NCL buffer at pH 6.8. B) HPLC of the reaction after 24 hours.

1.2. Effect of the acidic character of the substituted phenol leaving group on the hydrolysis of the aryloxycarbonyl-MeDbz peptides

The next factor studied was the acidic character of the *p*-nitrophenol group. We searched for other substituents on the phenol ring that endowed different acidity and cover a wide range of pKa values. The selected substituents, encompassing phenol derivatives with pKa values ranging 9.90 to 7.83, are depicted in Table 1.[48] The synthesis of these substituted phenylchloroformates is shown below.

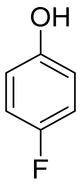
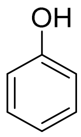
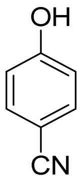
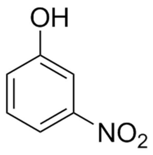
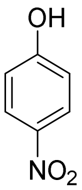
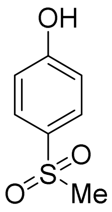
Substituent						
pKa value	9.90	9.88	8.2	8.13	7.9	7.83

Table 1. Substituents on the phenyl ring and their pKa values.[48]

1.2.1. *Synthesis of phenylchloroformates*

The phenylchloroformates were successfully prepared by reaction of the different phenols with triphosgene in dry dichlormethane and pyridine (**Figure *4***).[49]

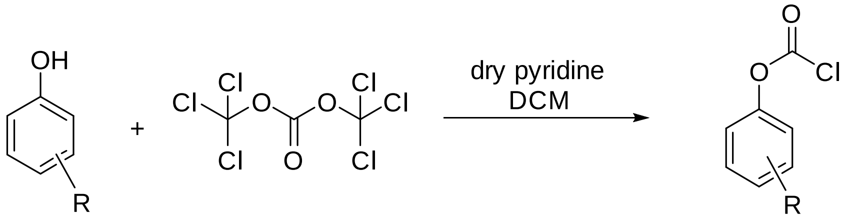


Figure 4. Scheme of the synthesis of phenylchloroformates from the corresponding phenols. R= F, H, CN, NO_2, SO_2Me.

The chloroformates prepared were used to acylate the MeDbz peptides on resin to afford the final *p/m*X-Phoc-MeDbz peptide. The corresponding peptides were dissolved in NCL buffer containing 100 mM of MPOH thiol, and the pH was adjusted to 7.5 using a solution of $NaOH_{(aq)}$ 1 M. The reactions were monitored by HPLC to detect the peptide thioester formation or any hydrolysis.

1.2.2. *mNO2-Phoc-MeDbz*

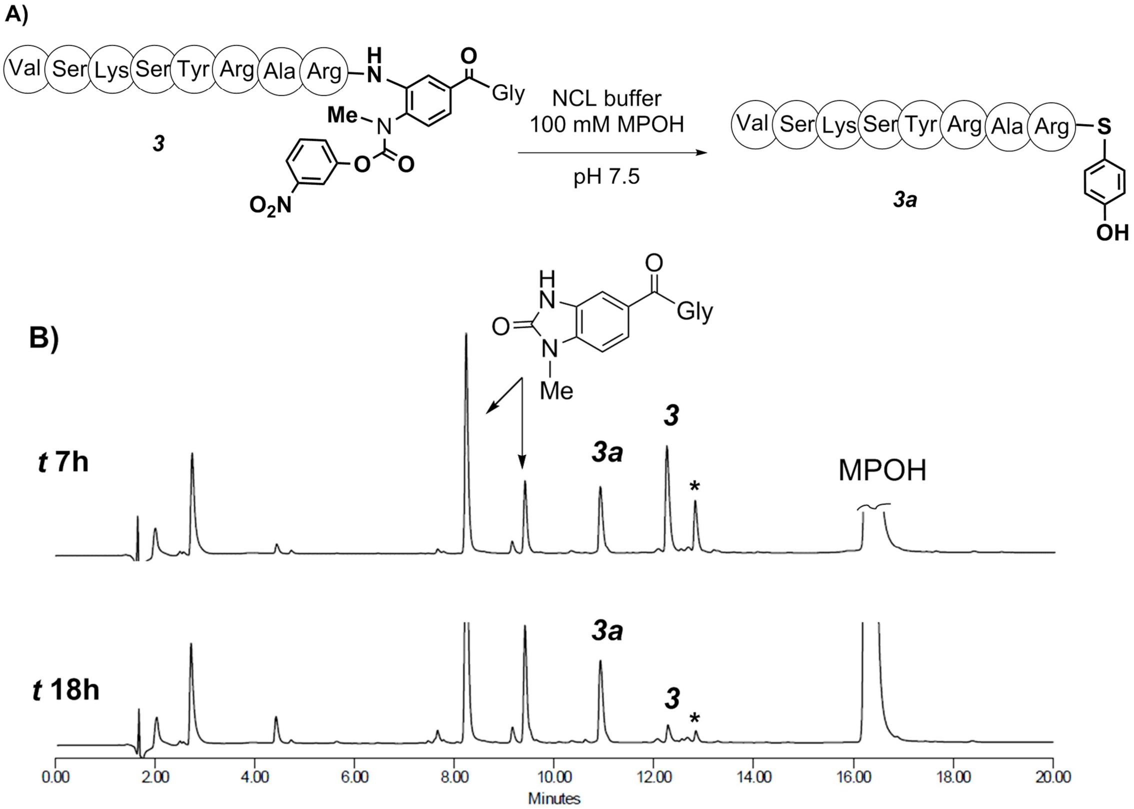


Figure 5. A) Scheme of the reaction of model $m$$NO_2$-Phoc-MeDbz peptide in NCL buffer at pH 7.5. B) HPLC of the reaction at different reaction times. (*) impurity from the synthesis of the peptide.

mNO$_2$-Phoc-MeDbz peptides did not produce any significant or detectable amount of hydrolysis. This absence of hydrolysis could be the result of the increase in the pKa value when the nitro group is in meta position. The mNO$_2$-Phoc-MeDbz appears to be a suitable linker for the generation in aqueous solution of MeNbz peptides (**Figure *5***).

1.2.3. Phoc-MeDbz

Phenol has a pKa value higher than *p*-NO$_2$-phenol . Therefore, it was expected the absence of hydrolysis, but also a slow cyclization rate to the MeNbz. Indeed, after 11 hours the main product was still the starting Phoc-MeDbz peptide (**Figure *6***). Due to the slow rate of MeNbz formation, this linker should have applications for KCL in where the C-terminal amino acid of the MeNbz fragment ligates slowly, such as Val, Ile or Pro residues.

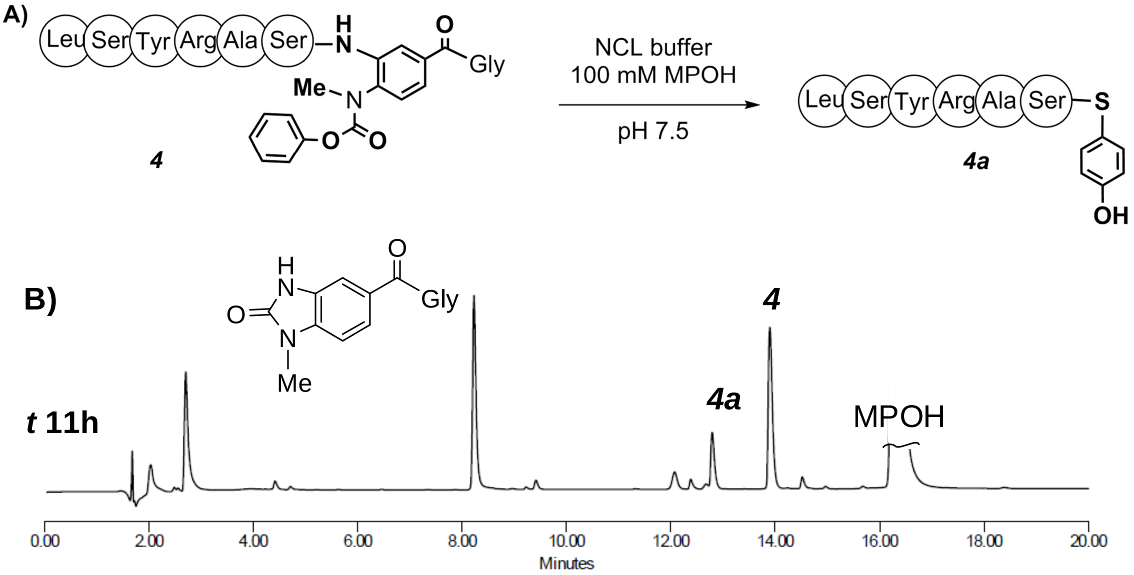

A)
Leu Ser Tyr Arg Ala Ser
H
N
O
Gly
Me
N
O
O
4
NCL buffer
100 mM MPOH
pH 7.5
Leu Ser Tyr Arg Ala Ser
S
4a
OH
B)
H
N
O
N
Me
O
Gly
4
4a
MPOH
t 11h
0.00
2.00
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00
Minutes

Figure 6. A) Scheme of the reaction of model Val terminal Phoc-MeDbz peptide in NCL buffer at pH 7.5. B) HPLC of the reaction after 24 h.

1.2.4. pF-Phoc-MeDbz

The pKa of the *p*-fluorophenol is similar to phenol (~9.9), so a similar behaviour should be expected. As predicted, hydrolysis of the carbamate moiety was not detected and the intramolecular cyclization to render the MeNbz peptide was slow. After 24 hours the main product observed was the starting *p*F-Phoc-MeDbz peptide which makes this linker suitable for slow ligations (**Figure *7***).

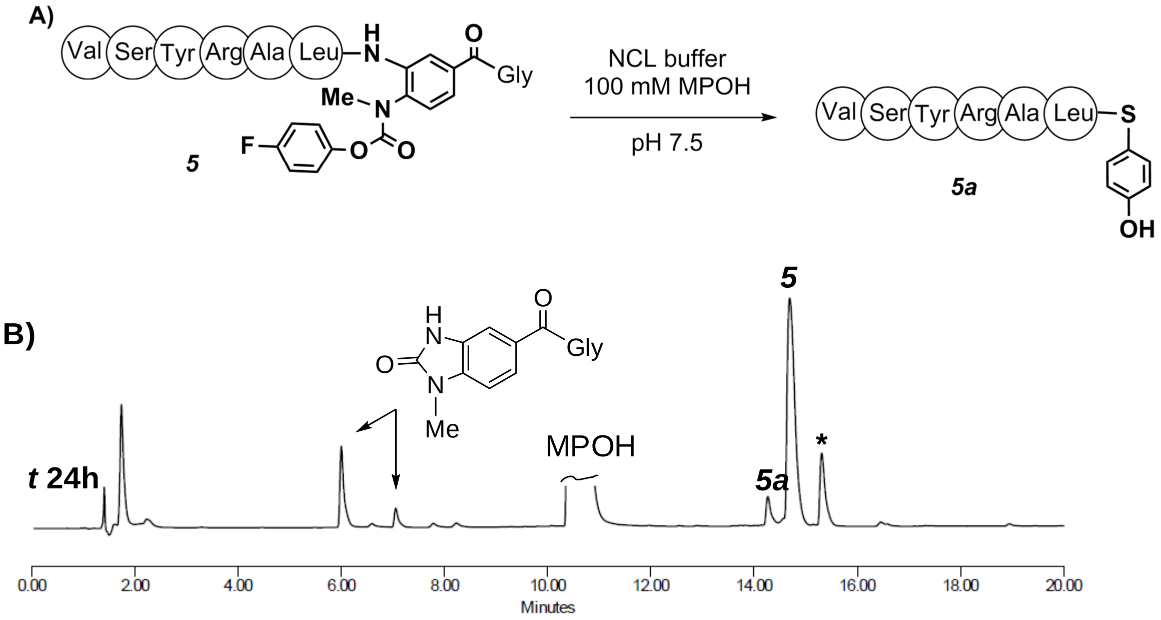

A)
Val Ser Tyr Arg Ala Leu
H
N
O
Gly
Me
N
F
O
O
5
NCL buffer
100 mM MPOH
pH 7.5
Val Ser Tyr Arg Ala Leu
S
5a
OH
B)
H
N
O
Gly
O
N
Me
5
MPOH
5a
*
t 24h
0.00
2.00
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00
Minutes

Figure 7. A) Scheme of the reaction of model Leu terminal *p*F-Phoc-MeDbz peptide in NCL buffer at pH 7.5. B) HPLC of the reaction after 24 h.

1.2.5. pSO$_2$Me-Phoc-MeDbz

p-(methylsulfonyl)phenol has a pKa similar to *p*-nitrophenol (~7.8-7.9). Because of its acid character, it was expected to observe hydrolysis of the peptide. After 4 hours it was already detected a little amount of hydrolysis of the carbamate which after 24 hours of reaction increased significantly (**Figure *8***).

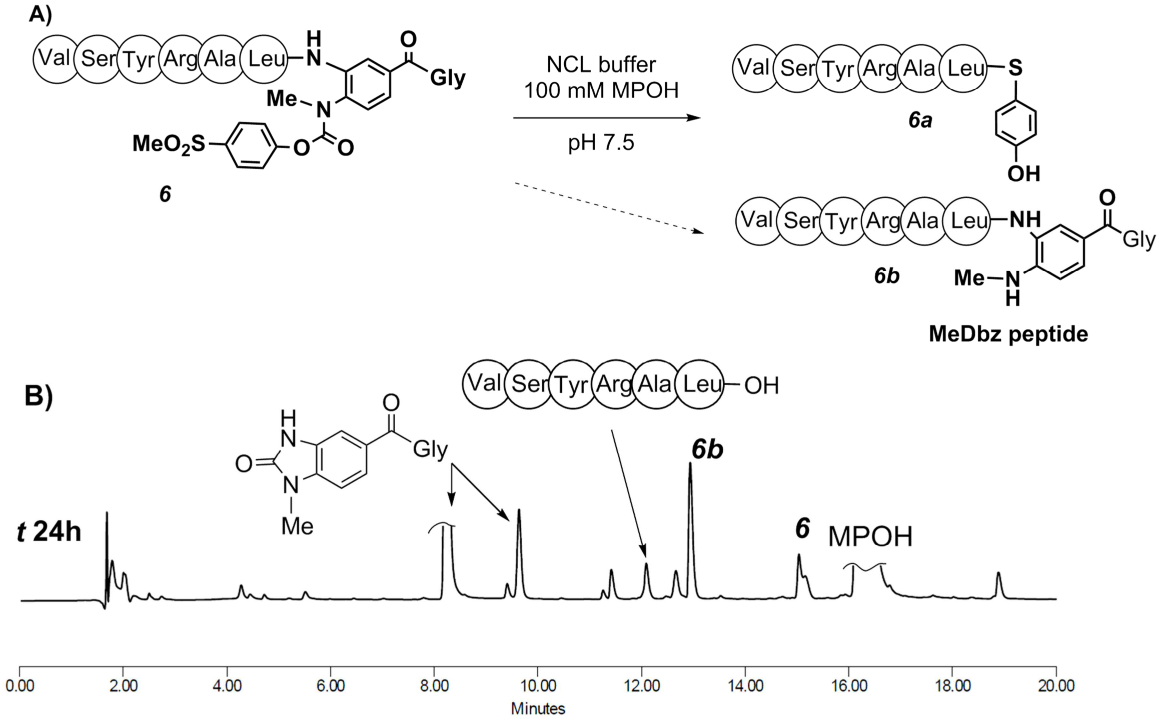

A)
Val Ser Tyr Arg Ala Leu
H
N
O
Gly
Me
N
MeO2S
O
O
6
NCL buffer
100 mM MPOH
pH 7.5
Val Ser Tyr Arg Ala Leu
S
6a
OH
Val Ser Tyr Arg Ala Leu
NH
O
Gly
6b
Me
N
H
MeDbz peptide
B)
Val Ser Tyr Arg Ala Leu
OH
H
N
O
Gly
O
N
Me
6b
6
MPOH
t 24h
0.00
2.00
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00
Minutes

Figure 8. A) Scheme of the reaction of model *p*SO$_2$Me-Phoc-MeDbz peptide in NCL buffer at pH 7.5. B) HPLC of the reaction after 24 h.

1.2.6. pCN-Phoc-MeDbz

Finally, the *p*CN-Phoc-MeDbz (pKa ~8.2) was explored. This derivative efficiently achieves a near complete thioester conversion in a model peptide bearing a C-terminal Ala after 7 hours of reaction (**Figure *9***). Moreover, no hydrolysis of the peptide was detected.

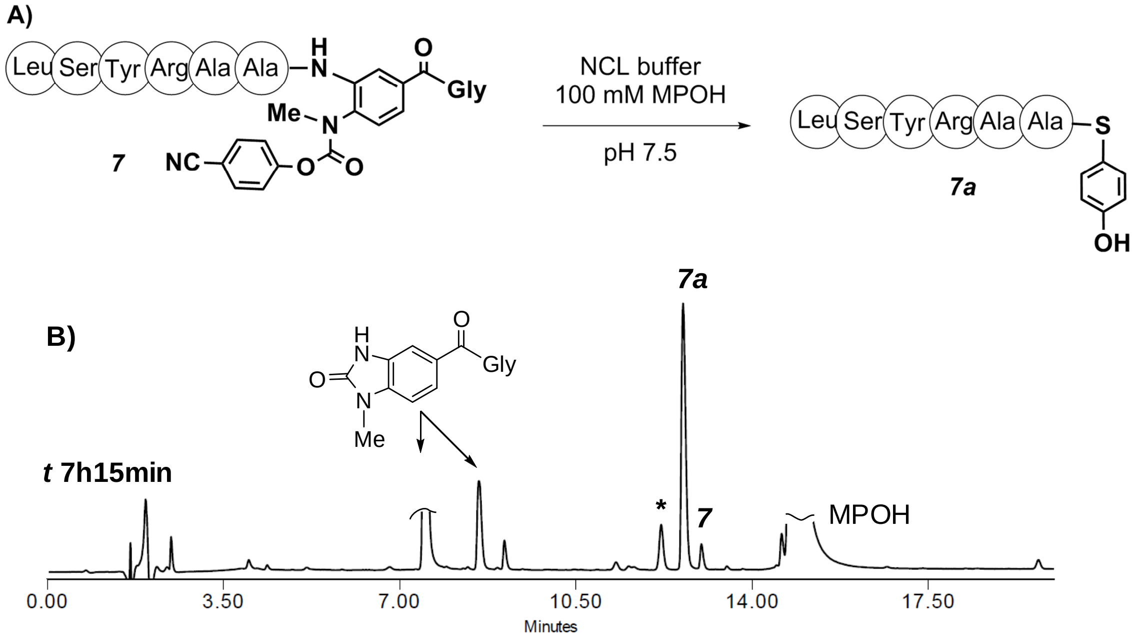


Figure 9. A) Scheme of the reaction of model *p*CN-Phoc-MeDbz peptide in NCL buffer at pH 7.5. B) HPLC of the reaction after 7h 15 minutes of reaction rT

After analyzing the results, the *p*CN-Phoc-MeDbz was selected as the alternative to the *p*NO_2-Phoc-MeDbz, although *m*NO_2-Phoc-MeDbz offers similar results. The lower price of *p*-cyanophenol represents a little advantage for a bulk-scale synthesis.

2. NCL assay using model *p*CN-Phoc-MeDbz peptides

The next objective was to explore the use of *p*CN-Phoc-MeDbz peptides for *in situ* NCL. With this purpose, we designed a ligation between a *p*CN-Phoc-MeDbz peptide containing a Phe at the C-terminus and an N-terminal Cys fragment. Both peptides were dissolved in NCL buffer containing 100 mM MPOH and the pH was adjusted to 7.5. The starting peptides were totally consumed after 6 hours yielding the desired ligated product (**Figure *10***).

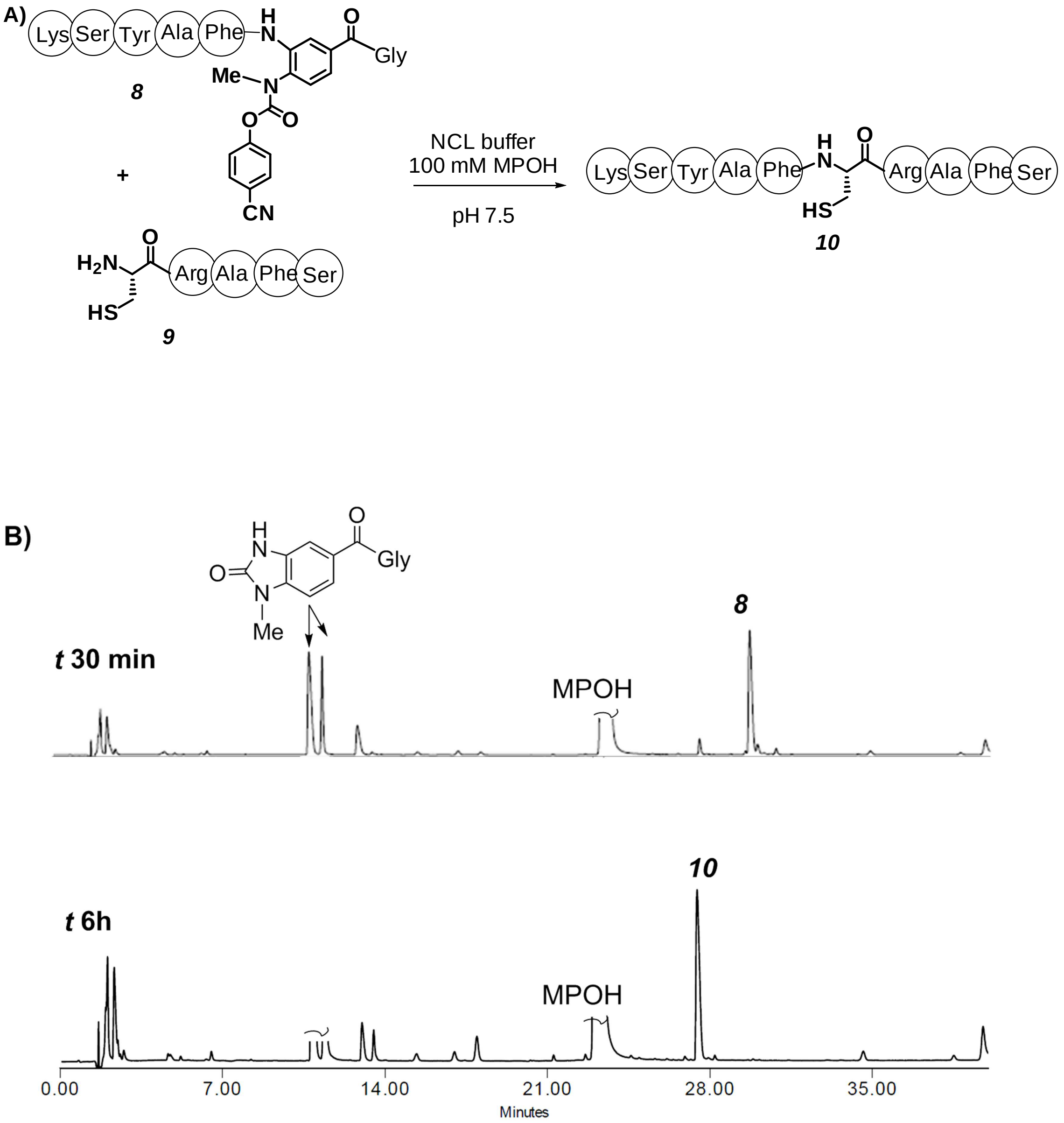

A)
Lys Ser Tyr Ala Phe
H N
O
Gly
Me N
O O
CN
8
+
H₂N
O
Arg Ala Phe Ser
HS
9
NCL buffer
100 mM MPOH
pH 7.5
Lys Ser Tyr Ala Phe
H N
O
Arg Ala Phe Ser
HS
10
B)
H N
O
Gly
O
N
Me
t 30 min
MPOH
8
t 6h
MPOH
10
0.00
7.00
14.00
21.00
28.00
35.00
Minutes

Figure 10. A) Scheme of the NCL between Phe terminal model *p*CN-Phoc-MeDbz peptide and CRAFS peptide. B) HPLC of the reaction at t = 30 min and 6 h.

It has to be pointed out that what determines the rate of the ligation is the kinetics of the MeNbz formation, which is highly dependent on the amino acid present at the C-terminus. So it would be expected ligation times about 6-8 hours for the rest of amino acids, except Pro and ß-branched which may require longer times according the kinetics observed in MeNbz peptides. [35] Based on previous observations,[35] once the MeNbz is formed, the thioesterification occurs rapidly (with the exception of Pro, Ile and Val). So, the rate of the NCL is essentially dependent on the time required for the peptide to fully cyclized into MeNbz form.

3. Racemization studies

The *N*-acylurea approach for NCL is a racemization-free method.[32,35] Since our methodology is based on the *N*-acylurea approach, it was expected the same result in the *p*CN-Phoc-MeDbz strategy. In order to further confirm it, it was run two parallel NCL with *p*CN-Phoc-MeDbz peptides bearing an L or D-Tyr at the C-terminus. So, the “L” and “D”-*p*CN-Phoc-MeDbz peptides were dissolved in the presence of Cys peptide in NCL buffer containing 100 mM MPOH and pH adjusted to 7.5 (**Figure *11***). The ligations were monitored by HPLC giving the desired ligated product in each case. The level of racemization was below of 0.5%, as quantified by the photodiode array detector of the HPLC system, thus confirming that in the case of Tyr the NCL using the *in situ* generation of MeNbz peptides occurs without racemization.

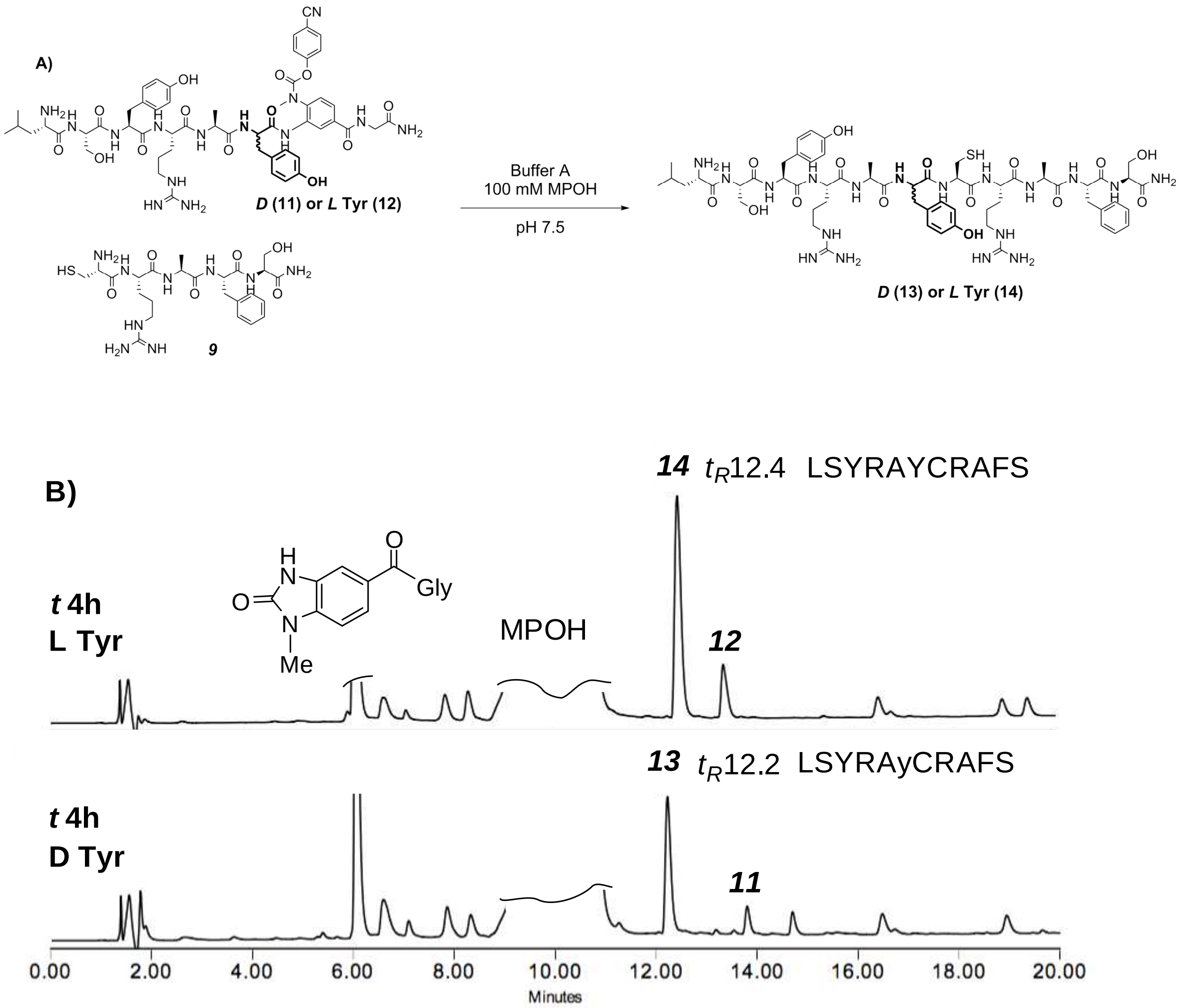

A)
D (11) or L Tyr (12)
9
Buffer A
100 mM MPOH
pH 7.5
D (13) or L Tyr (14)
B)
14 t_R12.4 LSYRAYCRAFS
t 4h
L Tyr
Gly
Me
MPOH
12
13 t_R12.2 LSYRAyCRAFS
t 4h
D Tyr
11
0.00
2.00
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00
Minutes

Figure 11. A) Scheme of the NCL between a D or L Tyr terminal pCN-Phoc-MeDbz peptide with CRAFS peptide. In blue Tyrosine moiety either L or D. B) HPLC of both ligations after 4h of reaction.t_R = retention time.

4. KCL using *p*CN-Phoc-MeDbz peptides

After assessing that *p*CN-Phoc-MeDbz peptides efficiently gave the desired product after NCL, their scope in KCL was further explored. Actually, it was envisaged their use as masked or crypto-thioester precursors, that upon selective activation of the linker would give C-terminal thioester peptides. Therefore, the presence of a Cys at the N-terminus of the same peptide would not interfere or react with the thioester precursor at the C-terminus, as it remained inert until activation (**Figure *12)***.

To confirm this hypothesis, it was designed a three-piece one-pot sequential ligation: a *p*CN-Phoc-MeDbz peptide displaying a C-terminal Phe and a Cys at the N-terminus would react in a first place at pH 7.0 with a MeNbz peptide having a C-terminal Leu. Once the ligation was completed, then a third Cys peptide was added and the pH was carefully raised up to 7.5 in order to speed up the MeNbz formation, and subsequently the next ligation, to finally give the three-fragment peptide.

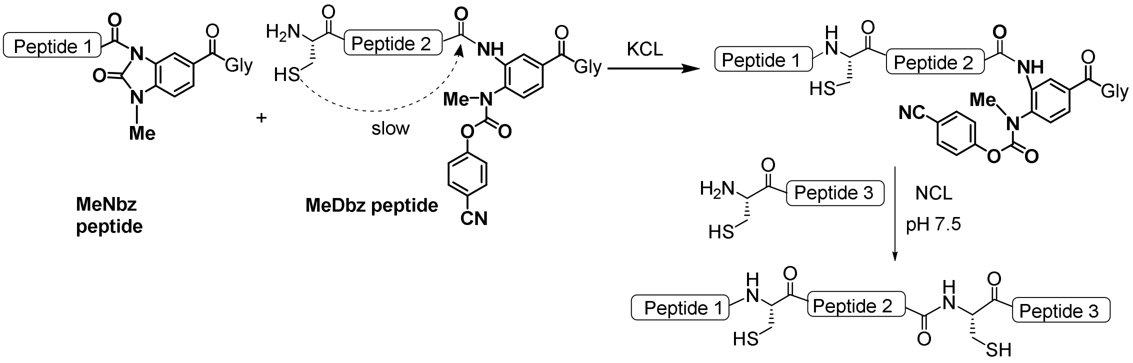


Figure 12. Concept of the KCL using pCN-Phoc-MeDbz peptides.

The chromatograms of the reaction (**Figure *13***) show that the first ligation gave the desired ligated product ***17*** in 3h 30 min. When the pH was raised to 7.5 to promote the cyclization of the linker and, thus the subsequent thioesterification, the desired KCL product ***18*** from the ligation between the peptide ***17*** and the peptide ***9*** was observed as the main product. Several other byproducts were observed in lower ratios, such as the ligation product from ***15*** and ***9***. (**Figure *13***). These results confirmed that *p*CN-Phoc-MeDbz peptides could be used in KCL as masked thioester precursors.

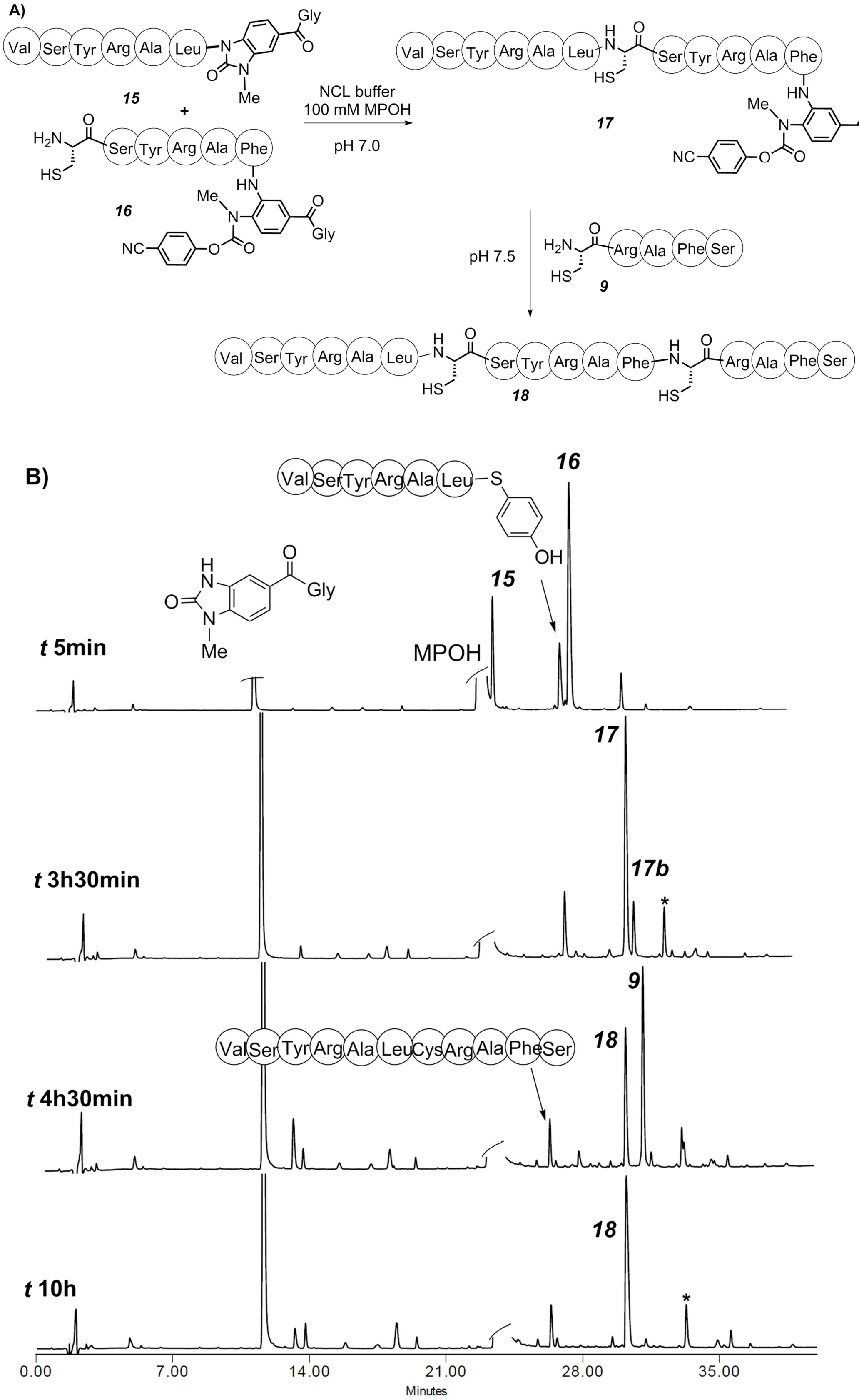


Figure 13. A) Scheme of the designed KCL using MeNbz and pCN-Phoc-MeDbz peptides. B) HPLC at different reaction times of the model KCL reaction. **17b** = MPOH thioester from **17**.

5. Stability of *p*CN-Phoc-MeDbz peptides under common protein synthesis reactions

To further assess that this new methodology could be used for the convergent chemical synthesis of proteins, it was explored the stability of the *p*CN-Phoc-MeDbz group in common reactions used in protein synthesis, i. e., desulfurization, S-acetamidomethyl (Acm) and thiazolidine (Thz) removal.

5.1. Acm removal

The Acm protecting group is widely used to protect natural Cys in order to avoid undesired desulfurization after NCL at non Cys residues. There are reported different conditions to remove the Acm, such as the use of 2,2'-dithiobis(5-nitropyridine) (DTNP),[50] iodine oxidation[51] or the use of heavy metals such as thallium,[52] mercury[53] or silver.[54] In our hands, silver trifluoromethanesulfonate (AgOTf) in acetic acid solution gave the best results.

A model *p*CN-Phoc-MeDbz peptide carrying a Acm-protected Cys was dissolved at room temperature in an acetic aqueous buffer (50%) containing 50 mM of AgOTf (**Figure *14***). After the complete deprotection of the Cys the peptide remained stable even after the work up. Therefore, the *p*CN-Phoc-MeDbz peptides are orthogonal to Acm deprotection conditions using silver salts.

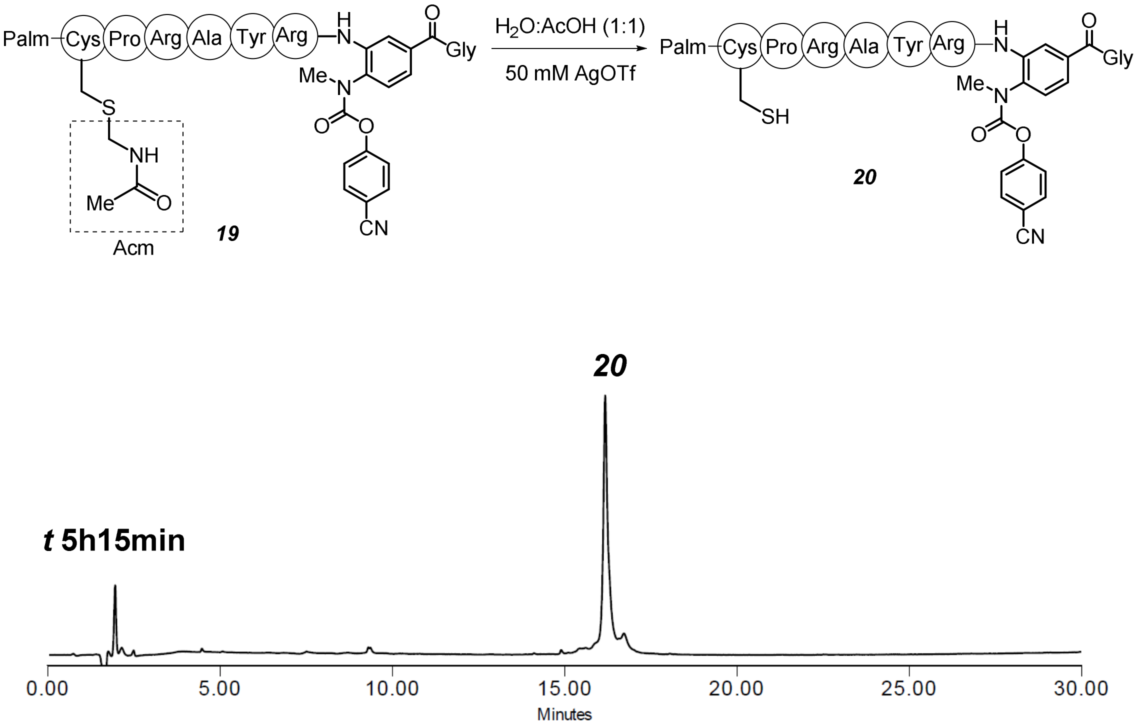


Figure 14. A) Scheme of the Acm removal reaction. B) HPLC of the reaction after 5h.

5.2. Desulfurization

Due to the lack of Cys in some proteins, or their unsuitable position for NCL, some residues are usually changed by Cys, or other thiol auxiliaries, to facilitate their synthesis. Therefore, desulfurization is a frequent reaction carried out during chemical synthesis of proteins to restore the original residue in that position.

Several desulfurization conditions have been described to remove thiol groups in peptides, from metal based reactions i.e. Nickel Raney,[12,55,56] to radical-based conditions,[57] including the most recent described methodology using $NaBH_4$.[58] Homogeneous radical desulfurization has gained a lot of attention in the last years, principally due to its smooth conditions and high recovery yield.

Following the radical desulfurization conditions described in the literature, it was carried out optimization assays on model peptides containing a cysteine (**Figure *15***). Different additives and parameters were screened, such as temperature, pH, concentration of the radical initiatior and thiols (**Table *2***), having into account that high temperature or pH could lead to MeNbz conversion[a]. The best desulfurization conditions were obtained when dissolving the peptide at 2 mM in **desulfurization buffer D**: 6 M Gn·HCl, 0.2 M phosphate, 250 mM MESNa, 250 mM TCEP and 200 mM VA-044 (2,2'-Azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride) at pH 5.0 and 37 ºC. After 2 hours, the desulfurization reached a conversion of 91% (**Figure *16***).

Once the desulfurization reaction was optimized, the stability of the *p*CN-Phoc-MeDbz was checked under these conditions. After 7 hours it was found that the peptide remained stable without observing any side-reaction in the Phoc-MeDbz moiety (**Figure *17***). Therefore, the *p*CN-Phoc-MeDbz group is orthogonal to radical based desulfurization conditions.

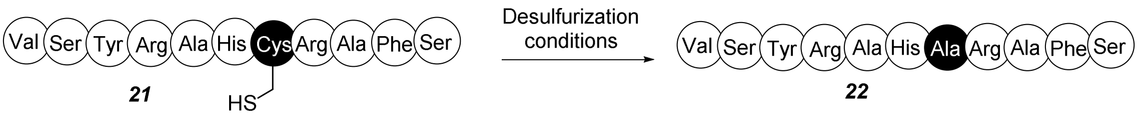


Figure 15. Scheme of desulfurization reaction optimization on model peptide.

---

[a] In principle, the MeNbz conversion should be avoided. However, the presence of MESNa would generate the thioester in situ, that is stable under desulfurization conditions.

Buffer, pH	[TCEP]	[VA-044]	Thiol scavenger	Temp	% Cys	% Ala
Buffer* pH 6.8	0.5 M	0.1 M	0.3 M C(CH$_3$)$_3$SH	rT	t 6h, 23%	t 6h, 77%
Buffer* pH 6.8	0.5 M	0.1 M	0.3 M C(CH$_3$)$_3$SH	40 °C	t 5h, 25%	t 5h, 75%
Buffer* pH 5.3	0.5 M	0.1 M	0.3 M C(CH$_3$)$_3$SH	rT	t 10 h, 53%	t 10 h, 46%
Buffer* pH 5.3	0.5 M	0.1 M	0.3 M C(CH$_3$)$_3$SH	40 °C	t 10 h, 51%	t 10 h, 49%
Buffer* pH 5.3	0.5 M	0.2 M	40 mM Gluthathione	rT	t 20 h, 92%	t 20h, 8%
Buffer* pH 5.3	0.5 M	--	40 mM Gluthathione	rT	t 5h, 90%	t 5h, 10%
Buffer* pH 5.3	0.25 M	0.2 M	200 mM MESNa	rT	t 6h, 15%	t 6h, 84%
Buffer* pH 6.85	0.25 M	0.2 M	100 mM MESNa	rT	t 4h, 89%	t 4h, 11%
Buffer* pH 6.85	0.25 M	--	--	60 °C	t 4h, 90%	t 4h, 10%
Buffer* pH 5	0.25 M	0.2 M	250 mM MESNa	rT	t 4h, 19%	t 4h, 81%
Buffer* pH 5	0.25 M	0.2 M	250 mM MESNa	37 °C	t 2h, 9%	t 2h, 91%

Table 2. Desulfurization conditions assayed on model peptide VSYRAHCRAFS. * = buffer containing 6 M Gn·HCl, 0.2 M phosphate

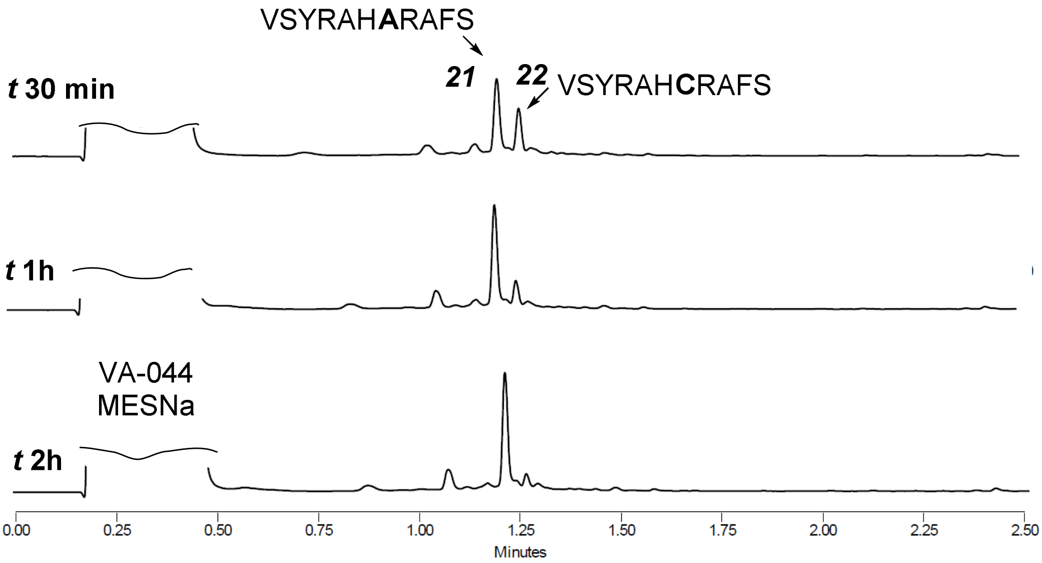


Figure 16. HPLC traces at indicated times of the desulfurization reaction using conditions highlighted in Table 2.

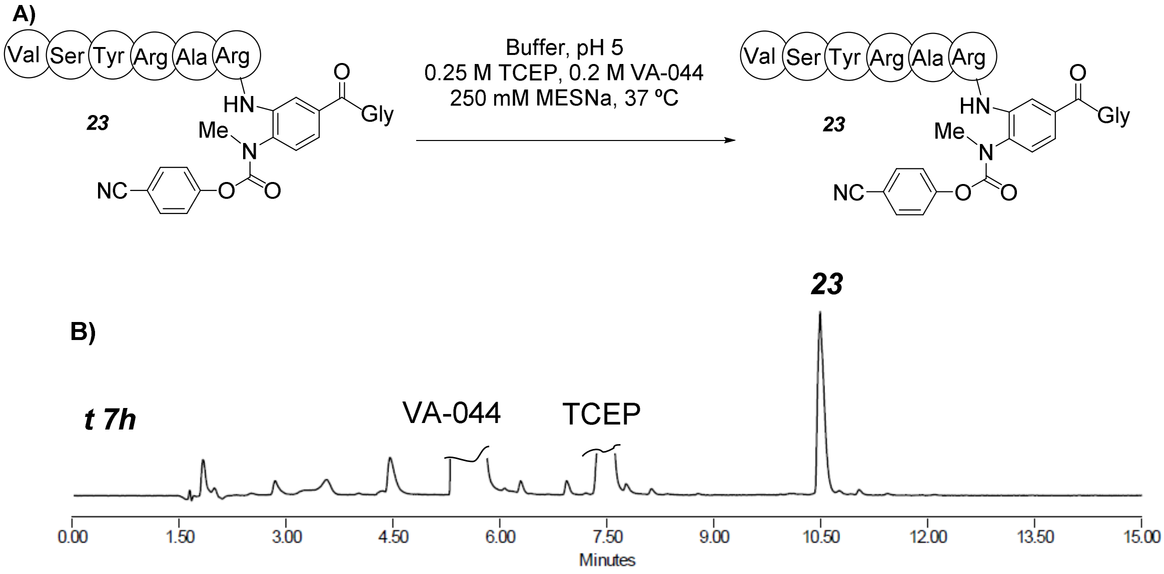


Figure 17. A) Stability of *p*CN-Phoc-MeDbz under desulfurization conditions. B) HPLC of the reaction after 7 hours.

5.3. Thz conversion to Cys

Other strategy widely used to selectively protect Cys at the N-terminus during NCL is the thiazolidine, which is removed in aqueous solution using methoxylamine.[59] Most of the protocols reported the use of methoxylamine at 0.3 M in aqueous buffer. Thereby, it was decided to carry out the screening reactions dissolving the peptide (**Figure *18***) in 6 M Gn·HCl, 0.2 M phosphate, 0.2 M TCEP and 0.3 M methoxylamine and changing the pH or the temperature (**Table *3***). Complete Thz removal was observed after 4 hours when the reaction was carried out at pH 4 and 40 °C.

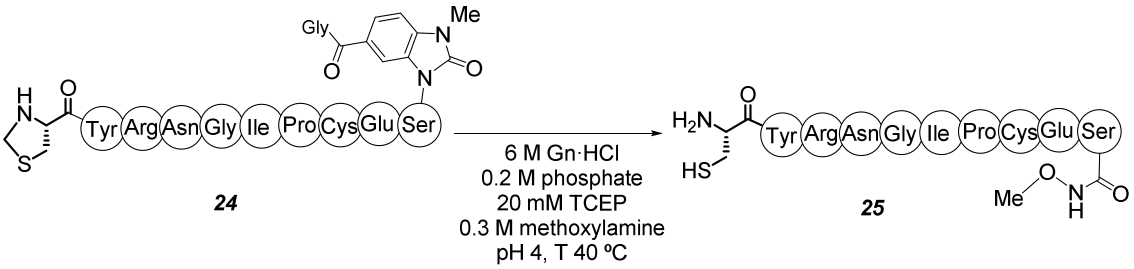

H
N
S
O
Tyr
Arg
Asn
Gly
Ile
Pro
Cys
Glu
Ser
Gly
O
Me
N
N
O
24
6 M Gn·HCl
0.2 M phosphate
20 mM TCEP
0.3 M methoxylamine
pH 4, T 40 ºC
H_2N
HS
O
25
Me
O
N
H
O

Figure 18. Scheme of Thz removal conditions on MeNbz model peptide harboring a Thz residue at the N-terminus.

pH	Temp	Observations
5	40	Still Thz product observed after 4 h
5	rT	Still Thz product observed after 30 h
6	40	Still Thz product observed after 26 h
6	rT	Still Thz product observed after 30 h
7	40	Still Thz product observed after 5 h
7	rT	Still Thz product observed after 6 h 30 min
4	40	All Thz converted into Cys after 4h

Table 3. Optimization of the conditions for Thz removal. All the reactions were carried out in a buffer containing 6 M Gn·HCl, 0.2 M NaPi, 20 mM TCEP, 0.3 M $MeONH_2$.

When a model *p*CN-Phoc-MeDbz peptide was dissolved under these optimized conditions (highlighted in black) (**Figure *19***), both the carbamate and the CN groups did not react with methoxylamine, supporting the ortoghonality with the Thz group.

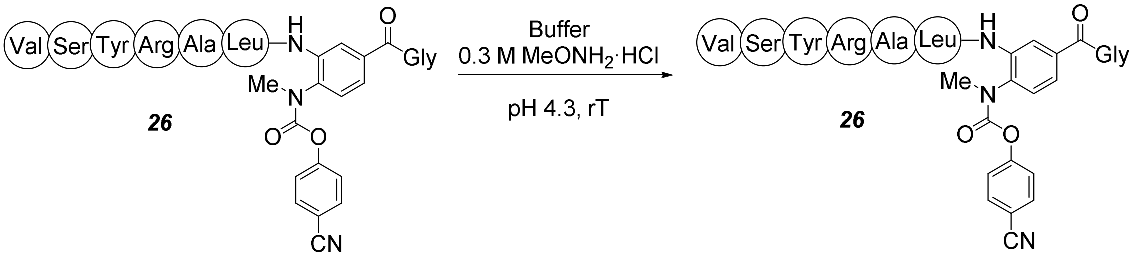


Figure 19. Scheme Thz removal reaction on model peptide.

Based on the results obtained, the *p*CN-Phoc-MeDbz linker is orthogonal to desulfurization, Acm and Thz removal conditions. Therefore, the new methodology described in the present thesis is suitable for the chemical synthesis of proteins. In the next chapter we are going to apply the *p*CN-Phoc-MeDbz methodology to the chemical synthesis of the Sonic Hedgehog, a protein involved in tissue patterning, neurodevelopment or stem cell differentiation.

CONCLUSIONS

The conclusions of this chapter adressed to the synthesis of MeDbz linker derivatives are:

1. We have successfully synthetized *p*-nitrophenyloxycarbonyl MeDbz (*p*NO_2-Phoc-MeDbz) peptides after acylation of the corresponding MeDbz linker with *p*-nitrophenyl chloroformate and cleavage off the resin.

2. The cyclization of the *p*NO_2-Phoc-MeDbz peptides to the corresponding MeNbz in solution was achieved raising the pH to 7.5. The MeNbz peptides underwent rapid thioesterification in the NCL buffer. Therefore, a new methodology to prepare thioester peptides in solution from *p*NO_2-Phoc-MeDbz peptides was developed. However, thioesterification in the presence of MPOH leads to significant hydrolysis of the carbamate linker.

3. To avoid this undesired reaction, different aryloxycarbonyl-MeDbz derivatives were screened. The *p*-cyano-phenyloxycarbonyl-MeDbz (*p*CN-Phoc-MeDbz) was found to successfully ligate to Cys peptides with no hydrolysis of the linker in the presence of MPOH.

4. The new *p*CN-Phoc-MeDbz derivatives are orthogonal to homogeneous radical based desulfurization conditions, stable under Acm removal by silver salts, and thiazolidine conversion into cysteine using methoxylamine. Therefore, the *p*CN-Phoc-MeDbz methodology is a new approach for the synthesis of proteins by convergent strategies and KCL

5. A sequential three-piece one-pot was carried out using a MeNbz peptide, another *p*CN-Phoc-MeDbz fragment displaying a Cys at the N-terminus and a third Cys peptide. This experiment confirmed that *p*CN-Phoc-MeDbz peptides can be used as masked thioester precursors in KCL reactions.

REFERENCES

(1) Fruton, J. S. *A Skeptical Biochemist*, 1st ed.; Fruton, J. S., Ed.; Harvard University Press: Cambridge, Massachusetts; London, England, 1992.

(2) du Vigneaud, V.; Resslbr, C.; Swan, J. M.; Roberts, C. W.; Katsoyannis, P. G. The Synthesis of Oxytocin. *J. Am. Chem. Soc.* **1954**, *76* (12), 3115–3121.

(3) Mayer, J. P.; Zhang, F.; DiMarchi, R. D. Insulin Structure and Function. *Pept. Sci.* **2007**, *88*, 687.

(4) Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **1963**, *85* (14), 2149–2154.

(5) Kenner, G. W. Bakerian Lecture - Towards Synthesis of Proteins. *Proc. R. Soc. London Ser. B-Biological Sci.* **1977**, *197* (1128), 237–253.

(6) Kent, S. B. H. Total Chemical Synthesis of Proteins. *Chem. Soc. Rev.* **2009**, *38* (2), 338–351.

(7) Schnölzer, M.; Kent, S. B. H. Constructing Proteins by Dovetailing Unprotected Synthetic Peptides: Backbone-Engineered HIV Protease. *Science (80).* **1992**, *256* (5054), 221–225.

(8) Milton, R.; Milton, S.; Kent, S. B. H. Total Chemical Synthesis of a D-Enzyme: The Enantiomers of HIV-1 Protease Show Reciprocal Chiral Substrate Specificity. *Science (80).* **1992**, *257*, 147.

(9) Dawson, P. E.; Muir, T. W.; Kent, S. B. H. Synthesis of Proteins by Native Chemical Ligation. *Science (80).* **1994**, *266* (925), 776–779.

(10) Hackeng, T. M.; Griffin, J. H.; Dawson, P. E. Protein Synthesis by Native Chemical Ligation: Expanded Scope by Using Straightforward Methodology. *Biochem. Chem.* **1999**, *96* (August), 10068–10073.

(11) Agouridas, V.; El Mahdi, O.; Cargoët, M.; Melnyk, O. A Statistical View of Protein Chemical Synthesis Using NCL and Extended Methodologies. *Bioorganic Med. Chem.* **2017**, *25* (18), 4938–4945.

(12) Yan, L. Z.; Dawson, P. E. Synthesis of Peptides and Proteins without Cysteine Residues by Native Chemical Ligation Combined with Desulfurization. *J. Am. Chem. Soc.* **2001**, *123* (4), 526–533.

(13) Rohde, H.; Seitz, O. Ligation-Desulfurization: A Powerful Combination in the Synthesis of Peptides and Glycopeptides. *Biopolymers* **2010**, *94* (4), 551–559.

(14) Wong, C. T. T.; Tung, C. L.; Li, X. Synthetic Cysteine Surrogates Used in Native Chemical Ligation. *Mol. Biosyst.* **2013**, *9* (5), 826.

(15) Hondal, R. J.; Nilsson, B. L.; Raines, R. T. Selenocysteine in Native Chemical Ligation and Expressed Protein Ligation Department of Chemistry , Uni V Ersity of Wisconsin-Madison L -Selenocysteine (Sec or U) Has Been Called the “ 21st Amino Acid ”. 1 Like the Twenty Common Amino Acids , Selenocyst. *J. Am. Chem. Soc.* **2001**, (123), 5140–

5141.

(16) Metanis, N.; Keinan, E.; Dawson, P. E. Traceless Ligation of Cysteine Peptides Using Selective Deselenization. *Angew. Chemie - Int. Ed.* **2010**, *49* (39), 7049–7053.

(17) Li, X.; Kawakami, T.; Aimoto, S. Direct Preparation of Peptide Thioesters Using an Fmoc Solid-Phase Method. *Tetrahedron Lett.* **1998**, *39* (47), 8669–8672.

(18) Hojo, H.; Haginoya, E.; Matsumoto, Y.; Nakahara, Y.; Nabeshima, K.; Toole, B. P.; Watanabe, Y. The First Synthesis of Peptide Thioester Carrying N-Linked Core Pentasaccharide through Modified Fmoc Thioester Preparation: Synthesis of an N-Glycosylated Ig Domain of Emmprin. *Tetrahedron Lett.* **2003**, *44* (14), 2961–2964.

(19) Clippingdale, a B.; Barrow, C. J.; Wade, J. D. Peptide Thioester Preparation by Fmoc Solid Phase Peptide Synthesis for Use in Native Chemical Ligation. *J Pept Sci* **2000**, *6* (5), 225–234.

(20) Zhu, J.; Wan, Q.; Ragupathi, G.; George, C. M.; Livingston, P. O.; Danishefsky, S. J. Biologics through Chemistry: Total Synthesis of a Proposed Dual-Acting Vaccine Targeting Ovarian Cancer by Orchestration of Oligosaccharide and Polypeptide Domains. *J. Am. Chem. Soc.* **2009**, *131* (11), 4151–4158.

(21) Von Eggelkraut-Gottanka, R.; Klose, A.; Beck-Sickinger, A. G.; Beyermann, M. Peptide ??thioester Formation Using Standard Fmoc-Chemistry. *Tetrahedron Lett.* **2003**, *44* (17), 3551–3554.

(22) Aboye, T. L.; Clark, R. J.; Craik, D. J.; Goransson, U. Ultra-Stable Peptide Scaffolds for Protein Engineering - Synthesis and Folding of the Circular Cystine Knotted Cyclotide Cycloviolacin O2. *ChemBioChem* **2008**, *9* (1), 103–113.

(23) Mende, F.; Seitz, O. 9-Fluorenylmethoxycarbonyl-Based Solid-Phase Synthesis of Peptide ??-Thioesters. *Angew. Chemie - Int. Ed.* **2011**, *50* (6), 1232–1240.

(24) Ficht, S.; Payne, R. J.; Guy, R. T.; Wong, C. H. Solid-Phase Synthesis of Peptide and Glycopeptide Thioesters through Side-Chain-Anchoring Strategies. *Chem. - A Eur. J.* **2008**, *14* (12), 3620–3629.

(25) Alsina, J.; Yokum, T. S.; Albericio, F.; Barany, G. Backbone Amide Linker (BAL) Strategy for N-9- Fluorenylmethoxycarbonyl (Fmoc) Solid-Phase Synthesis of Unprotected Peptide P-Nitroanilides and Thioesters. *J. Org. Chem.* **1999**, *64* (24), 8761–8769.

(26) Ingenito, R.; Bianchi, E.; Fattori, D.; Pessi, A. Solid Phase Synthesis of Peptide C-Terminal Thioesters by Fmoc/t-Bu Chemistry. *J. Am. Chem. Soc.* **1999**, *121* (49), 11369–11374.

(27) Shin, Y.; Winans, K. A.; Backes, B. J.; Kent, S. B. H.; Ellman, J. A.; Bertozzi, C. R. Fmoc-Based Synthesis of Peptide-(α)thioesters: Application to the Total Chemical Synthesis of a Glycoprotein by Native Chemical Ligation. *J. Am. Chem. Soc.* **1999**, *121* (50), 11684–11689.

(28) Camarero, J. A.; Hackel, B. J.; De Yoreo, J. J.; Mitchell, A. R. Fmoc-Based Synthesis of

Peptide α-Thioesters Using an Aryl Hydrazine Support. *J. Org. Chem.* **2004**, *69* (12), 4145–4151.

(29) Tofteng, A. P.; Sørensen, K. K.; Conde-Frieboes, K. W.; Hoeg-Jensen, T.; Jensen, K. J. Fmoc Solid-Phase Synthesis of C-Terminal Peptide Thioesters by Formation of a Backbone Pyroglutamyl Imide Moiety. *Angew. Chemie Int. Ed.* **2009**, *48* (40), 7411–7414.

(30) Elashal, H. E.; Sim, Y. E.; Raj, M. Serine Promoted Synthesis of Peptide Thioester-Precursor on Solid Support for Native Chemical Ligation. *Chem. Sci.* **2017**, *8* (1), 117–123.

(31) Terrier, V. P.; Adihou, H.; Arnould, M.; Delmas, A. F.; Aucagne, V. A Straightforward Method for Automated Fmoc-Based Synthesis of Bio-Inspired Peptide Crypto-Thioesters. *Chem. Sci.* **2016**, *7* (1), 339–345.

(32) Blanco-Canosa, J. B.; Dawson, P. E. An Efficient Fmoc-SPPS Approach for the Generation of Thioester Peptide Precursors for Use in Native Chemical Ligation. *Angew. Chemie - Int. Ed.* **2008**, *47* (36), 6851–6855.

(33) Shu, K.; Noguchi, T.; Honda, K.; Kondoh, Y.; Osada, H.; Ohno, H.; Fujii, N.; Oishi, S. Synthesis of the Src SH2 Domain and Its Application in Bioassays for Mirror-Image Screening. *RSC Adv.* **2017**, *7* (61), 38725–38732.

(34) Sueiras-Diaz, J.; Zhang, Y.; Velentza, A.; Santoso, B.; Yang, S. Total Chemical Synthesis of a Biologically Active and Homogeneous Analog of Human Growth Hormone [Nle14,125,170,Glu29,91,Gln74,Asn107,Asp109]hGH-NH2 by Sequential Native Chemical Ligation. *Tetrahedron Lett.* **2017**, *58* (25), 2448–2455.

(35) Blanco-Canosa, J. B.; Nardone, B.; Albericio, F.; Dawson, P. E. Chemical Protein Synthesis Using a Second-Generation N-Acylurea Linker for the Preparation of Peptide-Thioester Precursors. *J. Am. Chem. Soc.* **2015**, *137* (22), 7197–7209.

(36) Wang, J. X.; Fang, G. M.; He, Y.; Qu, D. L.; Yu, M.; Hong, Z. Y.; Liu, L. Peptide O-Aminoanilides as Crypto-Thioesters for Protein Chemical Synthesis. *Angew. Chemie - Int. Ed.* **2015**, *54* (7), 2194–2198.

(37) Fang, G. M.; Li, Y. M.; Shen, F.; Huang, Y. C.; Li, J. Bin; Lin, Y.; Cui, H. K.; Liu, L. Protein Chemical Synthesis by Ligation of Peptide Hydrazides. *Angew. Chemie - Int. Ed.* **2011**, *50* (33), 7645–7649.

(38) Ollivier, N.; Vicogne, J.; Vallin, A.; Drobecq, H.; Desmet, R.; El Mahdi, O.; Leclercq, B.; Goormachtigh, G.; Fafeur, V.; Melnyk, O. A One-Pot Three-Segment Ligation Strategy for Protein Chemical Synthesis. *Angew. Chemie - Int. Ed.* **2012**, *51* (1), 209–213.

(39) Sato, K.; Shigenaga, A.; Kitakaze, K.; Sakamoto, K.; Tsuji, D.; Itoh, K.; Otaka, A. Chemical Synthesis of Biologically Active Monoglycosylated GM2-Activator Protein Analogue Using N-Sulfanylethylanilide Peptide. *Angew. Chemie - Int. Ed.* **2013**, *52* (30), 7855–7859.

(40) Bang, D.; Pentelute, B. L.; Kent, S. B. H. Kinetically Controlled Ligation for the Convergent Chemical Synthesis of Proteins. *Angew. Chemie Int. Ed.* **2006**, *45* (24), 3985–3988.

(41) Lee, J.; Kwon, Y.; Pentelute, B. L.; Bang, D. Use of Model Peptide Reactions for the Characterization of Kinetically Controlled Ligation. *Bioconjug. Chem.* **2011**, *22* (8), 1645–1649.

(42) Beligere, G. S.; Dawson, P. E. Conformationally Assisted Protein Ligation Using C-Terminal Thioester Peptides *J. Am. Chem. Soc.* **1999**, *121* (26), 6332–6333.

(43) Muir, T. W.; Sondhi, D.; Cole, P. A. Expressed Protein Ligation: A General Method for Protein Engineering. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (12), 6705–6710.

(44) Muralidharan, V.; Muir, T. W. Protein Ligation: An Enabling Technology for the Biophysical Analysis of Proteins. *Nat. Methods* **2006**, *3* (6), 429–438.

(45) Muir, T. W. Semisynthesis of Proteins by Expressed Protein Ligation. *Annu. Rev. Biochem.* **2003**, *72* (1), 249–289.

(46) Henager, S. H.; Chu, N.; Chen, Z.; Bolduc, D.; Dempsey, D. R.; Hwang, Y.; Wells, J.; Cole, P. A. Enzyme-Catalyzed Expressed Protein Ligation. *Nat. Methods* **2016**, *13* (11), 925–927.

(47) Warden-Rothman, R.; Caturegli, I.; Popik, V.; Tsourkas, A. Sortase-Tag Expressed Protein Ligation: Combining Protein Purification and Site-Specific Bioconjugation into a Single Step. *Anal. Chem.* **2013**, *85* (22), 11090–11097.

(48) Liptak, M. D.; Gross, K. C.; Seybold, P. G.; Feldgus, S.; Shields, G. C. Absolute pKa Determinations for Substituted Phenols. *J. Am. Chem. Soc.* **2002**, *124* (22), 6421–6427.

(49) Wang, Q.; Huang, R. Synthesis and Biological Activity of Novel N -Tert-Butyl-N -Substituted Benzoyl-N- (Substituted Phenyl) Aminocarbonylhydrazines and Their Derivatives. **2001**, *42*, 8881–8883.

(50) Harris, K. M.; Flemer, S. J.; Hondal, R. J. Studies on Deprotection of Cysteine and Selenocysteine Side-Chain Protecting Groups. *J. Pept. Sci.* **2007**, *13*, 81–93.

(51) Otaka, A.; Koide, T.; Shide, A.; Fuji, N. Application of Simethylsulphoxide (DMSO)/trifluoroacetic Acid (TFA) Oxidation to the Synthesis of Cystine-Containing Peptide. *Tetrahedron Lett.* **1991**, *32*, 1223–1226.

(52) Fujii, N.; Otaka, A.; Funakoshi, S.; Bessho, K.; Yajima, H. New Procedure for the Synthesis of Cystine-Peptides by Oxidation of S-Substituted Cysteine-Peptides with thallium(III) Trifluoroacetate. *J. Chem. Soc.{,} Chem. Commun.* **1987**, *5* (3), 163–164.

(53) Fuji, N.; Otaka, A.; Funakoshi, S.; Bessho, K.; H, Y. New Method for Removing the S-P-Methoxybenzyl and S-T-Butyl Groups of Cysteine Residues with Mercuric Trifluoroactetate. *Chem. Pharm. Bull.* **1978**, *26*, 1576–1585.

(54) Kawakami, T.; Toda, C.; Akaji, K.; Nishimura, T.; Nakatsuji, T.; Ueno, K.; Sonobe, M.; Aimoto, S. No Title. *J. Biochem.* **2000**, No. 128, 455–461.

(55) Yang, Y. Y.; Ficht, S.; Brik, A.; Wong, C. H. Sugar-Assisted Ligation in Glycoprotein Synthesis. *J. Am. Chem. Soc.* **2007**, *129* (24), 7690–7701.

(56) Pentelute, B. L.; Kent, S. B. H. Selective Desulfurization of Cysteine in the Presence of Cys(Acm) in Polypeptides Obtained by Native Chemical Ligation. *Org. Lett.* **2007**, *9* (4), 687–690.

(57) Wan, Q.; Danishefsky, S. J. Free-Radical-Based, Specific Desulfurization of Cysteine: A Powerful Advance in the Synthesis of Polypeptides and Glycopolypeptides. *Angew. Chemie - Int. Ed.* **2007**, *46* (48), 9248–9252.

(58) Jin, K.; Li, T.; Chow, H. Y.; Liu, H.; Li, X. P−B Desulfurization: An Enabling Method for Protein Chemical Synthesis and Site-Specific Deuteration. *Angew. Chemie - Int. Ed.* **2017**, *56* (46), 14607–14611.

(59) Bang, D.; Kent, S. B. H. A One-Pot Total Synthesis of Crambin. *Angew. Chemie - Int. Ed.* **2004**, *43* (19), 2534–2538.

CHAPTER 2:

Synthesis of Sonic Hedgehog

INTRODUCTION

Insufficient activity of Hh during development leads to neonatal malformations such as holoprosencephaly or brachydactyly. In adults, remains mainly quiescent, but it is involved in the regulation of cell growth, survival and differentiation, and thereby, aberrant signaling of this pathway has been found in some types of cancers.[1,2]

The Hh cascade was discovered in the 1970s by Nüsslein-Volhard and E. F. Wieschaus, then awarded with the Nobel Prize, who identified more than 50 genes involved on the development of Drosophila Melanogaster.[3] The mutation of one of these genes displayed spikes-like in the wings on the fly larvae, what led to later call it **hedgehog gene.**[4]

In 1993, three paralogous genes related to hedgehog were identified in vertebrates:[5] indian hedgehog, desert hedgehog and sonic hedgehog, being this last one named after the famous video-game main character. Each of these genes codifies for a different protein: Indian Hedgehog (Ihh), Desert Hedgehog (Dhh) and Sonic Hedgehog (Shh) respectively. Hedgehog genes and proteins are highly conserved between species, from vertebrates to invertebrates, and play similar roles during embryo development.[1]

The Hedgehog pathway (Hh) is a crucial signaling cascade during embryogenesis. Insuficient but it is also known to be related to important diseases in adults when its signaling is altered. Insufficient activity of Hh during development leads to diseases such as holoprosencephaly or brachydactyly. While in adults, Hh is responsible of cell growth, survival and differentiation, and thereby, aberrant signaling of this pathway has been found in some types of cancers.

1. Hh signaling pathway

The first studies of the Hh pathway were carried out in Drosophila Melanogaster and mouse cell cultures in the late 1990s.[6,7] Although until the moment to understand and identify the different components of this complex pathway, there are still some steps in which the mechanism remains unknown.

The main proteins in Hh pathway are **Patched** (Ptc), **Smoothened** (Smo) and the **Glioma related** (Gli) **transcription factors**. Ptc, a twelve transmembrane protein,[8] is the receptor of ShhN and it concentrates mainly on the base of the cilia.[1] Actually, in vertebrates there exist Ptc1 and Ptc2, which are expressed in different tissues.[9]

The oncoprotein Smo is a seven transmembrane protein GPCR which can adopt three different conformations depending on its phosphorylation stage:[10] Smo A is a cytoplasmatic inactive form; Smo B is also an inactive conformation but is bound to the cilia; and finally Smo C is the active form and localizes at the cilia.[1] The change from inactive to active conformation in Smo, depends on sequenteial phosphorylation steps carried out by protein kinase A (PKA).[11]

Another key component of the Hh signaling pathway are the Gli proteins, which pertain to the zinc finger transcription factor family. In vertebrates there are three Gli proteins: Gli1, Gli2 and Gli3. All of them contain a highly conserved DNA binding domain that comprises five zinc finger domains. While Gli 1 and Gli 2 act as the main activators at the last step of Hh pathway, Gli 3 plays a repressor role.[12]

The activation of the Hh signaling pathway depends, on the presence or absence of ligand proteins, ShhN is the main ligand of the pathway.

In the **absence of ligand**, Ptc remains active and catalytically inhibits Smo, which is mainly localized in vesicles. The repression of Smo results in the lack of Hh signaling. Although the mechanism by which Ptc regulates the oncoprotein has not been elucidated, some hypothesis suggest the implication of a small molecule.[1,13] Recently, Huang *et al.* pointed that this molecule effector could be an oxysterol.[14] In their work, they found that oxysterols seemed to bind to Smo on its Cysteine Rich domain (CRD), and that this trigger the conformational change into the active Smo C.

On the other situation, when the **ligand Shh is present** (**Figure *1***), it is able to bind to Ptc causing its inactivation and internalization. Thus, Smo is no longer inhibited, and changes into its active conformation that translocates to the primary cilium on the cell membrane and then activates the signaling cascade. This finally results on the activation of genes Gli1 and Gli2, and decrease of Gli3, causing the accumulation of Gli1 protein transcription factor, that translocates to the nucleus and triggers the transcription of different genes.[13] The main genes activated by Gli are **Cyclin D2,**[12,15] which promotes the transition from a quiescent to a proliferative state in cells; **Nmyc1 proto-oncogen**, that induces cell proliferation;[12,16] the anti-apoptotic factor **BCL2;**[12] and **snail**, a gene that has been associated to increase of invasiveness and metastasis.[12,17] Other genes activated downstream are Gli1, Ptc1 and Hedgehog Interacting Protein (HhIP), which are negative regulators of the Hh pathway.[1]

How Shh inhibits Ptc is not clearly understood, although it is thought that the residues on the extended conformation may trap the receptor into its inactive form, causing its internalization [13]. It has been also suggested that Shh may also bind to co-receptors that are found forming a complex with Ptc, such as the glypican-family of cell surface proteoglycans, the Cell adhesion molecule-related/downregulated by oncogens (CDON), Brother of Cdo (Boc) and the glycosylphosphatidylinisotol (GPI).[18–20] The binding of ShhN to these molecules had been hypothesized to mediate the concentration of ShhN to the base of the cilia where the Ptc localizes. [13]

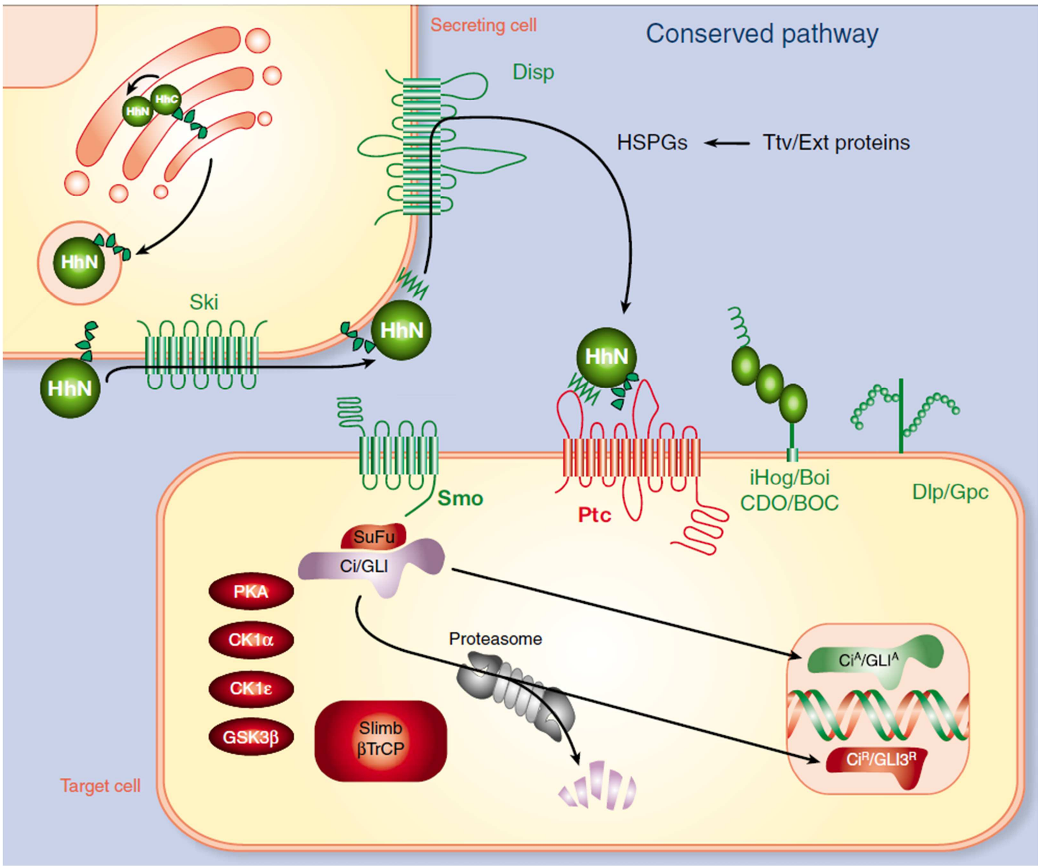

Secreting cell
Conserved pathway
Disp
HSPGs
Ttv/Ext proteins
HhN
HhC
Ski
Smo
Ptc
iHog/Boi
CDO/BOC
Dlp/Gpc
SuFu
Ci/GLI
PKA
CK1α
CK1ε
GSK3β
Slimb
βTrCP
Proteasome
Ci^A/GLI^A
Ci^R/GLI3^R
Target cell

Figure 1. Activation of the Hh pathway in the presence of Shh. Ski is the HHAT in invertebrates. Ci (Cubitus interruptus) is the paralog of Gli in invertebrates. Adapted from [21].

The mechanisms by which Gli genes are activated upon the presence of Hh ligands has not been described. However, it has been suggested that in the presence of Shh, the Supressor of protein Fused (SuFu), a protein that negatively regulates Gli, no longer recruit Gli proteins in the cytoplasm, and it allow their translocation to the nucleus.[1,12,22]

2. Biosynthesis of Sonic Hedgehog

Hedgehog proteins have to undergo post-translational modifications in order to become active. Actually, all Hh proteins suffer two post-translational modifications that culminate with the attachment of two lipid residues (**Figure 2**): a palmitic acid and a cholesterol molecule.[13] Since of Shh is the main ligand of the Hh pathway, we will mainly focus on this protein.

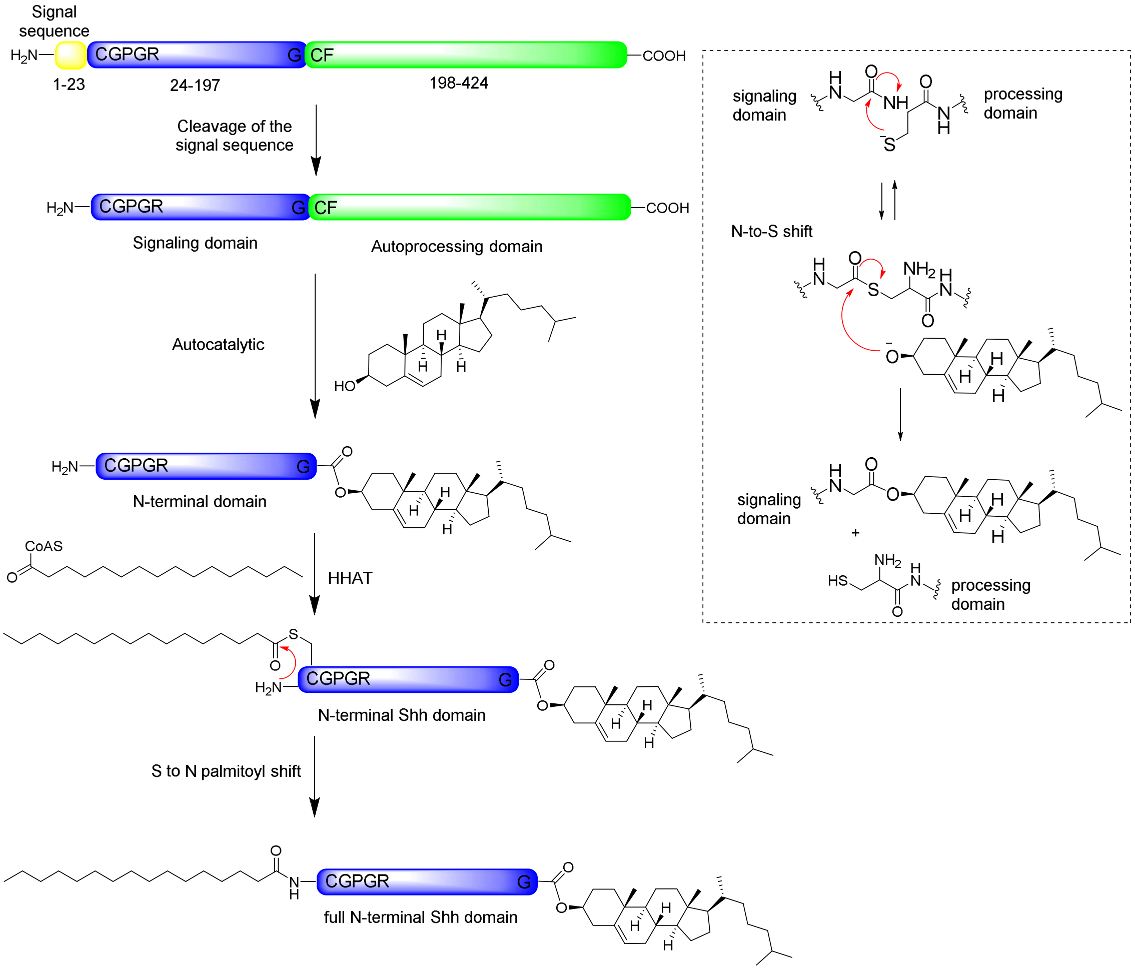

Signal sequence
H2N
CGPGR
G
CF
COOH
1-23
24-197
198-424
Cleavage of the signal sequence
Signaling domain
Autoprocessing domain
Autocatalytic
HO
H
N-terminal domain
CoAS
O
HHAT
S
N-terminal Shh domain
S to N palmitoyl shift
full N-terminal Shh domain
signaling domain
processing domain
NH
N-to-S shift
NH2
signaling domain
+
HS
processing domain

Figure 2. Post-translational modifications of Shh. Modified from [1].

Shh is biosynthesized originally as a 454-mer protein. It contains an N-terminal signal sequence that after removal releases the Shh precursor, a 45 kDa protein.[1] This still inactive protein would undergo the post-translational modifications in two steps. The first step consists on the splicing of the protein catalyzed by its C-terminal domain. This cleavage occurs between a Cys and a Gly residues, involving first an *N* to *S* acyl shift rearrangement of the amide bond to render a thioester. Next, the hydroxyl group of a cholesterol molecule attacks the carbonyl group of the thioester releasing on one hand the C-terminal domain and on the other, the N-terminal part of Shh (from now on we would refer at this active form as ShhN) with a cholesterol moiety attached through an ester bond.[1,2] The already released C-terminal domain does not display any further function or signaling activity in the Hh pathway and it is degraded by the proteasome.[23]

Next, the N-terminal active domain is translocated to the cytoplasmatic membrane, facilitating the next post-translational modification step. This is carried out by the Hedgehog Acyl Transferase (HHAT), an enzyme localized at the plasma membrane that catalyzes the attachment of a palmitic residue at the N-terminal Cys of the ShhN. HHAT requires a conserved sequence, CysGlyProGlyArg, on ShhN to recognize the protein and transfer the palmitic residue. Unlike other palmitoylated proteins, in which the palmitic residue is attached through the thiol of a Cys group, in ShhN, it is attached through an amide bond. In Shh, HHAT first attaches the palmitic residue at the thiol group, thus forming a thioester, which rearranges to finally yield an amide bond between the lipid and the protein domain.[1,2]

The final active Shh consists on a 20 kDa protein with two lipid residues attached: a cholesterol at the C-terminus and a palmitic at the N-terminus entailing the active ShhN.

ShhN secretion is carried out by a 12 transmembrane protein, called Dispatched.[2] Outside of the cell, ShhN has autocrine or paracrine activity, and also be transported to distant tissues. Although the proteins involved in ShhN transport have not been identified yet, it has been hypothetized that cholesterol carrying proteins may be implicated.[24]

3. Structure of Shh

Shh is mainly a globular protein consisting on a combination of α-helixes and β-sheets. However, its X-ray crystal (**Figure *3***) shows that the first N-terminal 15 residues project almost 30 Å away from the globular core.[25] Although the X-ray was determined from a murine ShhN lacking the lipid modifications, it is expected that the ShhN harboring the lipid moieties would adopt a similar structure. Thereby, the palmitic acid could extend even further away from the main core of the protein.

Salic *et al.* reported that while the globular part of ShhN is responsible for the tight binding to its receptor, the N-terminal residues (from Cys1 to Lys15) adopt an extended conformation that carry out the inactivation of the receptor.[13] Essentially, they hypothesized that the binding of ShhN towards its receptor occurs via these two regions that are separated from each other:

the globular part that bind to the high-affinity site of Ptc, and the extended arm, which binds to the loop2 from Ptc, known as the efector site.[13,26]

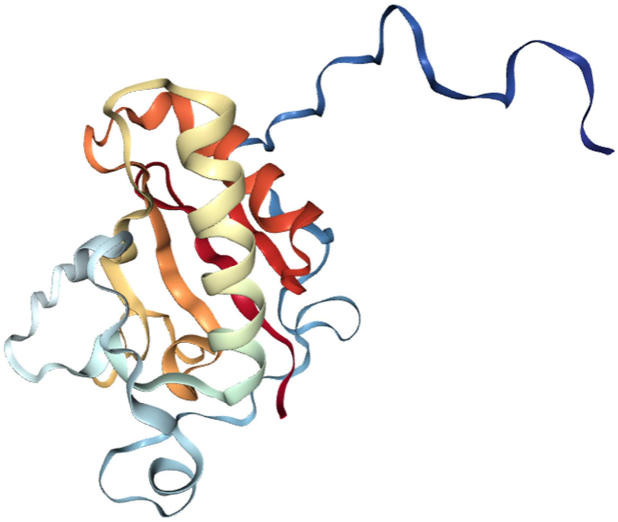


Figure 3. Crystal of the N-terminal domain of Shh without lipid modifications. Image extracted from Protein Data Bank [www.rcsb.org] Reference in PDB: 3M1N.[27]

3.1. Hydrophobic modifications in Shh

Extensive studies were carried out to determine the role of the lipid residues attached to ShhN.

3.1.1. Modifications at the N-terminus

It is known that the N-terminal arm of ShhN is the main responsible for its activity.[13] This is also supported by the fact that the first 15 residues of Hh proteins are highly conserved between species, thus, indicating that they may play an important role on the activity or interaction with the receptor.[28] Taylor *et al.* explored the effect of different hydrophobic modifications on ShhN (**Table *4***).[28] After the modification of the N-terminus of Shh with hydrophobic lipids of different length, they observed that ShhN activity increased along with the length of the chain. Moreover, they reported that the palmytoilated ShhN was a 100-fold more active than the protein lacking the palmitic residue. Accordingly, long chain fatty acids, the addition of hydrophobic aminoacids at the N-terminus appear to enhance the activity of ShhN when compared to the unpalmitoylated protein [28]. It was also shown that the N-terminal Cys was not required for ShhN activity, concluding that its main role is during palmitoylation.[13,28]

Chemical modification	EC_{50} (nM)	Relative to unmodified
Myristoyl	0.6	160x
Palmytoyl	2.5	40x
Lauroyl	2.5	40x
N-octylmaleimide	2.5	40x
N-pyrenylmaleimide	5	20x
Decanoyl	5	20x
Thiaproline	12	8x
N-hydroxyphenylmaleimide	15	7x
N-naphtylmaleimide	15	7x
Octanoyl	25	4x
N-isopropylacetamide	50	2x
N-ethylmaleimide	100	1x
C24II*	12	8x
Unmodified cysteine	99	1x

Table 4. Hydrophobic modifications at the N-terminus of Shh and their correlation with the protein activity. *ShhN containing isoleucine dipeptide at the N-terminus.[28]

3.1.2. Modifications at the C-terminus

Whereas modifications at the N-terminus directly correlated with ShhN activity, this trend was not observed when substitutions where conducted at the C-terminus. Hydrophobic modifications at the C-terminal part of ShhN did not enhance its actvity, thus confirming that the N-terminal region is playing the major role on ShhN activity.[28]

4. Functions of Hh and related diseases

Hh signaling is critical for embryogenesis during which controls the patterning of the body. The three Hh proteins, Dhh, Shh and Ihh carry out similar physiological functions, but they present different patterns of expression.[29] For instance, Dhh is found in gonads, sertoli cells in the testis, and also in granulosa cells in the ovaries.[30] Meanwhile Ihh is mainly present in the primitive endoderm, gut and prehypertrophic chrondocytes.[2] On the other side, ShhN, the most expressed protein of the Hh pathway, can be found in a wide variety of tissues during embryogenesis. For example, it is expressed in midline tissues like the node, notochord and floor plate, controlling the patterning of left-right and dorso-ventral axes of the embryo. Moreover, Shh is expressed in the zone of polarizing activity of the limb and thereby plays an important role on correct formation of limbs, skeleton, muscle, skin, eyes, lungs, teeth, nervous system, intestines, and differentiation of sperm and cartilage.[1,2] In later stages, during organogenesis ShhN is also found in epithelial tissues.[2]

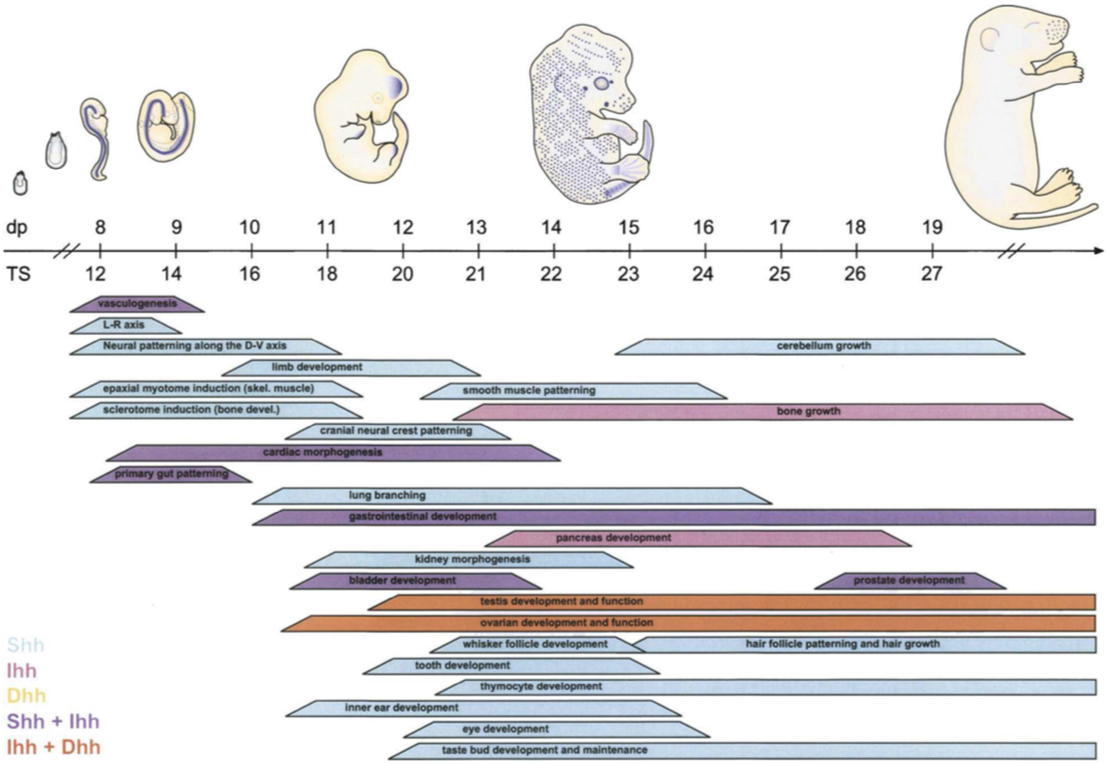

dp
8
9
10
11
12
13
14
15
16
17
18
19
TS
12
14
16
18
20
21
22
23
24
25
26
27
vasculogenesis
L-R axis
Neural patterning along the D-V axis
cerebellum growth
limb development
epaxial myotome induction (skel. muscle)
smooth muscle patterning
sclerotome induction (bone devel.)
bone growth
cranial neural crest patterning
cardiac morphogenesis
primary gut patterning
lung branching
gastrointestinal development
pancreas development
kidney morphogenesis
bladder development
prostate development
testis development and function
ovarian development and function
whisker follicle development
hair follicle patterning and hair growth
tooth development
thymocyte development
inner ear development
eye development
taste bud development and maintenance
Shh
Ihh
Dhh
Shh + Ihh
Ihh + Dhh

Figure 19. Functions carried out by Hh proteins during embryogenesis. The embryo cartoons (top) show aspects of the expression of the Hh receptor during mouse embryonic development. Bars in different colors show the embryonic stages when ShhN, Ihh and Dhh control developmental processes in the indicated tissues. Adapted from [2].

To ensure a proper signaling during this development, some of the functions carried out by the Hh proteins are redundant. As an example, Ihh and ShhN act together during vasculogenesis and cardiac morphogenesis in early developmental stages, and later, their synergy is responsible for the gastrointestinal development.[2]

Due to the important role of Hh during embryogenesis, leading mutations or alterations in Hh pathway results in insufficient signaling, it leads to severe malformations such as holoprosencephaly,[31] polydactyly, skeleton defects and craniofacial or limb malformations.[2]

Although in adults the level of Hh signaling is reduced, it still plays important functions. For instance, it is involved in the maintenance of stem cells and also regulates cell growth, survival and fate. In addition, Hh is reactivated when tissue damage occurs, thereby signaling for tissue repair and regeneration.[1,32] It is pretty obvious that alterations in such an important pathway, implied in cell growth and survival, would be related to cancer.

5. Role of Hh in cancer

Aberrant signaling in Hh has been related to different cancers such as basal cell carcinoma,[33] medulloblastoma,[34] rhabdomyosarcoma,[35] glioma,[36] breast,[37] prostrate,[38] esophageal,[39] gastric,[40] pancreatic,[41] and small cell lung carcinoma[42]. (**Figure *21***). Another disease related to constantly activation of the Hh pathway is the Gorlin syndrome.[43] Patients suffering from this disease, caused by a mutation on the Ptc receptor, present high incidence of basal cell carcinoma, medulloblastoma, and rhabdomyosarcoma.

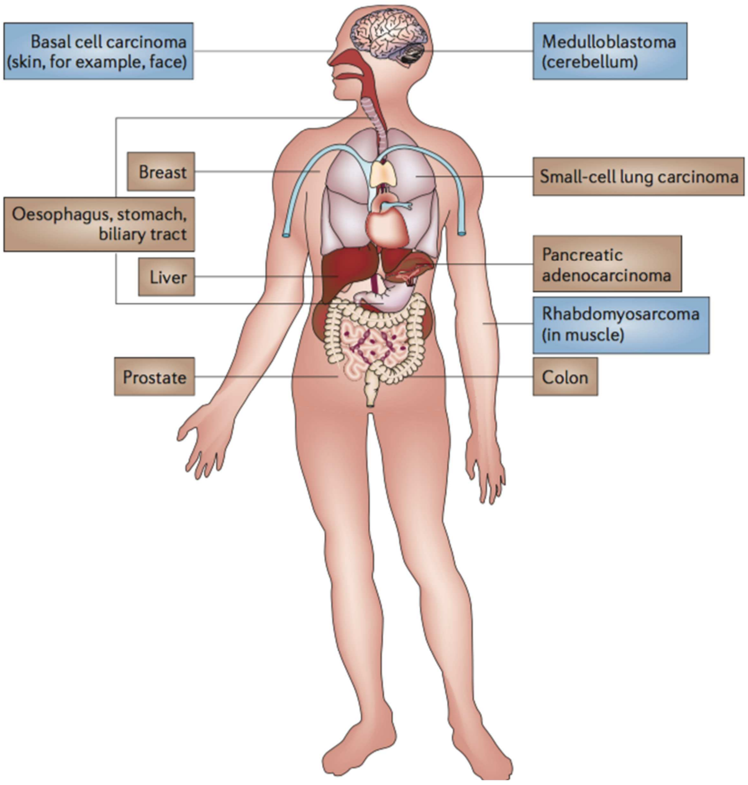

Basal cell carcinoma (skin, for example, face)
Medulloblastoma (cerebellum)
Breast
Small-cell lung carcinoma
Oesophagus, stomach, biliary tract
Pancreatic adenocarcinoma
Liver
Rhabdomyosarcoma (in muscle)
Prostate
Colon

Figure 21. Type of cancer with possible Hh deregulation. Adapted from [32].

The constant activation of Hh can be independent or dependant on ligand. ShhN ligand-dependent activation happens through three different mechanisms: autocrine, paracrine or inverse paracrine manner (**Figure *22***).

In the **autocrine** activation, the tumor cells are the responsible to secret Hh ligands that will act on themselves, thus creating a positive feedback loop. This type of activation has been related to breast, colorectal and non-small cell lung cancer.[20]

During the **paracrine** activation, the tumor cells are the ones secreting Hh ligands that will stimulate the stroma cells surrounding the tumor. Thereby, this environment nearby the tumor is the responsible for nurturing the cancer cells by the stimulation of angiogenesis, effects on the extracellular matrix, secretion of components of other molecular signaling pathways such as insulin growth factor (IGF) and Wnt signaling pathway.[20,44] This activation is mainly found in pancreatic cancer.

Finally, when the activation of Hh pathway is through an **inverse paracrine** manner, the stroma cells are the ones secreting Hh ligands that would bind and activate the signaling cascade on tumor cells. Multiple myeloma and lymphoma had been associated to this type of Hh activation.[20]

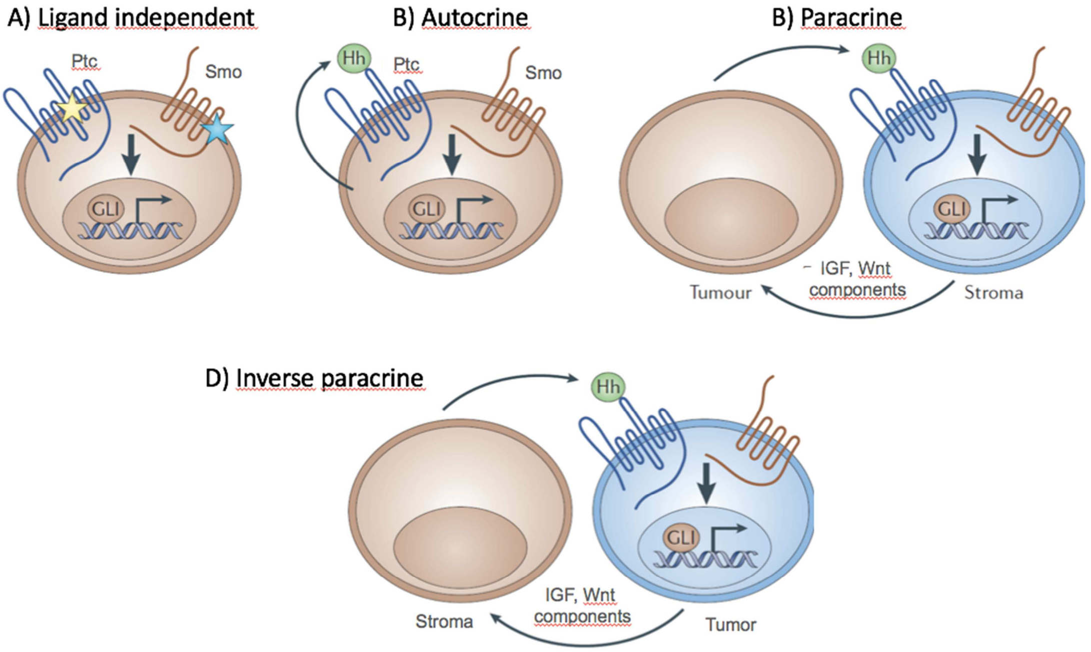
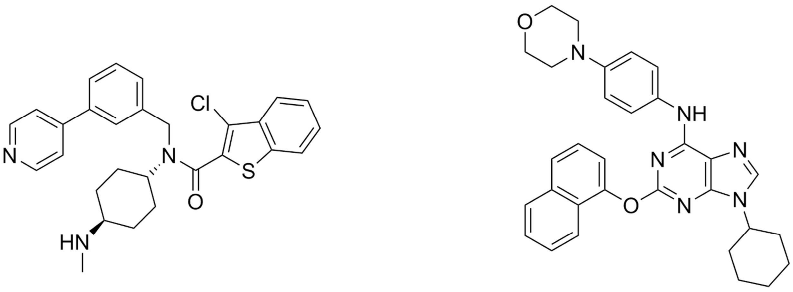


Figure 22. Models of Hh activation in cancer. A) independent on the ligand, mutations on Ptc or Smo lead to constant activation of Hh pathway. B) Autocrine: ShhN secreted for tumor cells acts on themselves. C) Paracrine: Tumor cells secret ShhN that act on the stroma which nurtures the cancer cells. D) Inverse paracrine: Stroma cells secret ShhN that activates Hh cascade in the tumor cells. Adapted from [32].

6. Inhibitors and agonists of Hh

In the last years, the Hh pathway has gained interest as a druggable pathway. On one side, inhibitors of this signaling cascade would be possible anticancer drugs, while on the other side, stimulators of Hh may find applications as angiogenesis inducers, promoters of wound healing and also become useful in stem cell research.[1] Although activators of Hh such as Smoothened

Agonist (SAG) and purmorphamine (**Figure *20***), have been reported, their use is restricted only to research purposes.[45]



SAG

Purmorphamine

Figure 20. SAG and Purmorphamine chemical structure.

In fact, there some drugs that inhibit the Hh pathway are currently being used for the treatment of cancer, while others have just entered on clinical trials. These outcomes have encouraged the search of new inhibitors of this pathway.

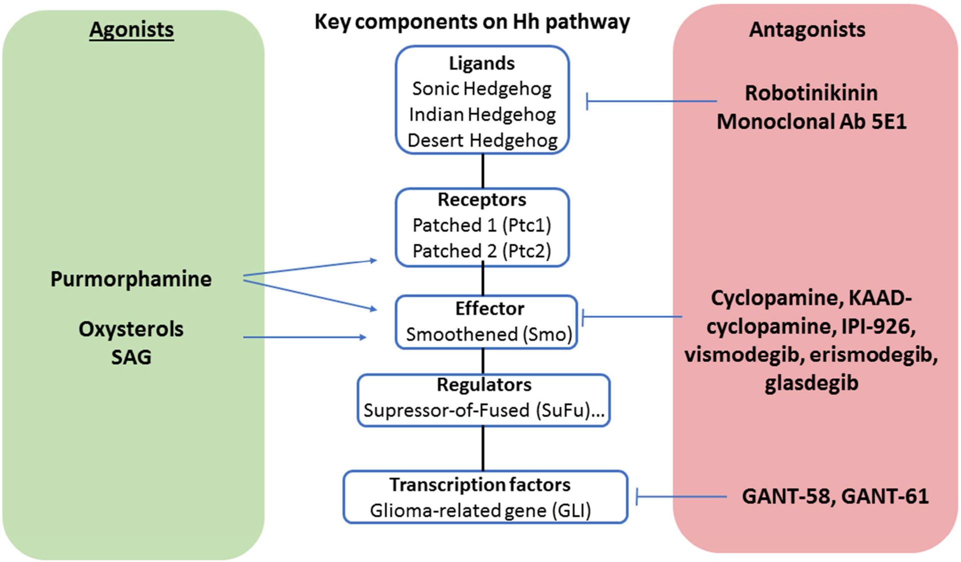

Agonists
Key components on Hh pathway
Antagonists
Ligands
Sonic Hedgehog
Indian Hedgehog
Desert Hedgehog
Robotinikinin
Monoclonal Ab 5E1
Receptors
Patched 1 (Ptc1)
Patched 2 (Ptc2)
Purmorphamine
Oxysterols
SAG
Effector
Smoothened (Smo)
Cyclopamine, KAAD-cyclopamine, IPI-926, vismodegib, erismodegib, glasdegib
Regulators
Supressor-of-Fused (SuFu)...
Transcription factors
Glioma-related gene (GLI)
GANT-58, GANT-61

Figure 21. Scheme of the main agonists and antagonists of the Hh pathway.[46]

All the Hh inhibitors described until the moment, with the exception of a monoclonal antibody,[46] are small molecules. Moreover, most of them target Smo oncoprotein (**Figure *21***). Inhibition of Smo and blocking Hh ligands become useful for cancers which aberrant signaling is caused upstream of Smo. Targeting components downstream Smo, like the Gli transcription factors, may be ideal to inhibit Hh pathway independently on the cause of its aberrant signaling. Smo, Ptc or Hh ligands localize outside or at the cell membrane and thereby, can be targeted without the need to enter the cell. Contrary, targeting components downstream Smo would require the inhibitor to enter into the cell, which become a limitation for some molecules.

6.1. Smo inhibitors

6.1.1. Cyclopamine and derivatives

Cyclopamine is a natural alkaloid found in veratrum californicum that blocks the activation of the Hh pathway.[47] This small molecule targets Smo preventing its translocation to the primary cilium.[1] Cyclopamine has been used in clinical trials for the treatment of basal cell carcinoma (BCC). However, due to its poor pharmacokinetic properties, derivatives with improved properties have been developed (**Figure** ***22***). For instance, KAAD-cyclopamine presents less cytotoxicity.[48] Another derivative from cyclopamine, **IPI-926**, is in ongoing clinical trials in phase I and II.[46] Actually IPI-926 has been assayed in patients with advanced or metastatic solid tumor, and in combination with other anticancer drugs for the treatment of pancreatic cancer.

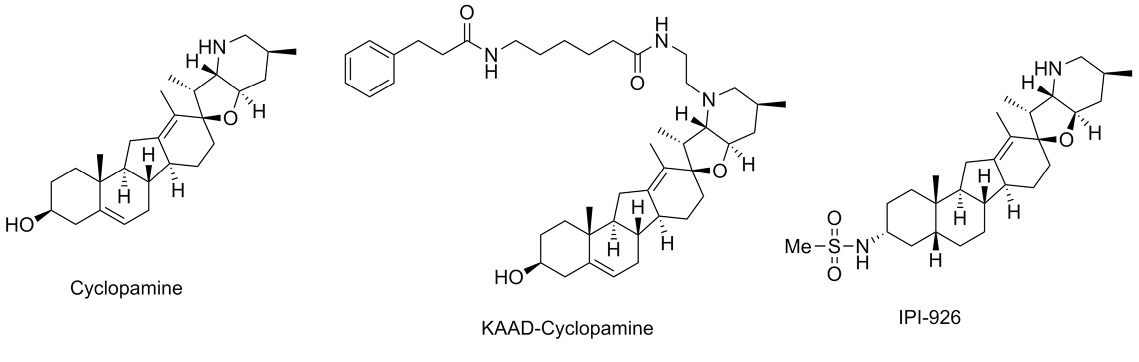


Figure 22. Cyclopamine and derivatives.

6.1.2. Other small molecules

In 2012, **Vismodegib** was the first Hh inhibitor approved by the FDA.[49] This small molecule targets Smo oncoprotein and it is indicated for the treatment of BCC. Another similar small molecule, **erismodegib**, was also approved by the FDA in 2015 for the treatment of BCC in patients with recurrent disease.[50] Similar to these already approved molecules, Pfizer has recently entered **glasdegib** on phase II clinical trials (**Figure *23***).

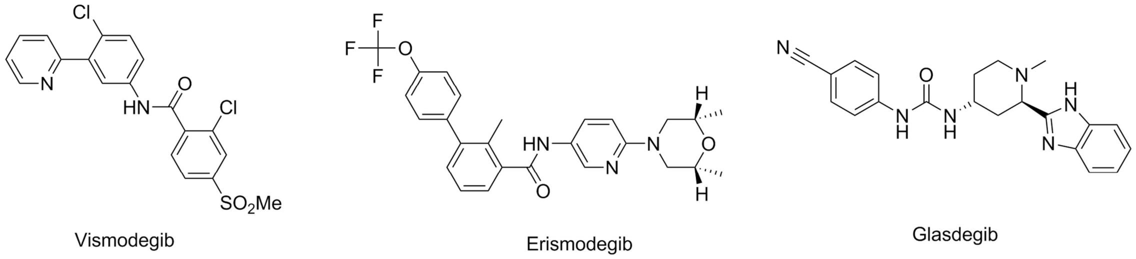


Figure 23. Chemical structure of vismodegib, erismodegib and glasdegib.

6.2. Shh inhibitors

6.2.1. Monoclonal antibody 5E1

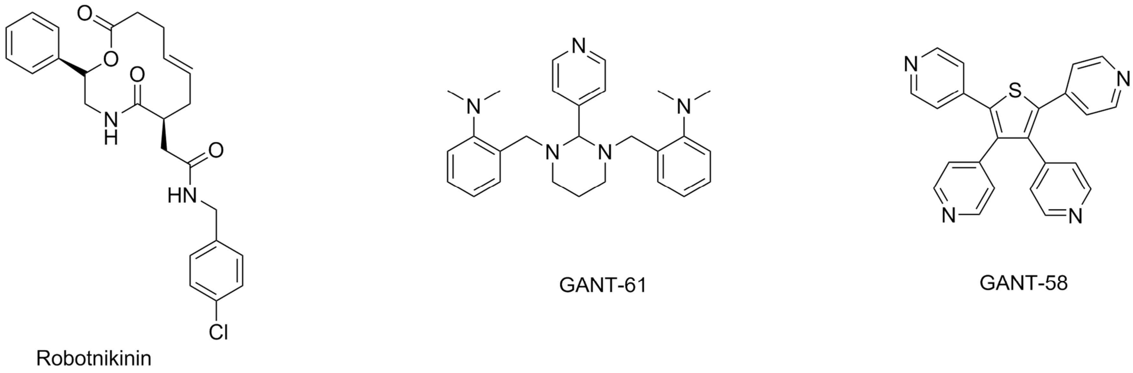
Monoclonal antibody **5E1** binds to Shh and prevents the interaction with its receptor Ptc.[51,52] Although it has been reported that 5E1 blocks the growth of tumors in mice models, so far it has just been used in research.[1]

6.2.2. Robotnikinin

Until the moment, **Robotnikinin** (**Figure *24***) is the only small molecule described to inhibit the Hh pathway by binding Shh. In cellular inhibition experiments it has shown a IC_{50} of 3.1 µM.[53]

6.3. GLI inhibitors

Molecules targeting GLI proteins have also been described. **GANT-58** and **GANT-61** (Error! Reference source not found.) block the activation of Hh signaling by preventing the binding of GLI proteins to DNA.[54]



Robotnikinin

GANT-61

GANT-58

Figure 24. Chemical structure of robotikinin and Gli antagonists.

7. Importance of Hh and ShhN

The implication of Hh pathway in several cancers, such as the BCC, the most common cancer in the western world,[32] have increased the interest in this signaling cascade. Very recently, it was discovered the significance of the Hh pathway on the renewal and differentiation of cancer stem cells (CSC)[55] which are the responsible of chemoresistance, and the reinitiation of the tumor. They can acquire migratory properties that contribute to metastasis, as well. So, potential Hh inhibitors be useful for the treatment of solid tumors and also find further applications in reducing metastasis.

The research resulted in the commercialization of two molecules for treatment of BCC, vismodegib and erismodegib, and the entrance of others in clinical trials. Indeed, because most of the cancers related to Hh are ligand dependent, suggested that the finding of a molecule that could block the interaction of the ligand to its receptor, could achieve broad use in the therapeutic field.

Due to the importance of ShhN, in the present thesis we envisaged that its chemical synthesis would remarkably contribute to study of Hh pathway. For instance, the synthesis of the active ShhN harboring the lipid residues would help to elucidate its structure and some unanswered mechanisms of the Hh signaling pathway. Moreover, since just two molecules are known to bind ShhN and block its interaction to Pct, we thought that the discovery of new ShhN binders would be an important contribution in the therapeutic field. In order to accomplish this goal, we considered phage display as a very suitable technique to identify possible ShhN antagonists.

OBJECTIVES

Aberrant Hh signaling has been related to several types of cancer, which has raised the interest in this pathway. During the last years, numerous efforts had been carried out to develop inhibitors of the Hh cascade. So far, only two inhibitors targeting Smo have successfully reached the market and are currently used for the treatment of BBC. Considering that the majority of the cancers related to Hh are ligand dependent, only one small molecule and a monoclonal antibody have been reported to efficiently recognize the ShhN protein and block its interaction with the receptor. Therefore, we thought that new ShhN antagonists could be promising Hh inhibitors.

We envisage that chemical synthesis of ShhN protein would efficiently afford the active ShhN. Moreover, it would allow us to carry out further studies in order to find new antagonists using techniques based on DNA encoded combinatorial peptide libraries, like phage display or mirror image phage display.

The main objectives of this chapter are:

1. The use of *p*CN-Phoc-MeDbz peptides for the synthesis of medium-size proteins (150-200 amino acids).

2. The chemical synthesis of the ShhN protein using NCL and KCL applying the *p*CN-Phoc-MeDbz methodology previously developed.

3. The chemical synthesis of a ShhN analog carrying a biotin residue, which would allow further biological studies.

4. The proper folding of the synthetic ShhN and analogs, ensuring their activity.

RESULTS AND DISCUSSION

The natural ShhN, a 174-mer protein bearing a palmitic acid at the N-terminus and a cholesterol moiety at the C-terminus, represents a challenging synthesis because of the presence of the two lipidic moieties. In general, the synthesis of lipidated or membrane proteins is problematic due to its hydrophobicity. Despite some chemical synthesis using NCL, most of the lipidated proteins have been prepared by expressed protein ligation,[56–60] in where the peptide building block containing the lipid was usually synthetized on solid or solution phase and then ligated to the recombinant protein. Several other strategies that allow the access to this type of proteins have been also described like the maleimidocaprolyl ligation, Diels-Alder, azide-alkyne Cu-catalyzed click and sortase-mediated protein ligations.[61] The limited solubility of the peptide fragment bearing the lipid residue is the common drawback in all of these approaches. In order to overcome the aggregation or precipitation during the synthesis of the protein, it has been extended the use of detergents, or the introduction of solubilizing tags like PEG, poly-arginine or maltose binding protein.[57,62]

1. Synthesis of natural ShhN: Palm-ShhN-cholesterol

1.1. Retrosynthesis of ShhN: strategy A

The main consideration when devising the retrosynthetic analysis was the determination of suitable ligation sites, either at natural Cys or other residues, and the length of the peptide fragments, which ideally should be around 25 to 50 residues. ShhN contains three Cys in positions 1, 79 and 160. Whereas Cys^{1} was discarded, the other natural cysteines could be potential ligation sites. Dividing the protein in positions 79 and 160 gives three fragments Cys^{1}-Arg^{78}, Cys^{79}-His^{159} and Cys^{160}-Gly^{174}. With the exemption of the last fragment, the others have a size larger than desired for SPPS. Therefore, it was searched for Ala residues located at suitable positions. Then, they would be exchanged by Cys, and reduced to natural Ala after ligation. In addition, to facilitate the synthesis of the protein and because it was known that Cys^{1} does not play an important role for binding to the receptor,[13] it was decided to mutate it into an Ala.

Finally, the ShhN sequence was divided into five peptide fragments: Ala^{1}-Gly^{34}, Ala^{35}-Arg^{78}, Cys^{79}-Arg^{121}, Ala^{122}-His^{159} and Cys^{160}-Gly^{174} (**Figure *1***). With the exemption of the last fragment, which is a 15-mer, the other sequences are in the range of 34 to 44 residues.

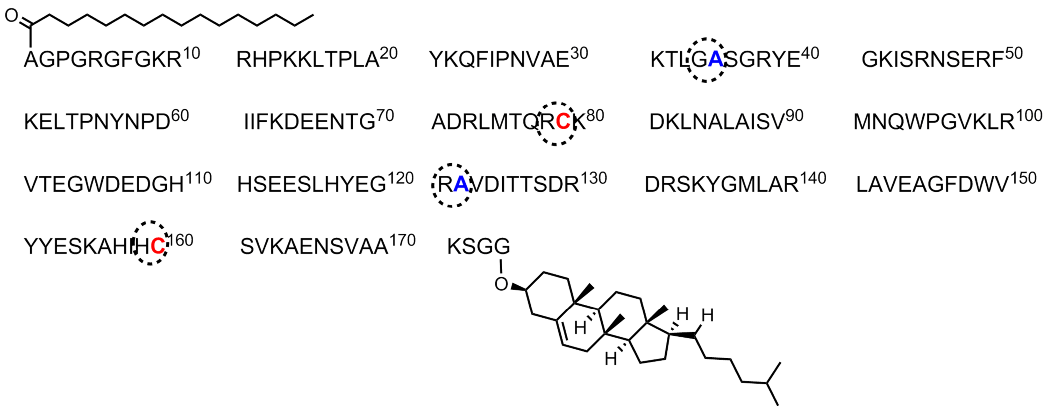

O
AGPGRGFGKR10
RHPKKLTPLA20
YKQFIPNVAE30
KTLGASGRYE40
GKISRNSERF50
KELTPNYNPD60
IIFKDEENTG70
ADRLMTQRCK80
DKLNALAISV90
MNQWPGVKLR100
VTEGWDEDGH110
HSEESLHYEG120
RAVDITTSDR130
DRSKYGMLAR140
LAVEAGFDWV150
YYESKAHHC160
SVKAENSVAA170
KSGG
O
H
H
H
H

Figure 1. Retosynthetic strategy A. Circles and letters in colors indicates ligation sites. In blue, the Ala residues which would be mutated into Cys for ligation purposes. In red, natural Cys used as ligation points.

1.2. Strategy A towards the Palm-ShhN-cholesterol

It was envisaged a convergent approach for the synthesis of Palm-ShhN-cholesterol: the N-terminal side encompassing residues Ala^{1}-Arg^{78} would be condensed in the final step to the C-terminal fragment spanning amino acids Cys^{79}-Gly^{174}

For the synthesis of the N-terminal fragment, the Palm-Ala^{1}-Gly^{34} segment (called **F1)** would be ligated to the Cys^{35}-Arg^{78} peptide (named **F2**). Finally, application of standard desulfurization conditions would reduce the Cys^{35} to the natural Ala^{35}.

In the C-terminal block, the N-terminal cysteine in the Cys^{79}-Arg^{121} (**F3**) fragment would be protected by a thiazolidine (Thz). The **F3** would be ligated to the Cys^{122}-His^{179} fragment (called **F4**). This ligation would give the Thz^{79}-His^{159} product (**F3-F4**), which following desulfurization would restore the natural Ala at position 122. Next, upon linker cyclization to the MeNbz peptide, the **F3-F4** segment would be ligated to the last fragment **F5** (Cys^{160}-Gly^{174}-cholesterol). This ligation would yield the Thz^{79}-Gly^{174}-cholesterol peptide (**F3-F4-F5**). Finally, the Thz would be removed using methoxylamine to give the natural Cys at position 79.

In the last step, ligation between the **F1-F2** and **F3-F4-F5** would give the full ligated Palm-ShhN-cholesterol (**Figure *2***).

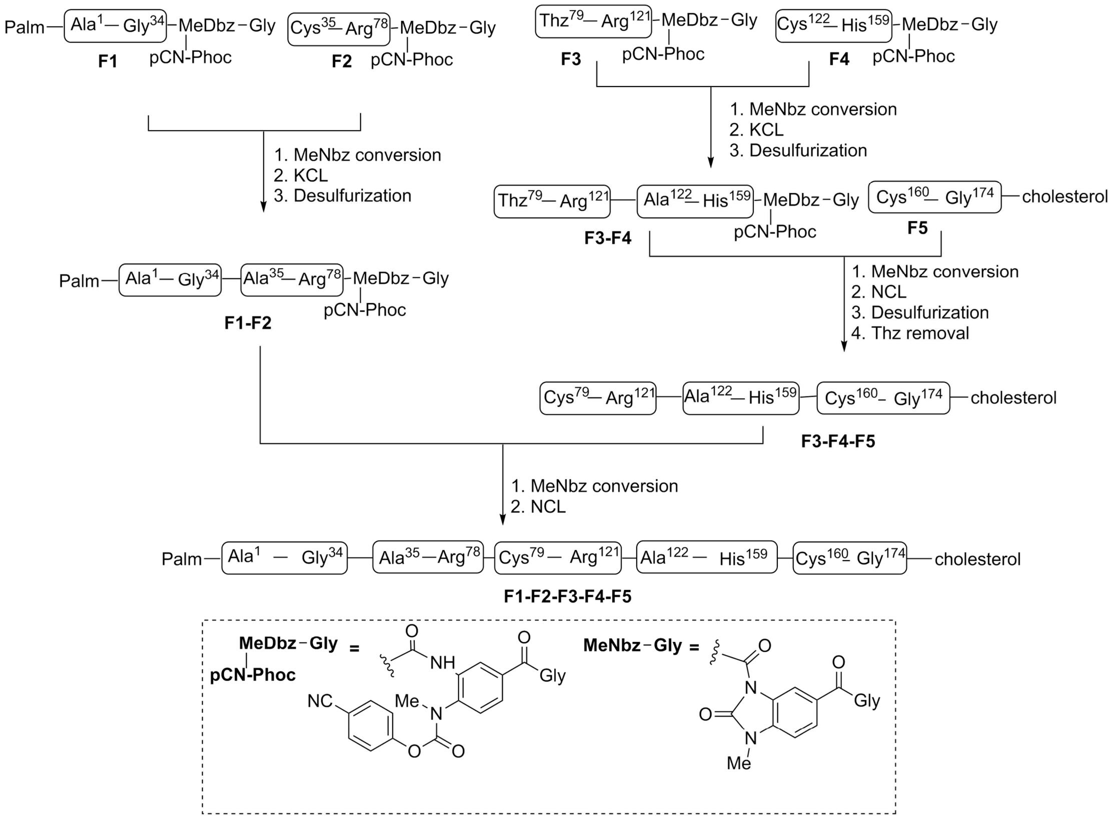

Palm—Ala¹— Gly³⁴–MeDbz–Gly pCN-Phoc F1
Cys³⁵– Arg⁷⁸–MeDbz–Gly pCN-Phoc F2
Thz⁷⁹— Arg¹²¹–MeDbz–Gly pCN-Phoc F3
Cys¹²²– His¹⁵⁹–MeDbz–Gly pCN-Phoc F4
1. MeNbz conversion
2. KCL
3. Desulfurization
Thz⁷⁹— Arg¹²¹—Ala¹²²– His¹⁵⁹–MeDbz–Gly pCN-Phoc F3-F4
Cys¹⁶⁰– Gly¹⁷⁴—cholesterol F5
Palm—Ala¹— Gly³⁴—Ala³⁵— Arg⁷⁸–MeDbz–Gly pCN-Phoc F1-F2
1. MeNbz conversion
2. NCL
3. Desulfurization
4. Thz removal
Cys⁷⁹— Arg¹²¹—Ala¹²²– His¹⁵⁹—Cys¹⁶⁰– Gly¹⁷⁴—cholesterol F3-F4-F5
1. MeNbz conversion
2. NCL
Palm—Ala¹ — Gly³⁴—Ala³⁵—Arg⁷⁸—Cys⁷⁹— Arg¹²¹—Ala¹²²— His¹⁵⁹—Cys¹⁶⁰– Gly¹⁷⁴—cholesterol
F1-F2-F3-F4-F5
MeDbz–Gly pCN-Phoc =
NH Gly NC Me N O O
MeNbz–Gly =
N O Gly N Me

Figure 25. Synthetic strategy toward Palm-ShhN-cholesterol.

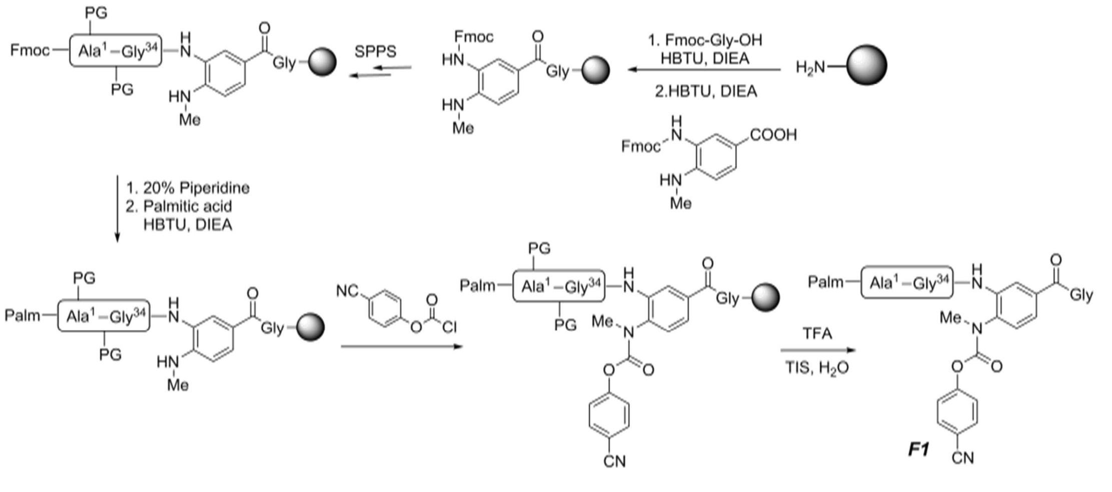
1.3. Synthesis of the peptide fragments

1.3.1. Synthesis of fragment F1: Palm-Ala^{1}-Gly^{34}-pCN-Phoc-MeDbz-Gly

The N-terminal fragment was synthetized following the standard procedure using Fmoc-Rink Amide Resin. After peptide elongation, the palmitic acid was coupled following the standard coupling conditions, but required the solubilization in a mixture of DCM and DMF. Next, the MeDbz linker was acylated with *p*-cyanophenylchloroformate (

Scheme 1.). The peptide was cleaved off the resin and lyophilized.

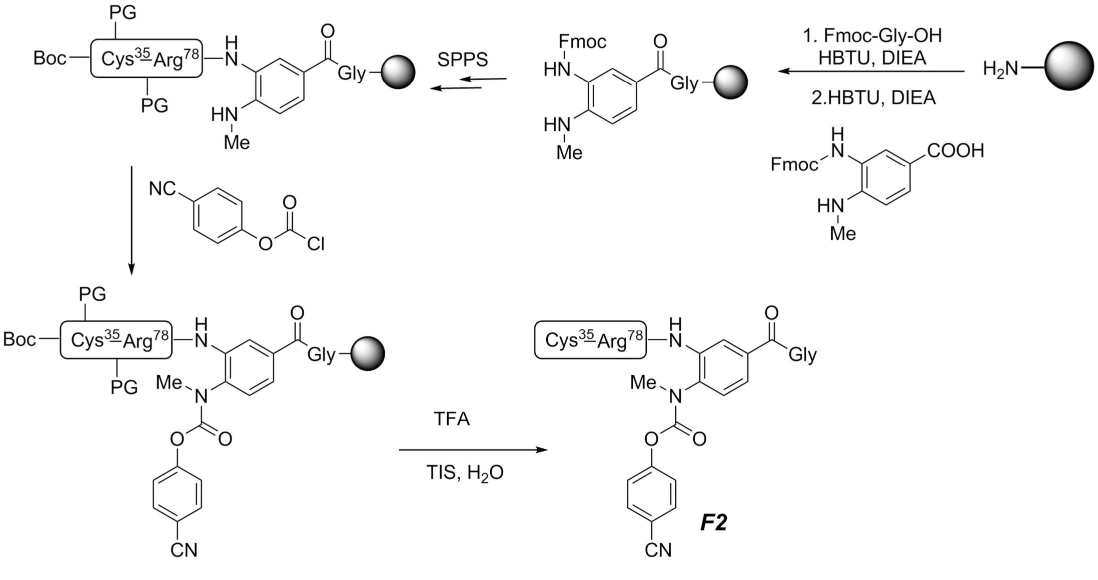
Although this fragment reacts in a KCL manner with **F2**, the main point under the synthesis of **F1** as the *p*CN-Phoc-MeDbz peptide is a major stability of this linker compared to MeNbz in large-scale cleavage. Then, by a simple thioesterification step it can be transformed in a thioester and readily used in NCL.


PG
Fmoc
Ala1–Gly34
Gly
SPPS
1. Fmoc-Gly-OH
HBTU, DIEA
2.HBTU, DIEA
H2N
COOH
HN
Me
1. 20% Piperidine
2. Palmitic acid
HBTU, DIEA
Palm
NC
Cl
TFA
TIS, H2O
CN
F1

Scheme 1. Synthesis of the F1. PG = protecting group.

1.3.2. *Synthesis of the F2: Cys^{35}-Arg^{78}-pCN-Phoc-MeDbz-Gly*

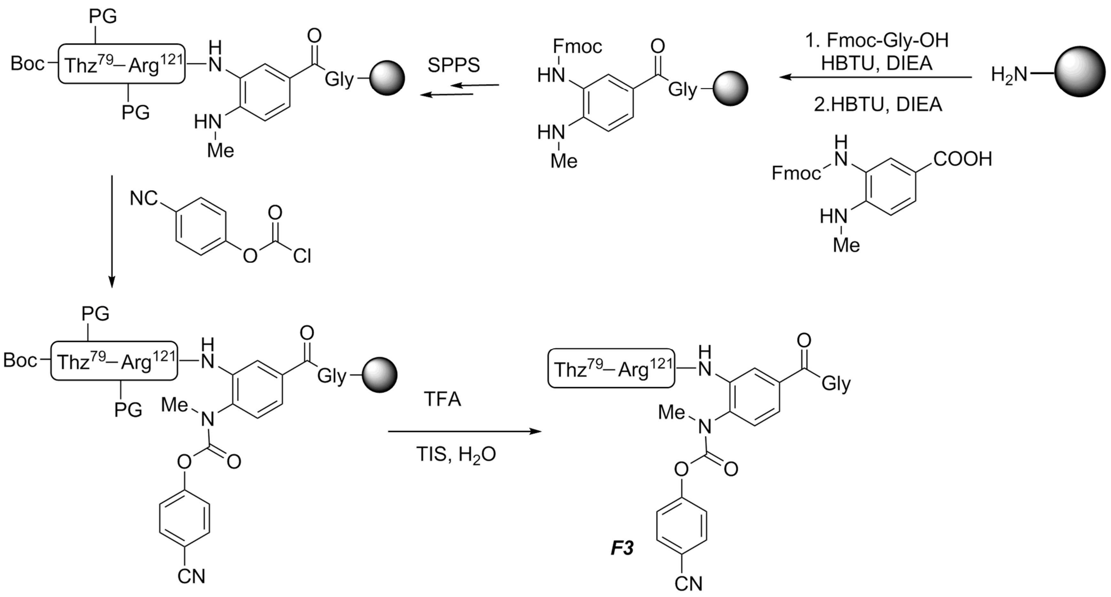
The second fragment is the longest peptide, 44 residues. The natural N-terminal Ala was mutated into a Cys and introduced as Boc-Cys(Trt). The synthesis of this fragment was carried following the standard SPPS protocol (**Scheme *2***).


PG
Boc
Cys35Arg78
Gly
SPPS
Fmoc
1. Fmoc-Gly-OH
HBTU, DIEA
2.HBTU, DIEA
H2N
COOH
HN
Me
NC
Cl
TFA
TIS, H2O
CN
F2

Scheme 2. Synthesis of F2. PG = protecting group.

1.3.3. *Synthesis of fragment F3: Thz^{79}-Arg^{121}-pCN-Phoc-MeDbz-Gly*

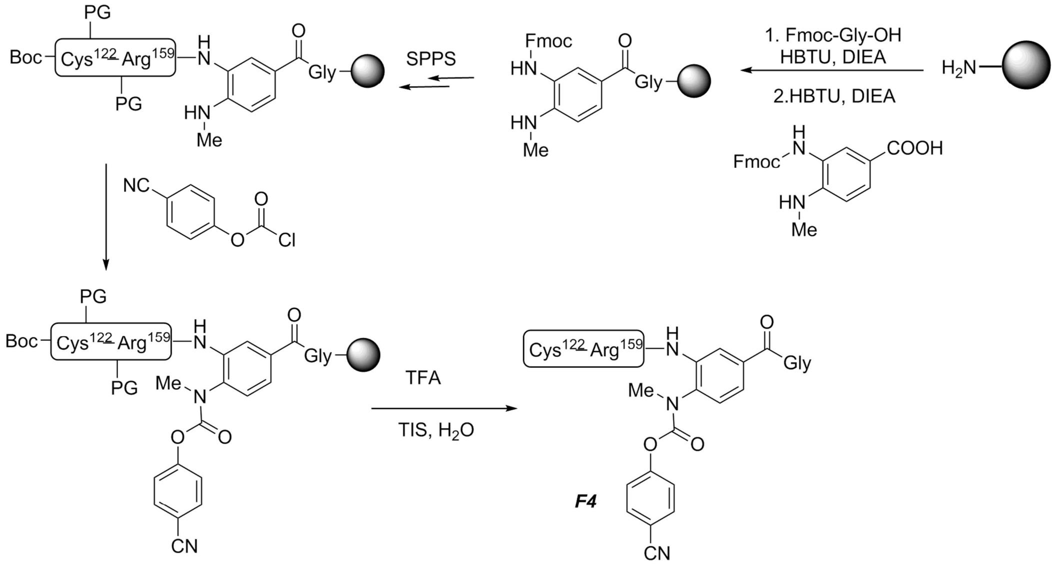
In this fragment the natural Cys at the N-terminus was introduced as a Thz in order to prevent undesired desulfurization. This segment was prepared following the standard procedure (**Scheme *3***).


PG
Boc
Thz79–Arg121
PG
SPPS
1. Fmoc-Gly-OH
HBTU, DIEA
2.HBTU, DIEA
Fmoc
COOH
TFA
TIS, H2O
F3

Scheme 3. Synthesis of F3. PG = protecting group.

1.3.4. Synthesis of Fragment F4: Cys^{122}-His^{159}-pCN-Phoc-MeDbz-Gly

The natural Ala^{122} was mutated into a Cys to allow the ligation with **F3**. This peptide was synthesized following the standard SPPS procedure (**Scheme 4**).

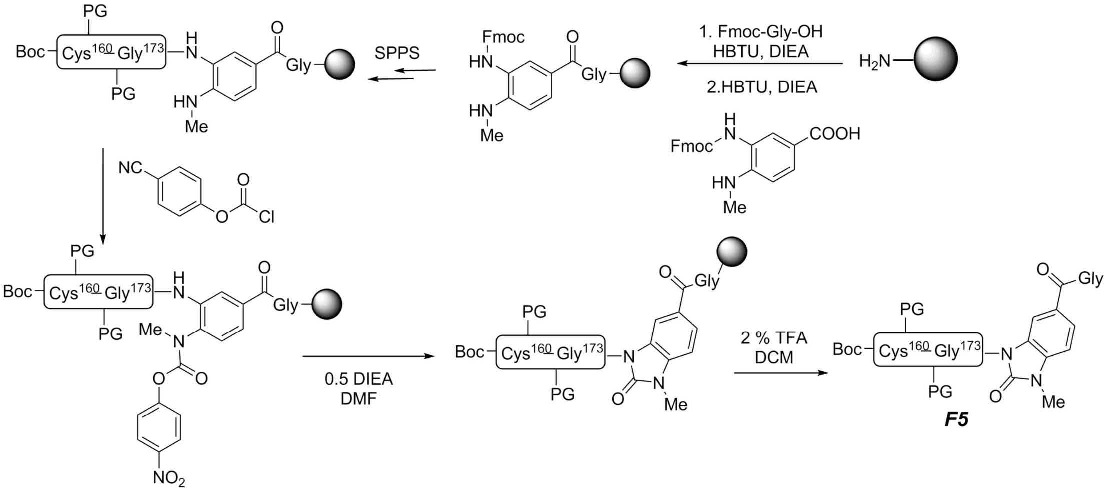

PG
Boc
Cys122–Arg159
PG
SPPS
1. Fmoc-Gly-OH
HBTU, DIEA
2.HBTU, DIEA
Fmoc
COOH
TFA
TIS, H2O
F4

Scheme 4. Synthesis of F4. PG = protecting group.

1.3.5. Synthesis of fragment F5: Cys^{160}-Gly^{174}-cholesterol

1.3.5.1. SPPS of the Cys160-Gly173-MeNbz-Gly

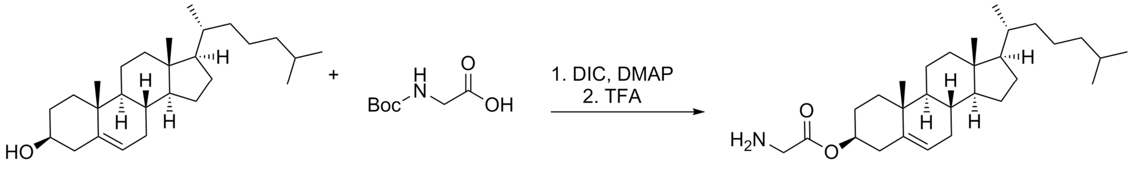
The last fragment is the shortest peptide, a 15-mer. For the synthesis of **F5** it was designed a condensation strategy between the fully protected Cys^{160}-Gly^{174}-MeNbz-Gly segment and Gly^{174}-cholesterol. Synthesis of Cys^{160}-Gly^{173}-MeNbz-Gly was carried out on a 2-chlorotrytil resin. The natural N-terminal Cys was coupled as Boc-Cys. After peptide elongation the MeDbz linker was acylated with *p*NO_2-phenylchloroformate and cyclized using 0.5 M DIEA in DMF to give the MeNbz peptide (**Scheme *5***). The protected peptide was cleaved from the resin using a low TFA cleavage cocktail: 2% TFA, 2% TIS, 2% H_2O, 94% DCM.


PG
Boc
Cys160-Gly173
H
N
O
Gly
PG
HN
Me
SPPS
Fmoc
HN
O
Gly
HN
Me
1. Fmoc-Gly-OH
HBTU, DIEA
2.HBTU, DIEA
H₂N
H
N
Fmoc
COOH
HN
Me
NC
O
O
Cl
PG
Boc
Cys160-Gly173
H
N
O
Gly
PG
Me
N
O
O
NO₂
0.5 DIEA
DMF
O
Gly
PG
Boc
Cys160-Gly173
N
PG
O
N
Me
2 % TFA
DCM
O
Gly
PG
Boc
Cys160-Gly173
N
PG
O
N
Me
F5

Scheme 5. Synthesis of F5. PG = protecting group**.**

1.3.5.2. Synthesis of the Gly-cholesterol building block

Gly^{174}-cholesterol has a cholesterol attached through an ester bond. It was synthesized by coupling the cholesterol to Boc-Gly followed by removal of the Boc protecting group (**Scheme *6***).


HO
H
H
H
H
+
Boc
H
N
O
OH
1. DIC, DMAP
2. TFA
H₂N
O
O
H
H
H
H

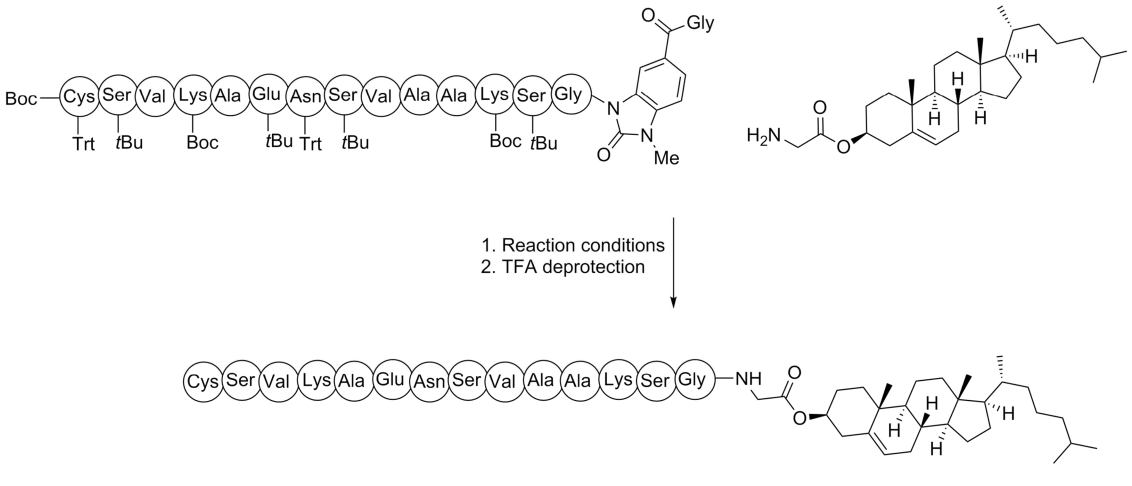
Scheme 6. Synthesis of the Gly-cholesterol building block

1.3.5.3. Condensation of the Gly-cholesterol building block with protected Cys^{160}-Gly^{173}-MeNbz-Gly.

Several reaction conditions were screened to condense the Gly-cholesterol building block with fully protected Cys^{160}-Gly^{173}-MeNbz-Gly (**Scheme *7***): solvents, temperature and equivalents of the building block and DIEA (**Table *1***).

Temp / time	Solvent	Gly-cholesterol	DIEA	Coupling agents	Observations*
rT / 18 h	THF, DMF	3 eq	3 eq	DMAP 2 eq	Product + SM
rT / 18 h	TFT, THF	5eq	30 eq	DMAP 30 eq	SM hydrolyzed
T 70 ^{0}C / 18 h	TFT, THF	5 eq	30 eq	DMAP 30 eq	SM hydrolyzed
rT / 48 h	TFT, THF, DMSO	5 eq	30 eq	DMAP 30 eq	Product + SM hydrolyzed
rT / 18 h	THF, DMF	10 eq	30 eq	DMAP 10 eq	Product + SM hydrolyzed
50 ^{0}C / 4 d	THF, DMF	1.5 eq	50 eq	--	Product + SM
50 ^{0}C / 3 d	THF, DMSO	1.5 eq	50 eq	--	Product + SM

Table 1. Reaction conditions screened. d=days; SM= starting material (Cys^{160}-Gly^{173}-MeNbz-Gly)
*Observations are based on HPLC and ESI/MS analysis.


Boc
Cys Ser Val Lys Ala Glu Asn Ser Val Ala Ala Lys Ser Gly
Trt tBu Boc tBu Trt tBu Boc tBu
Gly
N
O
N
Me
H2N
O
O
H
H
H
H
1. Reaction conditions
2. TFA deprotection
Cys Ser Val Lys Ala Glu Asn Ser Val Ala Ala Lys Ser Gly
NH
O
O
H
H
H
H

Scheme 7. Coupling of the Gly-cholesterol to Cys^{160}-Gly^{173}-MeNbz-Gly.

Unfortunately, none of the conditions assayed afforded the desired product in good synthetic yields. Most of the reaction conditions yielded the hydrolysis of the MeNbz group giving the Cys^{160}-Gly^{173}-OH peptide. In the best scenarios, the final product was formed but the reaction was not completed after long reaction times. Nevertheless, even when the assembled product was observed, it precipitated due to its high hydrophobicity, which made its purification difficult.

1.4. Retrosynthesis of srategy B towards the synthesis of Palm-ShhN-cholesterol

It was thought that the solubility problems could be overcome by coupling the cholesterol moiety to a shorter protected peptide. This forced us to rethink the retrosynthesis in order to have a shorter peptide. Thereby, Ala^{169} was selected as a new ligation point.

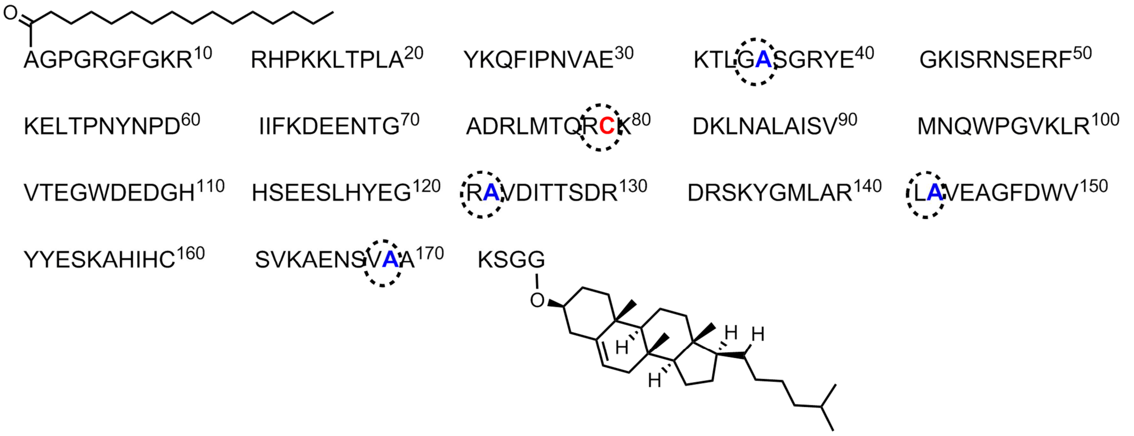

O
AGPGRGFGKR10
RHPKKLTPLA20
YKQFIPNVAE30
KTLGASGRYE40
GKISRNSERF50
KELTPNYNPD60
IIFKDEENTG70
ADRLMTQRCK80
DKLNALAISV90
MNQWPGVKLR100
VTEGWDEDGH110
HSEESLHYEG120
RAVDITTSDR130
DRSKYGMLAR140
LAVEAGFDWV150
YYESKAHIHC160
SVKAENSVAA170
KSGG
O
H
H
H
H

Figure 26. Retrosynthetic strategy B for Palm-ShhN-cholesterol. Circles and letters in colors indicate ligation sites. In blue, the Ala residues which would be mutated into Cys for ligation purposes. In red, natural Cys used as ligation points.

Due to this new fragment the protein was split into six peptide fragments (**Figure *3***). Fragments F1, F2, and F3 did not change, but the rest were redesigned. The natural Cys^{160} was no longer a ligation junction and it was protected with Acm. Natural Ala in places 122, 142 and 169, selected as new junction points, were mutated to Cys to allow the ligation of the fragments by KCL. The new F4 was shorter, consisting in 20 residues (**F4a** = Cys^{122}-Leu^{141}). The new F5 consisted in 25 residues (**F5a** = Cys ^{142}-Val^{168}). Finally, the last fragment bearing the cholesterol residue (**F6**) is the shortest, from Cys^{169} to Gly^{174} (**Figure *3***).

1.5. Strategy B toward natural Palm-ShhN-cholesterol

The change on the synthetic strategy just affected to the synthesis of the C-terminal block because fragments **F1**, **F2** and **F3** remained the same.

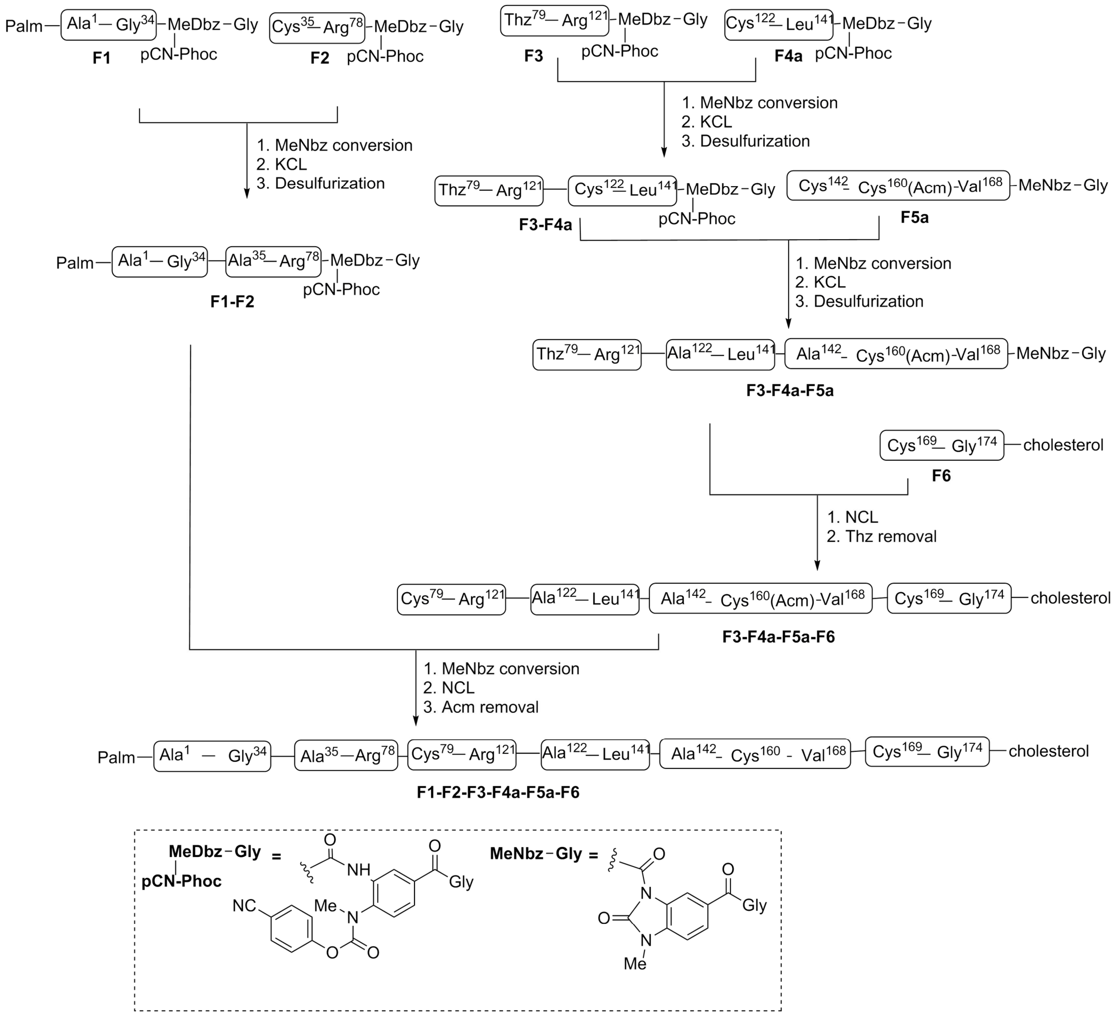

Palm
Ala1— Gly34
MeDbz-Gly
pCN-Phoc
F1
Cys35— Arg78
F2
Thz79— Arg121
F3
Cys122— Leu141
F4a
1. MeNbz conversion
2. KCL
3. Desulfurization
F1-F2
F3-F4a
Cys142— Cys160(Acm)-Val168
MeNbz-Gly
F5a
F3-F4a-F5a
Cys169— Gly174
cholesterol
F6
1. NCL
2. Thz removal
F3-F4a-F5a-F6
1. MeNbz conversion
2. NCL
3. Acm removal
F1-F2-F3-F4a-F5a-F6
MeDbz-Gly =
MeNbz-Gly =
NH
Gly
NC
Me
N
O
Me

Figure 4. Strategy B toward Palm-ShhN-cholesterol.

The new synthetic scheme draws first the ligation of **F3** to the new **F4a** by KCL and the **F5a** peptide with the cholesterol fragment **F6** resulting **F5a-F6**. Then, these two new segments will be condensed to afford **F3-F4a-F5a-F6**. Following ligation, the Cys at 122, 142 and 169 would be reduced to Ala and the Thz^{79} removed to render the native Cys (**Figure 4**).

The last step would consist in the ligation of this new C-terminal fragment with the already prepared N-terminal **F1-F2** sequence.

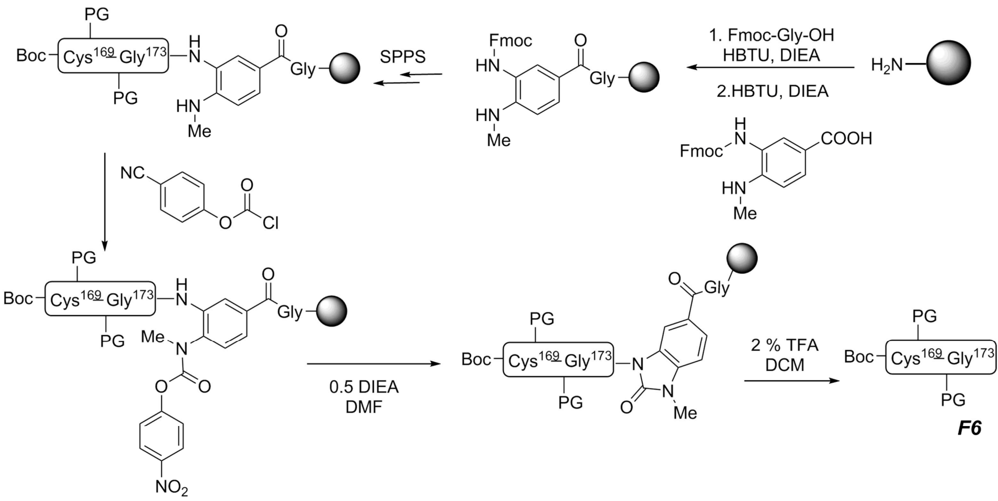
1.6. Synthesis of the peptide fragments

1.6.1. Synthesis of the F6: Cys^{169}-Gly^{174}-cholesterol

In the first place, it was decided to achieve the synthesis of the last **F6** fragment in order to figure out the solubility of the new cholesterol segment.

1.6.1.1. SPPS of the Cys^{169}-Gly^{173}-MeNbz-Gly

The natural Ala^{169} was substituted by a Cys, and introduced protected with Boc. The fragment was synthesized similarly to Cys^{169}-Gly^{173}-MeNbz-Gly (**Scheme *8***).



Scheme 8. Synthesis of F6. PG = protecting group.

1.6.1.2. Assembling of the Gly-cholesterol building block to protected Cys^{169}-Gly^{173}-MeNbz-Gly.

Similarly to the coupling of Gly-cholesterol to Cys^{160}-Gly^{173}-MeNbz-Gly, several reaction conditions were screened. In this case, it was possible to isolate the expected peptide in high purity and enough quantities (**Table *2***).

Temp/time	Solvent	Gly-Cholesterol	DIEA	Coupling agents	Observations
rT / 18 h	THF: DMF (1:1)	5 eq	0.5 M	DMAP 10 eq	SM hydrolyzed + Product
rT / 18 h	THF, DMF (1:1)	5 eq	30 eq	DMAP 10 eq	Product, yield = 33%

Table 2. Reaction conditions to condense Gly-cholesterol with Cys^{169}-Gly^{173}-MeNbz-Gly. d = days, SM = Cys^{169}-Gly^{173}-MeNbz-Gly.

The Cys^{169}-Gly^{174}-cholesterol protected product was obtained after 18 hours of reaction using 5 equivalents of the cholesterol building block, 30 equivalents of DIEA and 10 equivalents of DMAP. After final peptide deprotection, the **F6** fragment was isolated in 33 % yield (**Figure 5**). Given the success on the preparation of this challenging fragment, the synthesis of the **F4a** and **F5a** was next tackled.

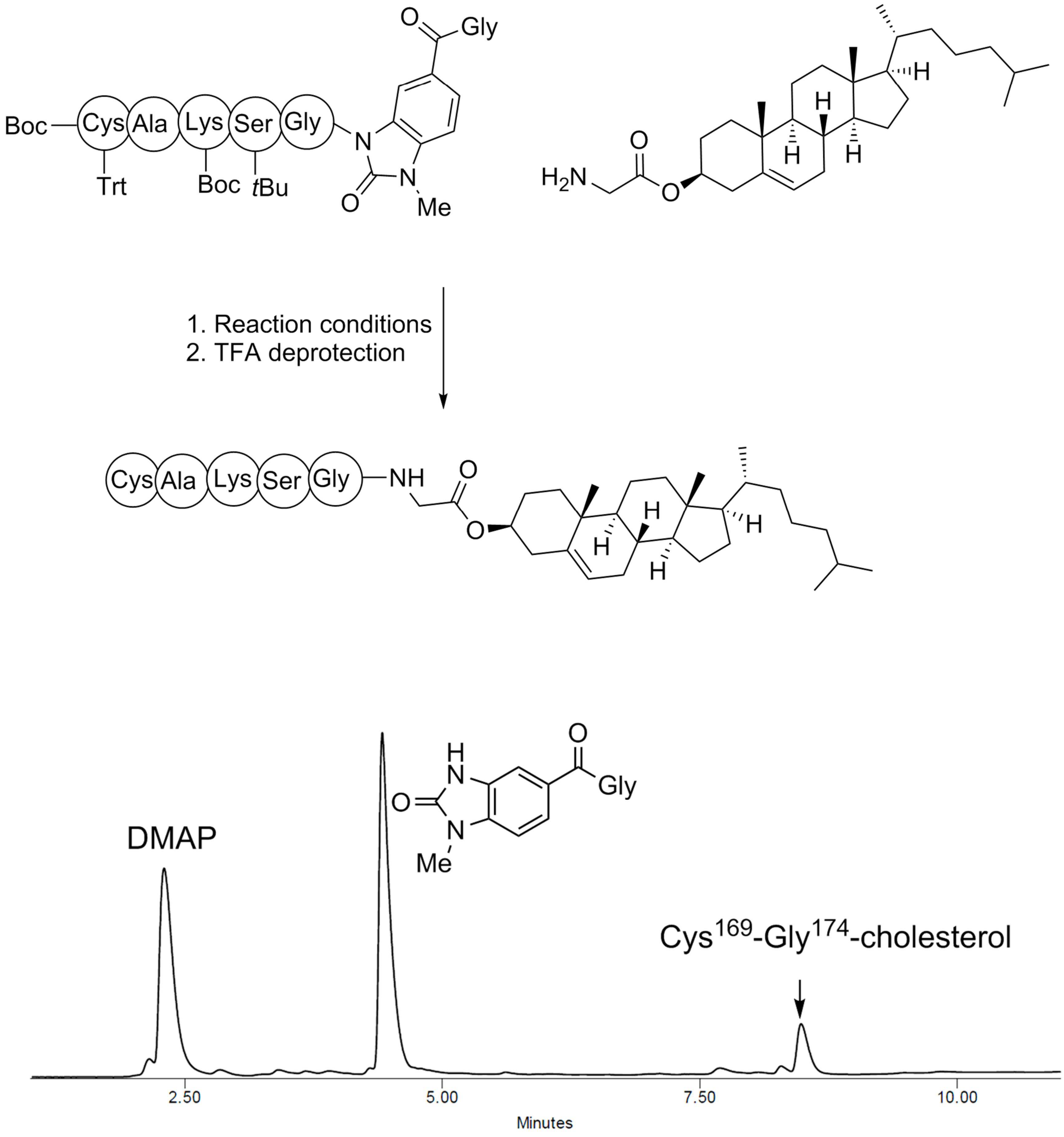
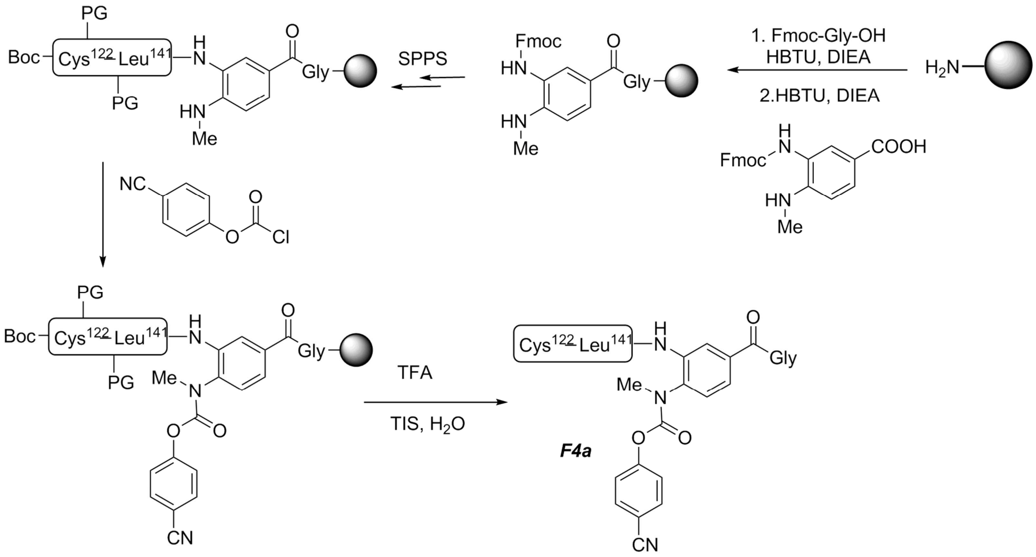

Boc
Cys
Ala
Lys
Ser
Gly
Trt
Boc
tBu
O
Gly
N
N
Me
O
H₂N
O
O
H
H
H
H
1. Reaction conditions
2. TFA deprotection
Cys
Ala
Lys
Ser
Gly
NH
O
O
H
H
H
H
DMAP
H
N
O
N
Me
O
Gly
Cys169-Gly174-cholesterol
2.50
5.00
7.50
10.00
Minutes

Figure 27. A) Top: scheme of the condensation reaction. B) Bottom: HPLC of the reaction.

1.6.2. Synthesis of F4a: Cys^{122}-Leu^{141}-pCN-Phoc-MeDbz-Gly

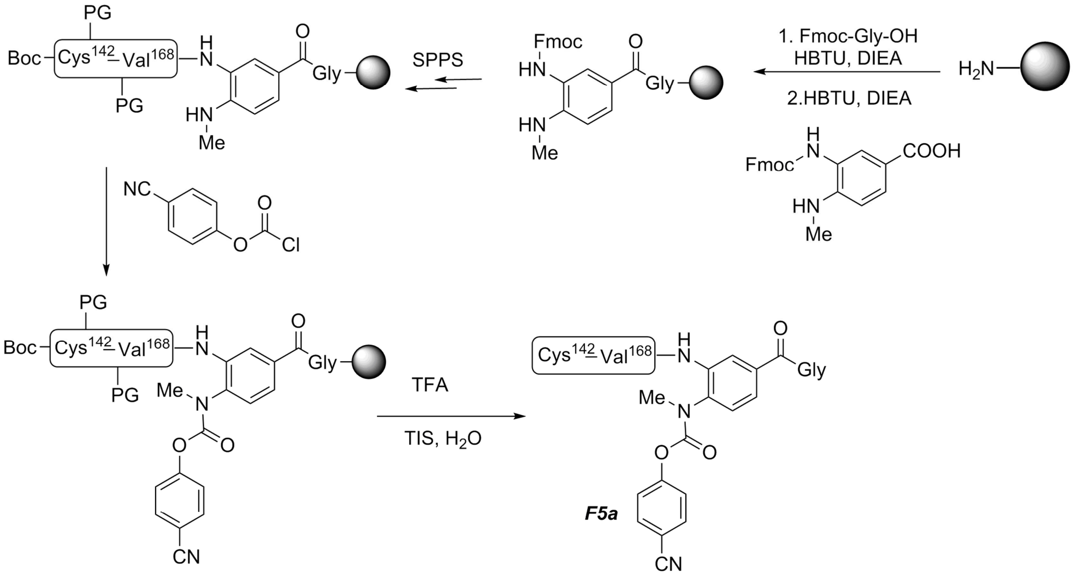
This fragment was synthetized following the standard procedure described for the synthesis of previous fragments (**Scheme 9**).



Scheme 9. Synthesis of F4a. PG = protecting group.

1.6.3. *Synthesis of F5a: Cys^{142}-Val^{168}-pCN-Phoc-MeDbz*

The peptide was synthetized following the same procedure as described above for the synthesis of **F5a** (**Scheme *10***).



Scheme 10. Synthesis of F5a. PG = protecting group.

1.7. KCL and NCL of the peptide fragments

With the fragments already in hand, the next step were the assembling of the different parts. The N-terminal half was built from the condensation of F1 and F2. On the other hand, the synthesis of the C-terminal part included the ligation of fragments F3, F4a, F5a, and F6.

The final ligation between the C and the N-terminus fragments would render the full Palm-ShhN-cholesterol protein.

1.7.1. *Synthesis of the N-terminal block: Palm-Ala^{1}-Arg^{78}-pCN-Phoc-MeDbz*

1.7.1.1. *KCL between F1 and F2*

Before the KCL between F1 and F2, the F1-*p*CN-Phoc-MeDbz-Gly peptide was thioesterified in the presence of MESNa (100 mM) at pH 7.5 and purified. The ligation proceeded upon the addition of F2 on buffer C containing MPOH (100 mM) at pH 6.8-7.0 (**Figure *6***). The ligation was completed after 1 hour 30 min affording the desired product in 72 % yield. It is important to remark that it was not observed solubility problems due to the palmitic acid.

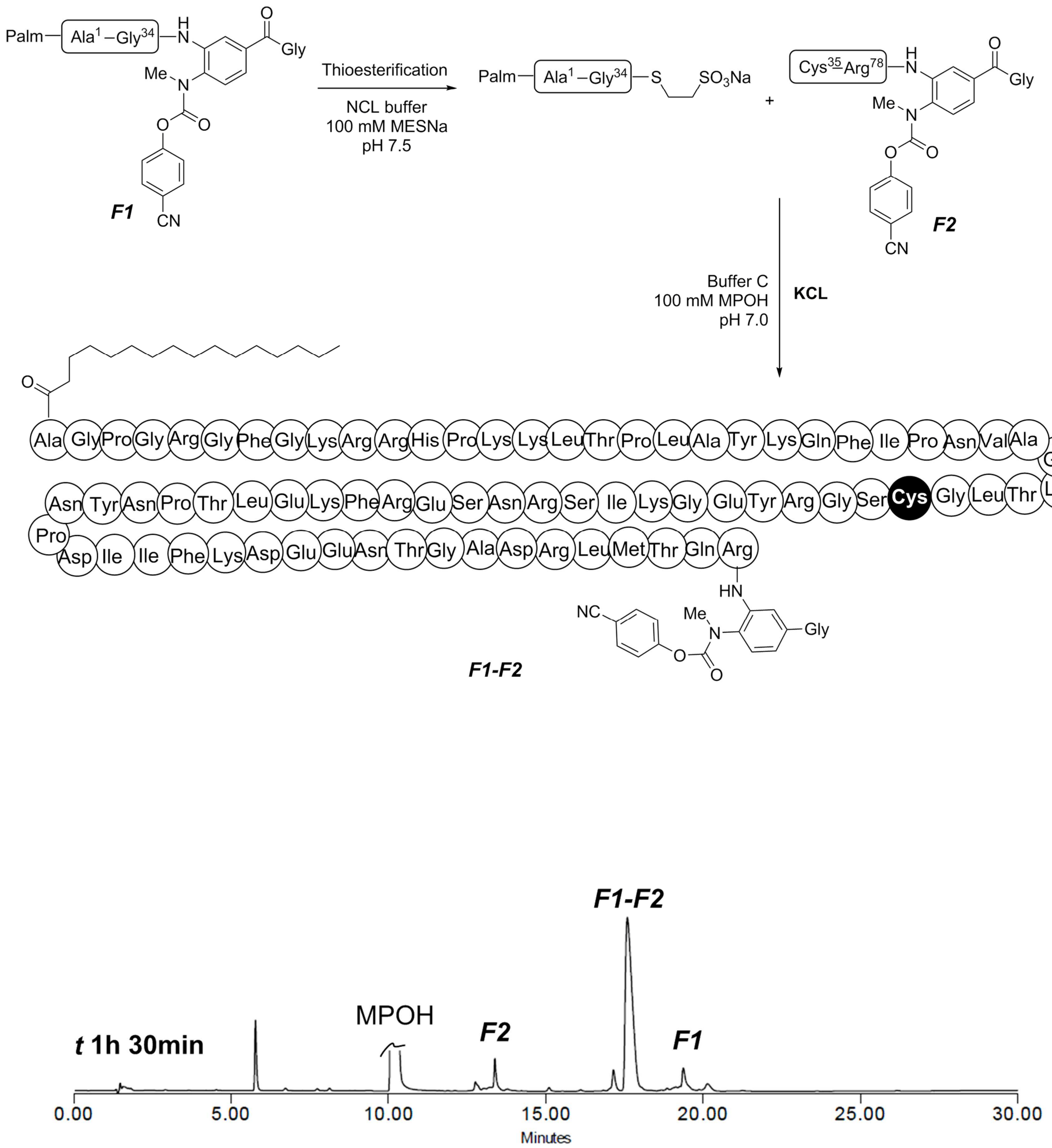

Palm
Ala^{1}–Gly^{34}
Gly
Me
CN
F1
Thioesterification
NCL buffer
100 mM MESNa
pH 7.5
Palm
Ala^{1}–Gly^{34}
S
SO$_3$Na
+
Cys^{35}–Arg^{78}
F2
Buffer C
100 mM MPOH
pH 7.0
KCL
Ala Gly Pro Gly Arg Gly Phe Gly Lys Arg Arg His Pro Lys Lys Leu Thr Pro Leu Ala Tyr Lys Gln Phe Ile Pro Asn Val Ala Glu Lys Thr Leu Gly Cys Ser Gly Arg Tyr Glu Gly Lys Ile Ser Arg Asn Ser Glu Arg Phe Lys Glu Leu Thr Pro Asn Tyr Asn Pro Asp Ile Ile Phe Lys Asp Glu Glu Asn Thr Gly Ala Asp Arg Leu Met Thr Gln Arg
HN
NC
Me
N
Gly
F1-F2
F1-F2
MPOH
F2
F1
t 1h 30min
0.00
5.00
10.00
15.00
20.00
25.00
30.00
Minutes

Figure 6. Top: scheme of KCL between F1 and F2Bottom: HPLC of the reaction.

1.7.1.2. Desulfurization of F1-F2

After purification of the ligated F1-F2 fragment, the natural Cys^{35} was reduced to the native Ala. The product was dissolved in buffer D (6 M guanidinium chloride, 0.2 M phosphate, 250 mM MESNa, 250 mM TCEP and 100 mM radical initiator VA-044) and the pH was adjusted to 6.5 with triethylamine, leaving the reaction at 37 °C for 5 hours (**Figure *7***). Once the desulfurization was completed, the peptide was obtained in 60 % yield after purification.

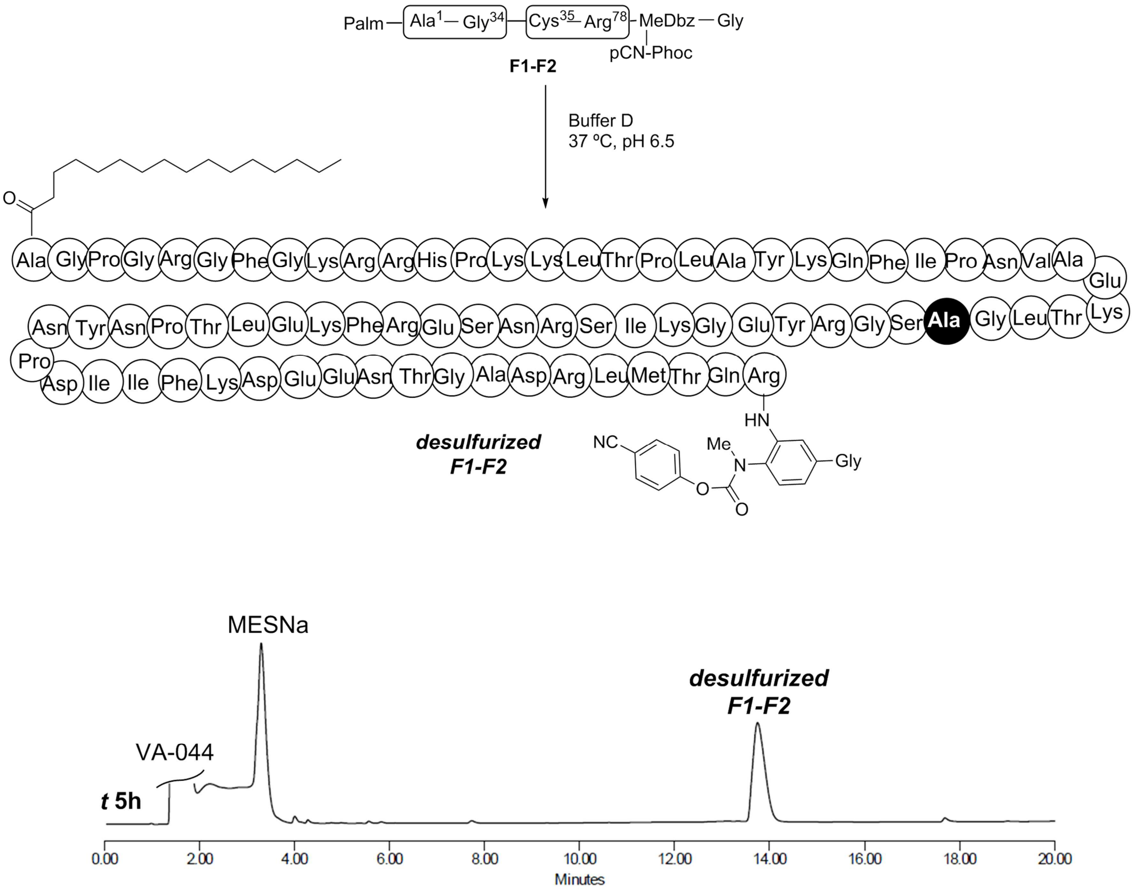

Palm
Ala1 — Gly34
Cys35 — Arg78
MeDbz — Gly
pCN-Phoc
F1-F2
Buffer D
37 °C, pH 6.5
Ala
desulfurized F1-F2
NC
Me
N
O
HN
Gly
MESNa
VA-044
t 5h
desulfurized F1-F2
0.00
2.00
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00
Minutes

Figure 7. Top: scheme of the desulfurization reaction. Bottom: HPLC of the reaction.

1.7.2. Synthesis of the C-terminal block: Cys^{79}-Gly^{174}-cholesterol

1.7.2.1. KCL between F3 and F4a

Fragment **F3** was first converted in a MESNa thioester using the same conditions employed for F1. Then, ligation with **F4a** proceeded in buffer C containing MPOH (100 mM) at pH 6.3. It was observed during this ligation significant amounts of **F3** hydrolyzed (**F3-OH**). After some troubleshooting (different thiols, pH and co-solvents), the optimized reaction was carried out in buffer C with MPOH (100 mM) as catalyst at pH 6.5 using a slight excess of **F4a** (1.4 equiv).

First, the **F4a** is dissolved in the ligation buffer, and then is added to the lyophilized **F3-MESNa** peptide. Using this procedure it was possible to decrease the percentage of hydrolysis to lower than 7 %, and obtain the **F3-F4a** product in a decent 51% yield (**Figure *8***).

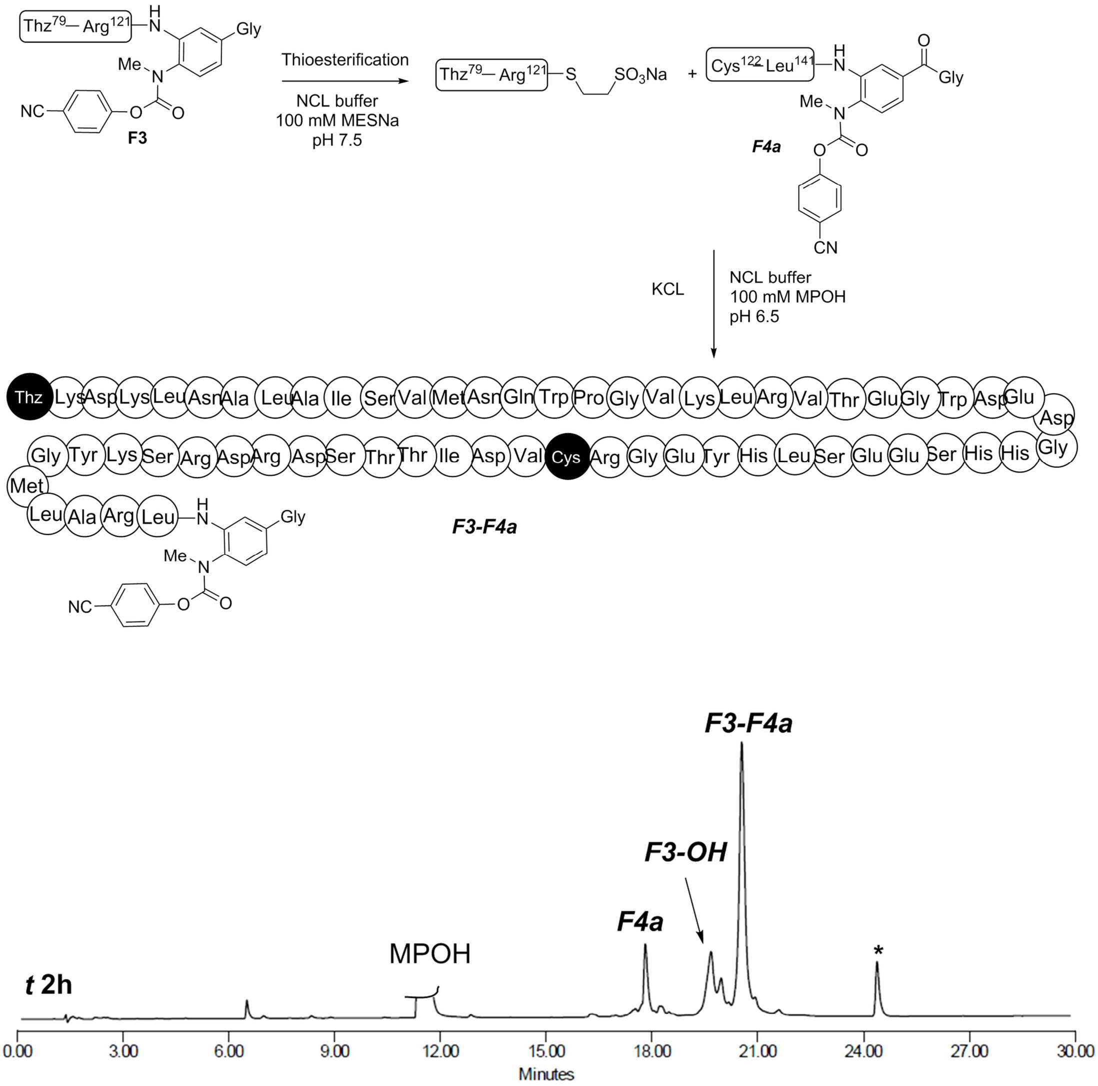

Thz79— Arg121
Gly
Me
NC
F3
Thioesterification
NCL buffer
100 mM MESNa
pH 7.5
Thz79— Arg121
SO3Na
+
Cys122 Leu141
Gly
F4a
CN
KCL
NCL buffer
100 mM MPOH
pH 6.5
Thz Lys Asp Lys Leu Asn Ala Leu Ala Ile Ser Val Met Asn Gln Trp Pro Gly Val Lys Leu Arg Val Thr Glu Gly Trp Asp Glu Asp Gly His His Ser Glu Glu Ser Leu His Tyr Glu Gly Arg Cys Val Asp Ile Thr Thr Ser Asp Arg Asp Arg Ser Lys Tyr Gly Met Leu Ala Arg Leu
F3-F4a
F3-F4a
F3-OH
F4a
MPOH
*
t 2h
0.00
3.00
6.00
9.00
12.00
15.00
18.00
21.00
24.00
27.00
30.00
Minutes

Figure 8. Top: Scheme of the KCL between F3 and F4a. (*) impurity from MPOH.

1.7.2.2. KCL between F3-F4a-pCN-Phoc-MeDbz-Gly and F5a

Previous to the ligation to F5b, the ligated peptide F3-F4b was converted into MeNbz to allow the thioesterification on its C-terminal. Then, after addition of F5a, the reaction became cloudy showing the appearance of a precipitate. The reaction did not give the desired KCL product, just the intramolecular cyclization of F5a (**Figure *9***).

In order to overcome this drawback, a third strategy involving a convergent condensation of **F3-F4a** and **F5a-F6** was attempted.

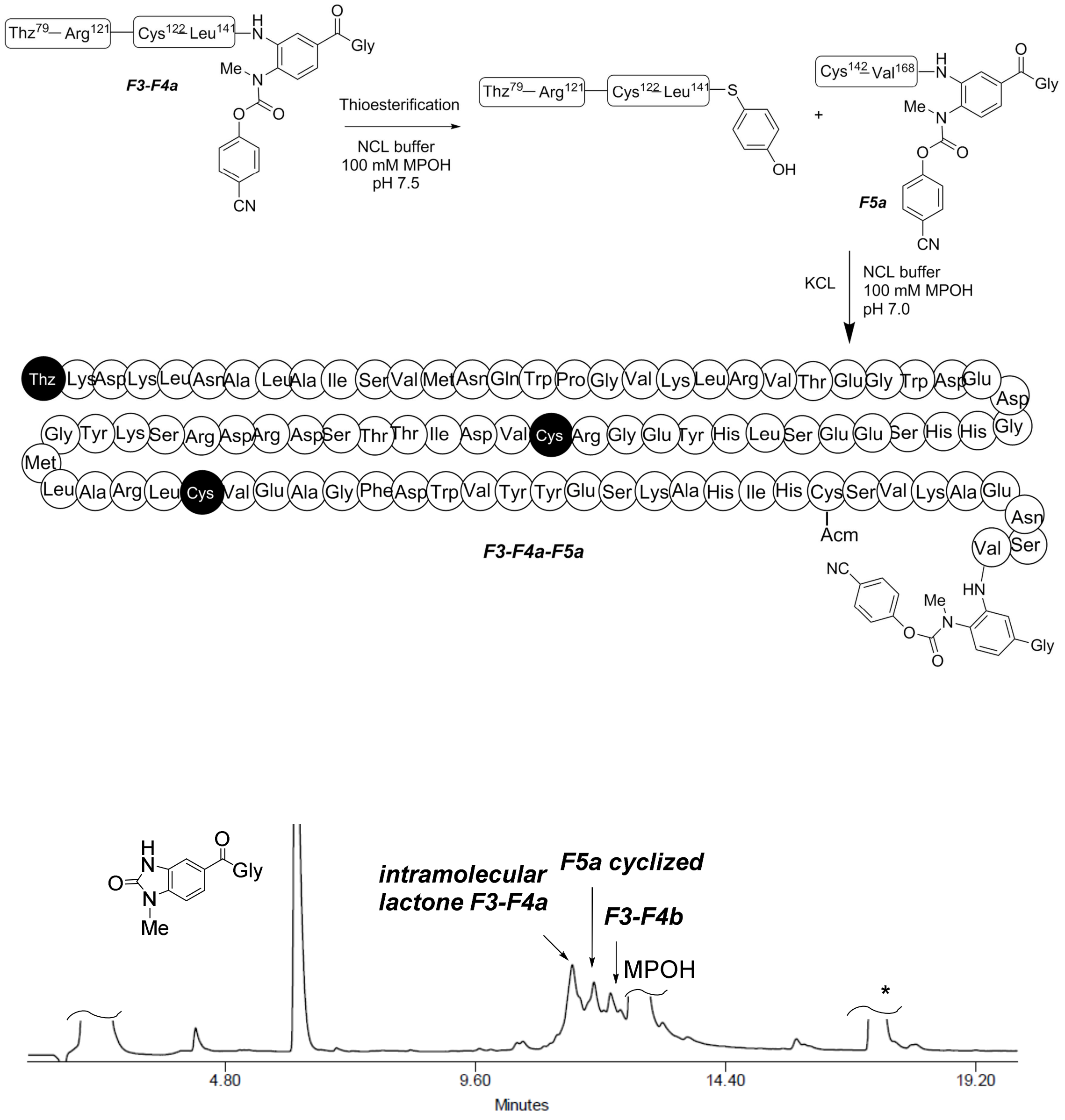

Thz79— Arg121
Cys122 Leu141
Gly
Me
F3-F4a
CN
Thioesterification
NCL buffer
100 mM MPOH
pH 7.5
Thz79— Arg121
Cys122 Leu141
S
OH
+
Cys142 Val168
Gly
F5a
KCL
NCL buffer
100 mM MPOH
pH 7.0
Thz Lys Asp Lys Leu Asn Ala Leu Ala Ile Ser Val Met Asn Gln Trp Pro Gly Val Lys Leu Arg Val Thr Glu Gly Trp Asp Glu Asp
Gly Tyr Lys Ser Arg Asp Arg Asp Ser Thr Thr Ile Asp Val Cys Arg Gly Glu Tyr His Leu Ser Glu Glu Ser His His Gly
Met
Leu Ala Arg Leu Cys Val Glu Ala Gly Phe Asp Trp Val Tyr Tyr Glu Ser Lys Ala His Ile His Cys Ser Val Lys Ala Glu Asn Ser Val
Acm
NC
HN
F3-F4a-F5a
intramolecular lactone F3-F4a
F5a cyclized
F3-F4b
MPOH
*
4.80
9.60
14.40
19.20
Minutes

Figure 9. Top: Scheme of the KCL between F3-F4a and F5a. Bottom: HPLC of the reaction. (*) thiol oxidized.

1.8. Strategy C towards the synthesis of Palm-ShhN-cholesterol

In this new synthetic strategy the N-terminal Cys of **F5a** was changed by a Thz (**F5b**). Thus on one side **F3** and **F4a** would be first ligated, and on the other side **F5b** and F6 would also be assembled. Then, the KCL products would finally be assembled giving the C-terminal block **F3-F4a-F5b-F6** (**Figure *10***).

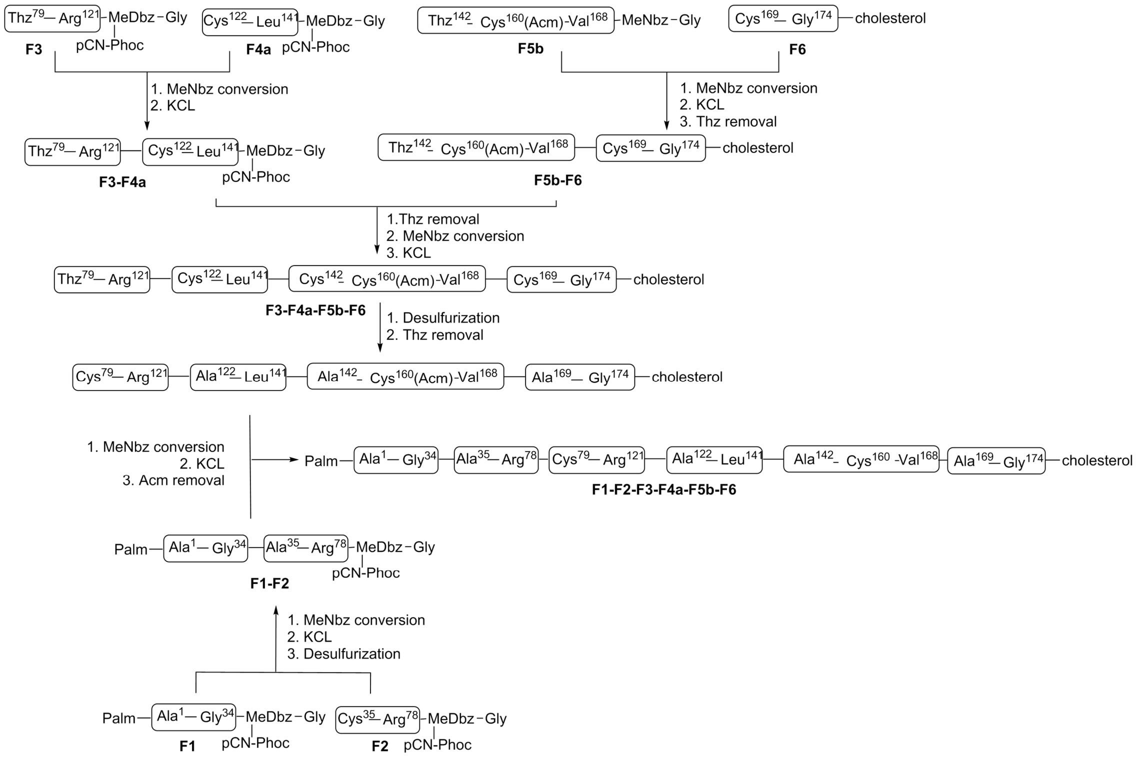
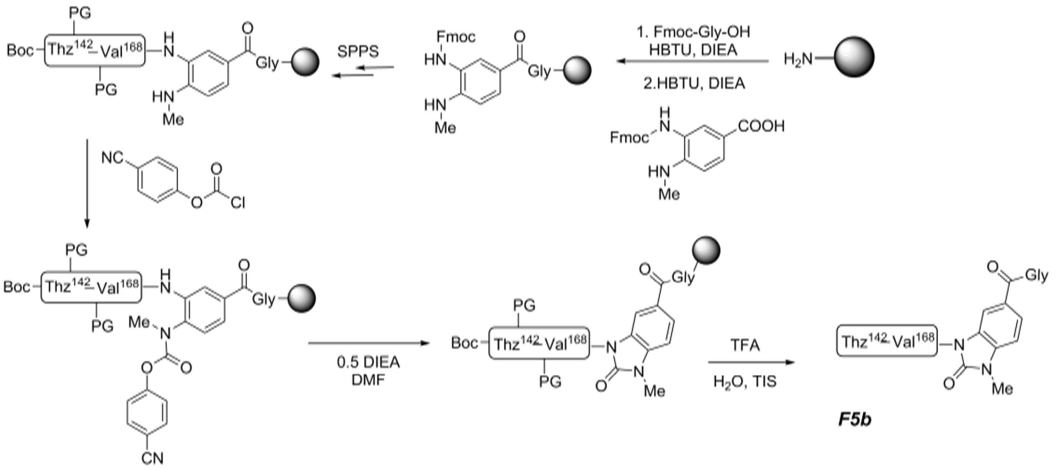

Thz79—Arg121–MeDbz-Gly
pCN-Phoc
F3
Cys122—Leu141–MeDbz-Gly
F4a
pCN-Phoc
Thz142—Cys160(Acm)-Val168–MeNbz-Gly
F5b
Cys169—Gly174—cholesterol
F6
1. MeNbz conversion
2. KCL
1. MeNbz conversion
2. KCL
3. Thz removal
F3-F4a
F5b-F6
1.Thz removal
2. MeNbz conversion
3. KCL
F3-F4a-F5b-F6
1. Desulfurization
2. Thz removal
1. MeNbz conversion
2. KCL
3. Acm removal
Palm—Ala1—Gly34—Ala35—Arg78—Cys79—Arg121—Ala122—Leu141—Ala142—Cys160-Val168—Ala169—Gly174—cholesterol
F1-F2-F3-F4a-F5b-F6
Palm—Ala1—Gly34—Ala35—Arg78–MeDbz-Gly
F1-F2
1. MeNbz conversion
2. KCL
3. Desulfurization
Palm—Ala1—Gly34–MeDbz-Gly
F1
Cys35—Arg78–MeDbz-Gly
F2

Figure10. Strategy C towards the natural Palm-ShhN-Cholesterol**.**

1.9. Synthesis of the peptide fragments

1.9.1. Synthesis of the F5b: Thz^{142}-Val^{168}-pCN-Phoc-MeNbz

The new **F5b** was synthetized following standard SPPS procedures, as previously reported **F5a**. The only difference was the introduction of N-terminal Thz (**Scheme *11***).

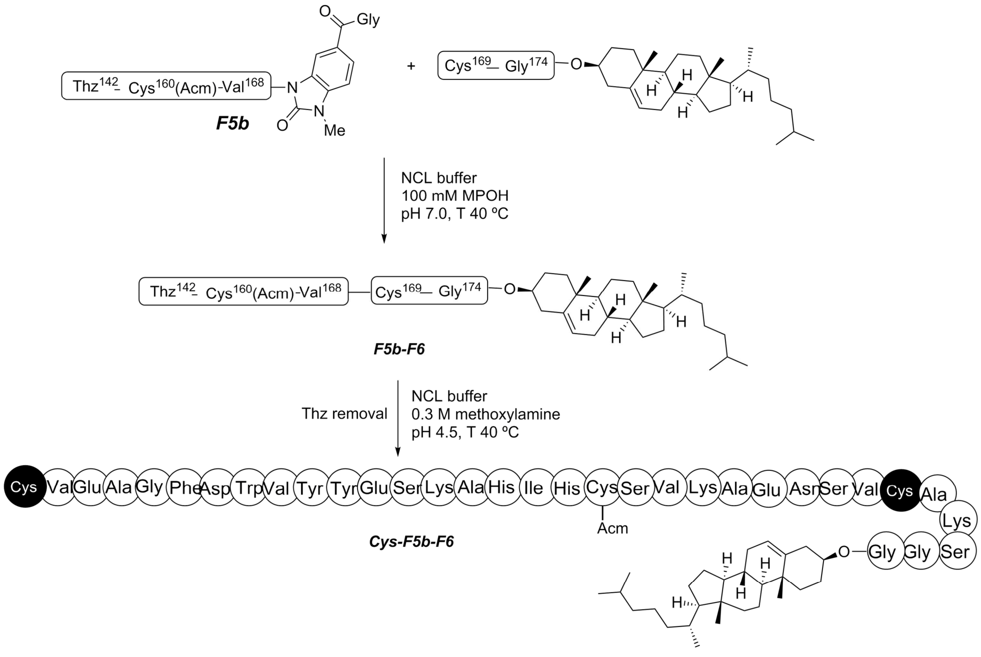

PG
Boc
Thz142 - Val168
SPPS
Fmoc
Gly
1. Fmoc-Gly-OH
HBTU, DIEA
2.HBTU, DIEA
COOH
HN
Me
NC
Cl
0.5 DIEA
DMF
TFA
H2O, TIS
F5b

Scheme 11. Synthesis of F5b. PG = protecting group.

1.9.2. Synthesis of the C-terminal block, strategy C

1.9.2.1. Ligation between F5b and F6 and methoxylamine removal

Both the **F5b** and **F6** fragments were dissolved in buffer C containing 100 mM MPOH at 40 °C and the pH was adjusted to 7.5 (**Scheme 12**). The ligated product was obtained after 18 hours of reaction (**Figure *11***). Since purification of the product was difficult due to its high hydrophobicity, it was decided to carry out the thiazolidine removal in the same pot. To this end, after extraction of the MPOH, methoxylamine and TCEP were added and the pH adjusted to 4.2. The desired Cys peptide was observed after 18 hours. After several attempts to purify the product, best results (62 % yield) were obtained by precipitation (**Figure 12**).


Thz142 - Cys160(Acm)-Val168
F5b
Cys169 - Gly174
NCL buffer
100 mM MPOH
pH 7.0, T 40 ºC
F5b-F6
Thz removal
NCL buffer
0.3 M methoxylamine
pH 4.5, T 40 ºC
Cys Val Glu Ala Gly Phe Asp Trp Val Tyr Tyr Glu Ser Lys Ala His Ile His Cys Ser Val Lys Ala Glu Asn Ser Val Cys Ala Lys Ser Gly Gly
Acm
Cys-F5b-F6

Scheme 12. KCL between F5b and F6, and following Thz removal of the F5b-F6 fragment.

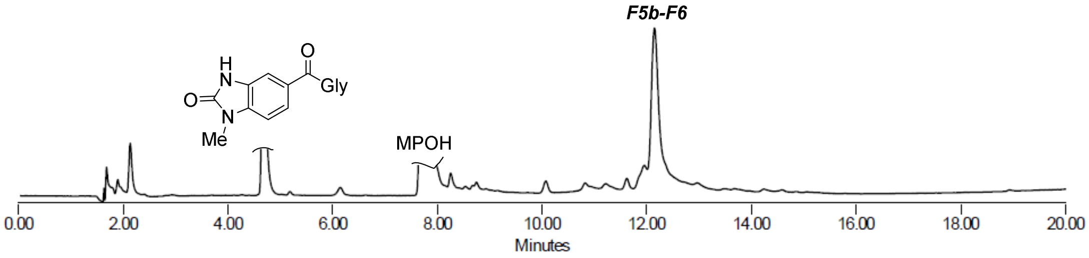


Figure 11. HPLC of the KCL between F5c and F6.

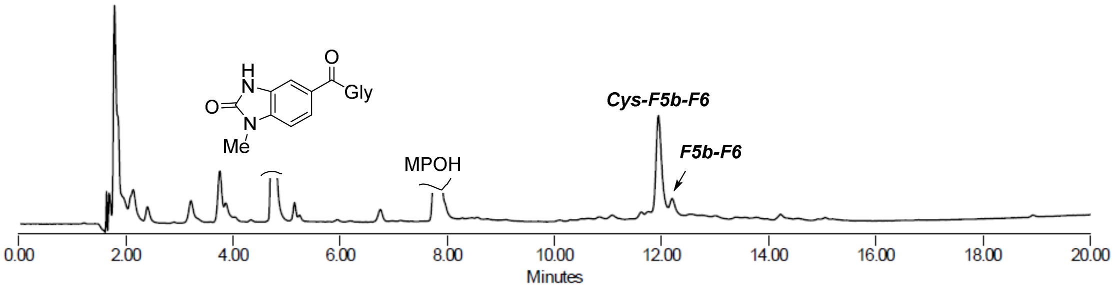


Figure 12. HPLC of the Thz removal of the F5c-F6 fragment.

1.9.2.2. Ligation between F3-F4a and F5a-F6

To obtain the ligated C-terminal block, F3-F4a and F5a-F6 peptides were dissolved in Buffer C containing 100 mM MPOH and pH was adjusted to 7.0. Unfortunately, a precipitate was observed and the desired product was not detected. Because of the identification of side products coming from the hydrolysis of the F3-F4a fragment or its intramolecular cyclization, we hypothesized that the fragment bearing the cholesterol was not soluble, and thus could not react to give the desired product.

Although this strategy was abandoned, the use of detergents during the ligation seemed promising to solubilize the lipid peptide fragment and avoid precipitation. Also the use of small ammounts of organic solvents, such as DMF could be an alternative.

2. Synthesis of the Palm-ShhN-biotin

It has been shown in different studies,[28] the activity of ShhN relies on its N-terminal region, whereas the C-terminal part, that includes the cholesterol, has mainly a protein translocation function. The removal of the cholesterol does not affect to the binding to Patched and posterior activation of the pathway. Therefore, it is possible to get rid of the cholesterol moiety, which would also avoid solubility problems. A biotin residue attached to a Lys side chain through a PEG linker would replace the cholesterol. The incorporation of a biotin can be useful in some biological studies in where it is interesting to attach the protein on a surface. Therefore, it was determined to prepare a synthetic analog of ShhN that would have a palmitoyl at the N-terminus and a pegylated biotin attached at the C-terminus.

2.1. Restrosynthesis of the synthetic analog Palm-Shh-biotin

In order to take advantage of the synthesized fragments the protein was split in five fragments. The first four of them were already prepared (**F1**, **F2**, **F3** and **F4a**), whereas **F5c** would span Ala^{142}-Gly^{174}. A Cys residue will substitute the native Ala^{142} to allow the ligation. Moreover, an extra Lys will be added at the C-terminus to attach the pegylated biotin in its side chain.

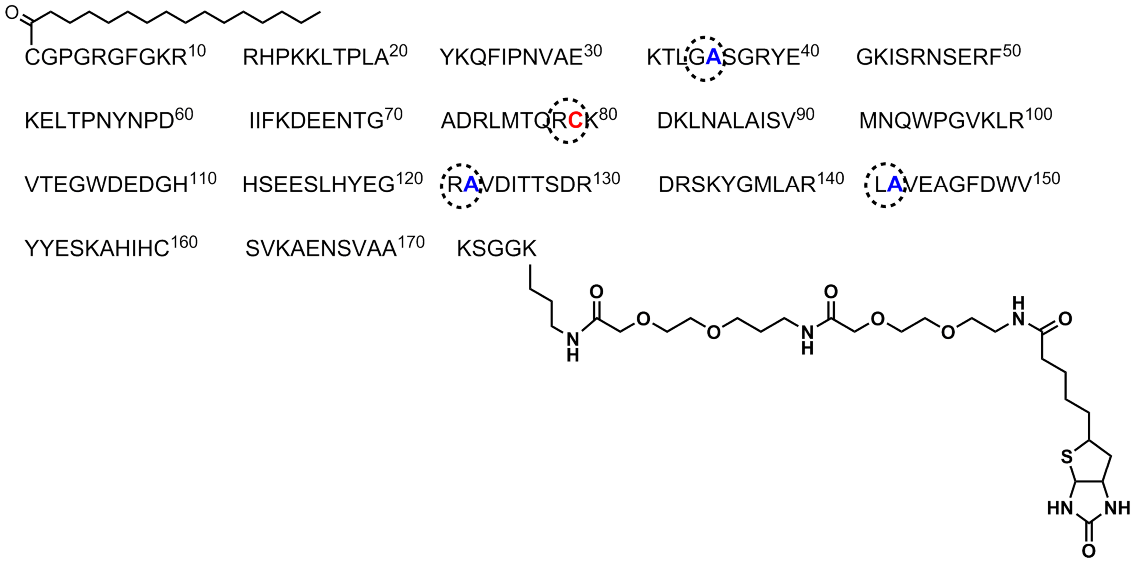

O
CGPGRGFGKR10
RHPKKLTPLA20
YKQFIPNVAE30
KTLGASGRYE40
GKISRNSERF50
KELTPNYNPD60
IIFKDEENTG70
ADRLMTQRCK80
DKLNALAISV90
MNQWPGVKLR100
VTEGWDEDGH110
HSEESLHYEG120
RAVDITTSDR130
DRSKYGMLAR140
LAVEAGFDWV150
YYESKAHIHC160
SVKAENSVAA170
KSGGK
O
O
O
N
H
O
O
O
N
H
H
N
O
S
HN
NH
O

Figure 13. Retrosynthetic analysis of the Palm-ShhN-biotin analog. Circles and letters in colors indicate ligation sites. In blue, the Ala residues which would be mutated into Cys for ligation purposes. In red, natural Cys used as ligation points.

2.2. Synthetic strategy towards the synthetic analog Palm-ShhN-biotin

Like in the previous strategies, the N-terminal half (**F1-F2**) would be ligated in the final step to the C-terminal part (**F3-F4a-F5c**, (**Figure 14**). Thus, fragments **F1-F2** and **F3-F4a** were already prepared from other approaches and the only difference was the ligation of **F5c** to **F3-F4a** for the synthesis of the C-terminal block. Then, following desulfurization to restore the natural

Ala^{122} and Ala^{142}, the Thz^{79} would be removed to give the Cys^{79}. The next step would consist on the ligation between both fragments and the final Acm removal to render the native Cys^{160}.

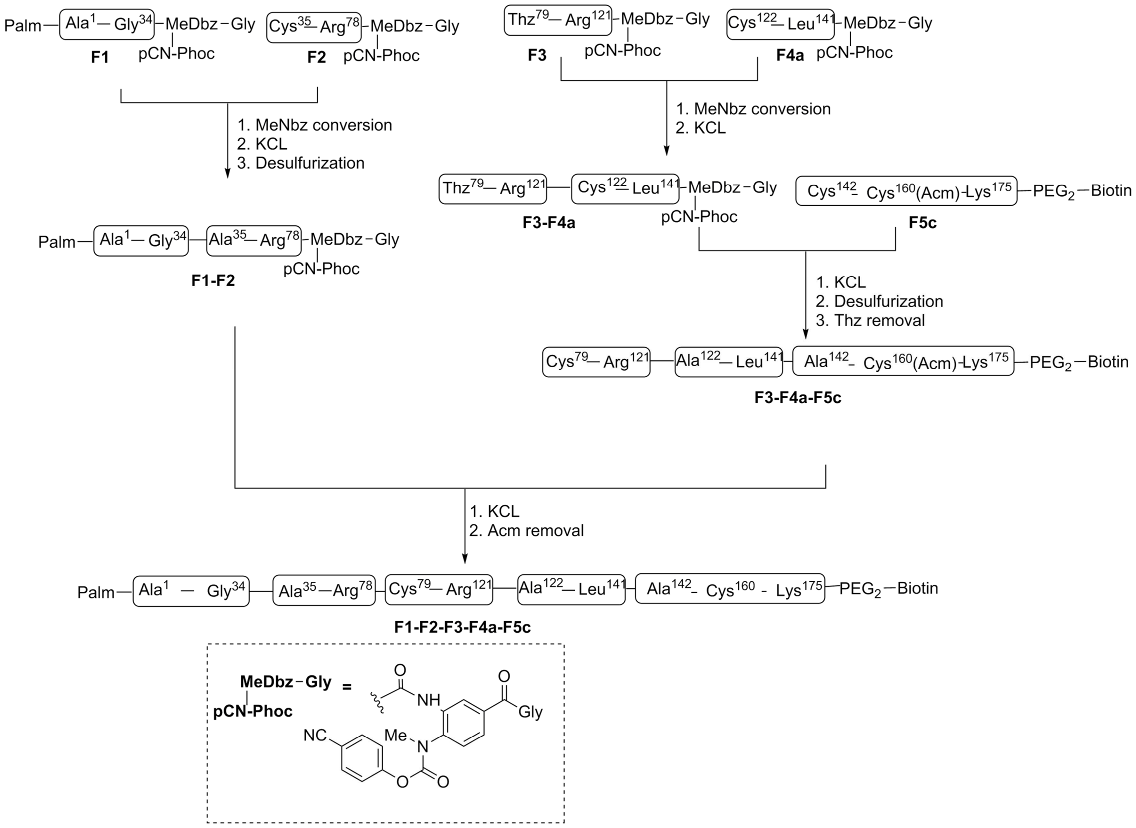
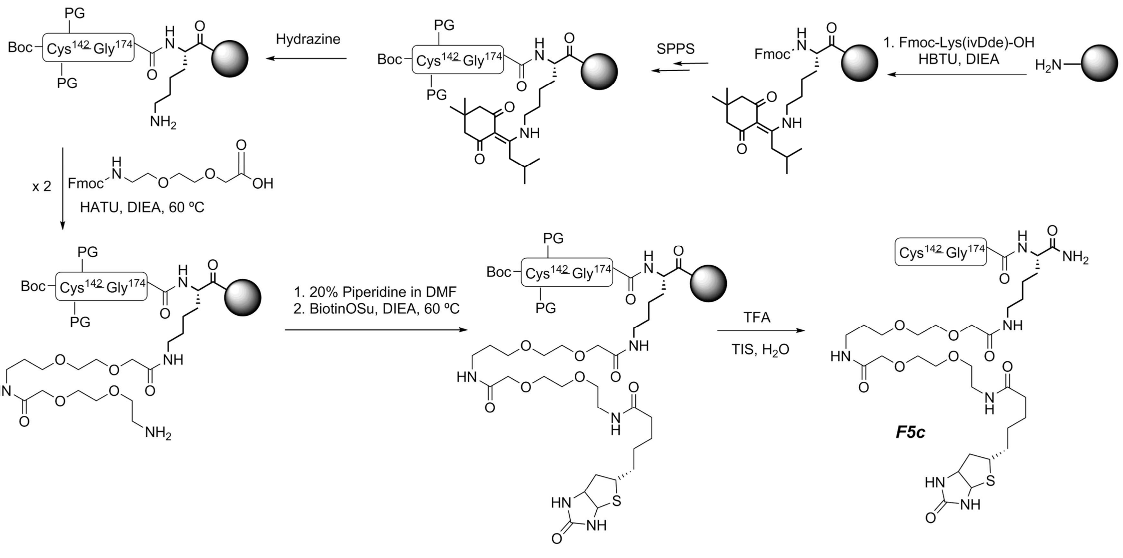

Palm—Ala¹—Gly³⁴—MeDbz-Gly / pCN-Phoc
F1
Cys³⁵—Arg⁷⁸—MeDbz-Gly / pCN-Phoc
F2
Thz⁷⁹—Arg¹²¹—MeDbz-Gly / pCN-Phoc
F3
Cys¹²²—Leu¹⁴¹—MeDbz-Gly / pCN-Phoc
F4a
1. MeNbz conversion
2. KCL
3. Desulfurization
1. MeNbz conversion
2. KCL
Thz⁷⁹—Arg¹²¹—Cys¹²²—Leu¹⁴¹—MeDbz-Gly / pCN-Phoc
F3-F4a
Cys¹⁴²—Cys¹⁶⁰(Acm)-Lys¹⁷⁵—PEG₂—Biotin
F5c
Palm—Ala¹—Gly³⁴—Ala³⁵—Arg⁷⁸—MeDbz-Gly / pCN-Phoc
F1-F2
1. KCL
2. Desulfurization
3. Thz removal
Cys⁷⁹—Arg¹²¹—Ala¹²²—Leu¹⁴¹—Ala¹⁴²—Cys¹⁶⁰(Acm)-Lys¹⁷⁵—PEG₂—Biotin
F3-F4a-F5c
1. KCL
2. Acm removal
Palm—Ala¹—Gly³⁴—Ala³⁵—Arg⁷⁸—Cys⁷⁹—Arg¹²¹—Ala¹²²—Leu¹⁴¹—Ala¹⁴²—Cys¹⁶⁰-Lys¹⁷⁵—PEG₂—Biotin
F1-F2-F3-F4a-F5c
MeDbz-Gly / pCN-Phoc =
O
NH
O
Gly
NC
Me
N
O
O

Figure 14. Synthetic strategy towards Palm-ShhN-biotin.

2.3. Synthesis of the peptide fragments

2.3.1. Synthesis of F5c: Cys^{142}-Lys^{175}-PEG$_2$-Biotin

The new F5c, a 33-mer peptide, has a similar length to the other fragments. The natural Ala^{142} was changed into a Cys for ligation purposes, while Cys^{160} was introduced protected with Acm on its side chain to avoid desulfurization. An extra Lys was added at the C-terminus to attach the pegylated biotin. Thereby, the Lys^{175} was introduced protected with ivDde on its side chain. After peptide elongation, the ivDde protecting group was removed using hydrazine and the pegylated biotin residue was attached (**Scheme *13***).


PG
Boc
Cys142 Gly174
Hydrazine
SPPS
1. Fmoc-Lys(ivDde)-OH
HBTU, DIEA
Fmoc
NH2
x 2
HATU, DIEA, 60 ºC
1. 20% Piperidine in DMF
2. BiotinOSu, DIEA, 60 ºC
TFA
TIS, H2O
F5c

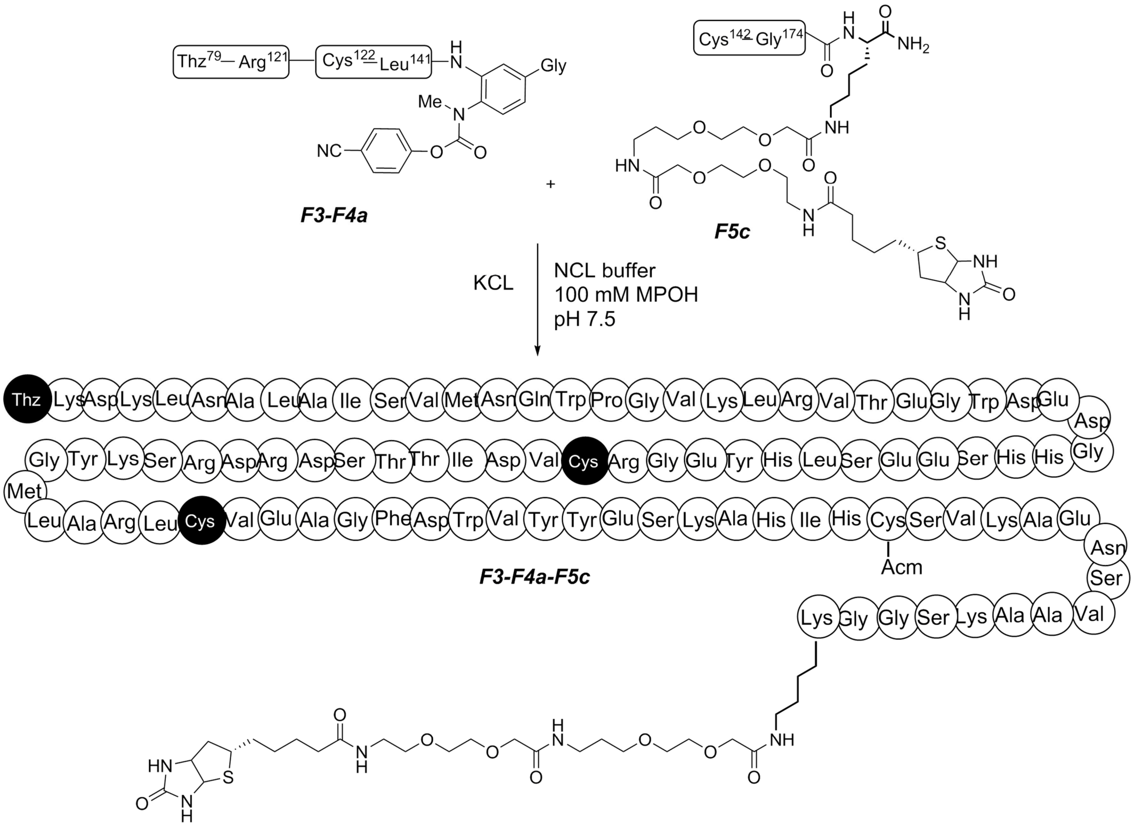
Scheme 13. Synthesis of the F5c. PG = Protecting group.

2.4.. Synthesis of the C-terminal block: Cys^{79}-Lys^{175}-PEG$_2$-biotin

The ligation between the **F3** and **F4a** was carried out following the same procedure reported for the last synthetic approach (Section 1.6.2.1). This section will be focused on the ligations between the **F3-F4a** and **F5c**, and the following reactions to finally afford the fully Palm-ShhN-biotin protein.

2.4.1. KCL between F3-F4a-pCN-Phoc-MeDbz-Gly and F5c

The ligated product **F3-F4a** and **F5c** were dissolved in NCL buffer containing 100 mM MPOH and the pH was raised to pH 7.5 (**Scheme 1*4***). The ligation was left at room temperature obtaining the desired product in 18 hours, in 56 % yield (**Figure *15***).


Thz79— Arg121
Cys122—Leu141
Gly
Me
NC
F3-F4a
+
Cys142—Gly174
NH2
F5c
KCL
NCL buffer
100 mM MPOH
pH 7.5
Acm
F3-F4a-F5c

Scheme 14. KCL between F3-F4a and F5c.

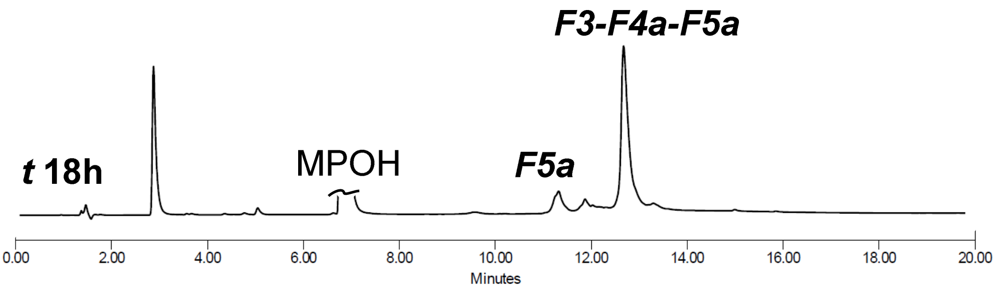
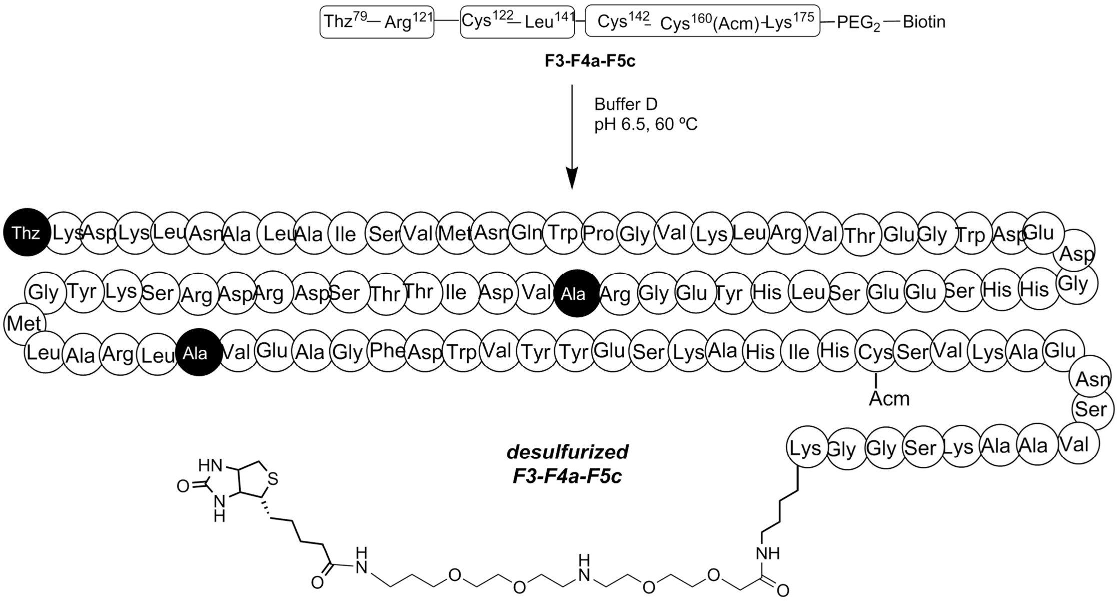


Figure 15. HPLC trace of the NCL between F3-F4a and F5c.

2.4.2. Desulfuritzation of the F3-F4-F5c

After ligation, the product **F3-F4a-F5c** was desulfurized to afford the native Ala^{122} and Ala^{142} (**Scheme *15***). The conditions that were used for the desulfurization of the other fragments (buffer D at 37 °C) did not render the completed desulfurized peptide even after prolonged reaction times. Instead, a mixture of Ala and Cys peptides in different ratios was observed. It was hypothesized that the desulfurization of two Cys would require higher temperature.

Luckily, when carried out at 60 °C, the two Cys were completely reduced to Ala. However, it was also detected that the Thz is partially removed at this temperature and the resulting Cys desulfurized, as well, especially in reaction times longer than 45 min. This unwanted reaction rendered a peptide no longer able to ligate with the N-terminal block. Fortunately, when dissolving the peptide in desulfurization buffer at pH 6.5, 60 °C for 45 minutes, the Thz elimination could be reduced to a minimum levels (less than 10 %) we obtained the desired desulfurized product with minimum undesired thiazolidine removal in 66 % yield (**Figure *16***).


Thz79— Arg121 — Cys122— Leu141 — Cys142 Cys160(Acm)-Lys175 — PEG2 — Biotin
F3-F4a-F5c
Buffer D
pH 6.5, 60 °C
Acm
desulfurized
F3-F4a-F5c

Scheme 15. Desulfurization of F3-F4a-F5c fragment.

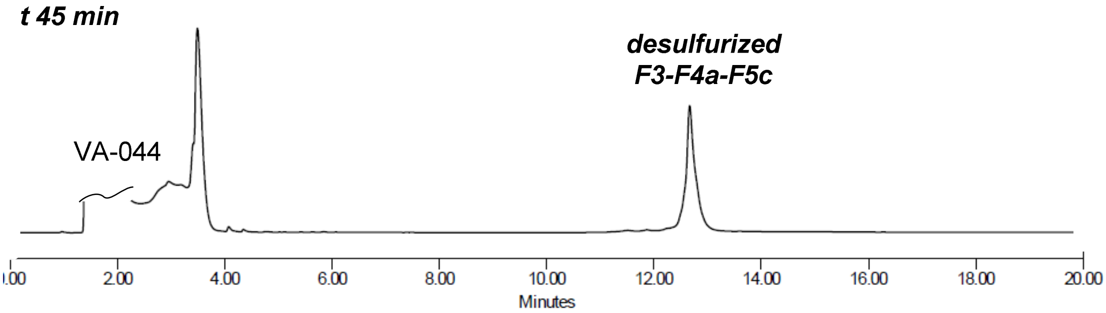


Figure 16. HPLC of the desulfurization reaction of F3-F4b-F5d fragment.

2.4.2.1.. Alternative using F3b with Cys(Acm) at the N-terminus: Cys(Acm)^{79}-Arg^{121}-pCN-Phoc-MeDbz

As an alternative to avoid undesired partially thiazolidine removal, we thought of the synthesis of a modified F3 fragment in which the Thz^{79}would be substituted for a Cys(Acm) (**Figure *17***). In addition to overcome problems during desulfurization, this strategy also avoid an extra step to remove the Thz. To confirm the feasibility of this new strategy, the new F3a was synthetized and used for the synthesis of the C-terminal block following the same strategy described above.

Thereby, F3a was ligated to F4b, and the ligation product was obtained in 51 % yield. Next, it the F3a-F4a was ligated to F5c rendering the desired product 43 % yield which was desulfurized restoring the natural Ala at position 123 and 143. Finally, the removal of the Acm protecting groups was achieved following the standard protocols using silver salt. The free Cys peptide was obtained in 70 % yield (**Figure 18**).

Unfortunately, the ligation between this C-terminal peptide and the already synthetized N-terminal was really slow which resulted in the hydrolysis of the F1-F2 fragment that was then no longer able to ligate. Therefore, the amount of desired product was really low. We suspected that traces of silver could interfere with the thiol additive from the buffer, and slow down the ligation. However, with further optimization on the Acm removal this strategy could be used for future synthesis of the protein.

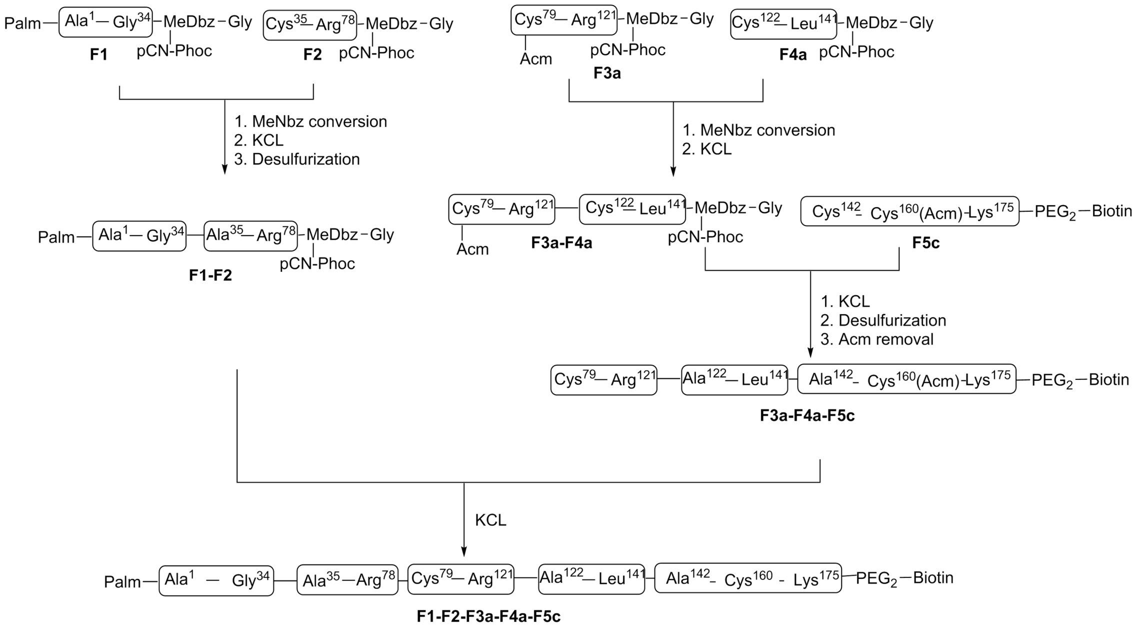

Palm—Ala¹— Gly³⁴—MeDbz-Gly
pCN-Phoc
F1
Cys³⁵— Arg⁷⁸—MeDbz-Gly
F2 pCN-Phoc
Cys⁷⁹— Arg¹²¹—MeDbz-Gly
Acm
pCN-Phoc
F3a
Cys¹²²— Leu¹⁴¹—MeDbz-Gly
F4a pCN-Phoc
1. MeNbz conversion
2. KCL
3. Desulfurization
1. MeNbz conversion
2. KCL
Palm—Ala¹— Gly³⁴—Ala³⁵— Arg⁷⁸—MeDbz-Gly
pCN-Phoc
F1-F2
Cys⁷⁹— Arg¹²¹—Cys¹²²— Leu¹⁴¹—MeDbz-Gly
Acm
F3a-F4a
pCN-Phoc
Cys¹⁴²– Cys¹⁶⁰(Acm)-Lys¹⁷⁵—PEG₂—Biotin
F5c
1. KCL
2. Desulfurization
3. Acm removal
Cys⁷⁹— Arg¹²¹—Ala¹²²— Leu¹⁴¹—Ala¹⁴²– Cys¹⁶⁰(Acm)-Lys¹⁷⁵—PEG₂—Biotin
F3a-F4a-F5c
KCL
Palm—Ala¹ — Gly³⁴—Ala³⁵—Arg⁷⁸—Cys⁷⁹— Arg¹²¹—Ala¹²²— Leu¹⁴¹—Ala¹⁴²– Cys¹⁶⁰ - Lys¹⁷⁵—PEG₂—Biotin
F1-F2-F3a-F4a-F5c

Figure 17. Alternative synthetic strategy using F3b.

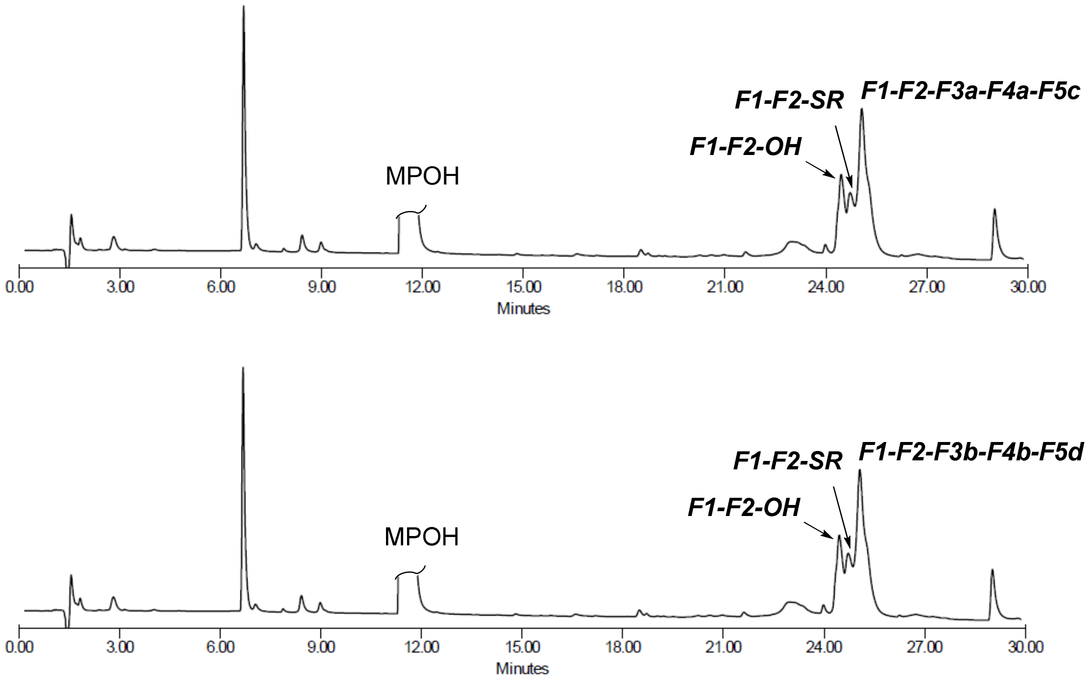


Figure 18. HPLC of the KCL between F1-F2 and F3a-F4b-F5c.

2.4.3. Thz removal of the F3-F4a-F5c

The next step in the synthesis of Palm-ShhN-biotin protein was the Thz removal. The peptide was dissolved in a buffer containing 0.3 M methoxylamine, and the pH was adjusted to 4.5. The reaction was carried out at 40 °C and was completed after 5 hours (**Figure *19***). After purification the product was obtained in 49 % yield.

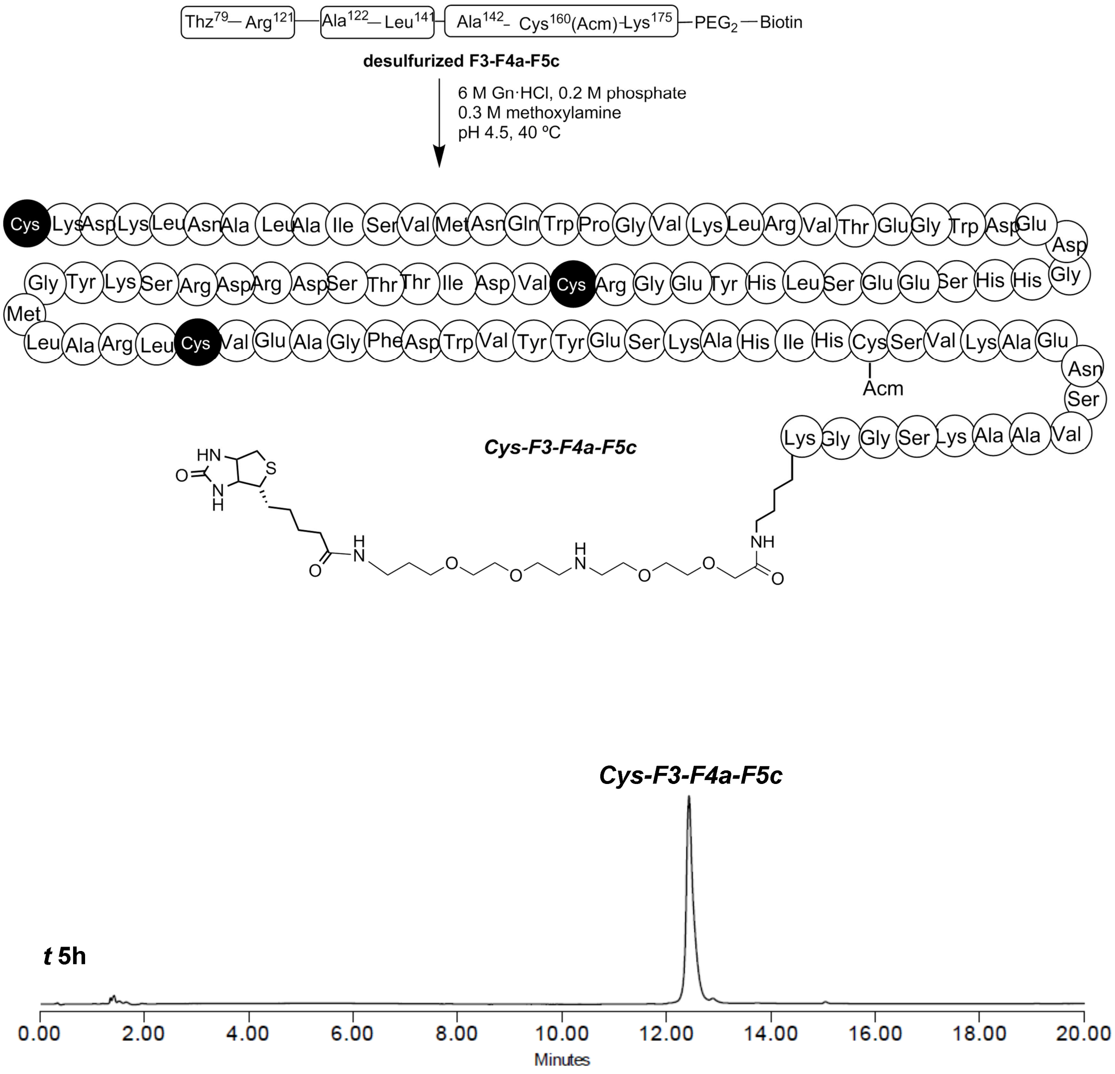

Thz79— Arg121 — Ala122— Leu141 — Ala142– Cys160(Acm)-Lys175 —PEG2—Biotin
desulfurized F3-F4a-F5c
6 M Gn·HCl, 0.2 M phosphate
0.3 M methoxylamine
pH 4.5, 40 ºC
Cys Lys Asp Lys Leu Asn Ala Leu Ala Ile Ser Val Met Asn Gln Trp Pro Gly Val Lys Leu Arg Val Thr Glu Gly Trp Asp Glu
Asp
Gly Tyr Lys Ser Arg Asp Arg Asp Ser Thr Thr Ile Asp Val Cys Arg Gly Glu Tyr His Leu Ser Glu Glu Ser His His Gly
Met
Leu Ala Arg Leu Cys Val Glu Ala Gly Phe Asp Trp Val Tyr Tyr Glu Ser Lys Ala His Ile His Cys Ser Val Lys Ala Glu
Acm
Asn
Ser
Lys Gly Gly Ser Lys Ala Ala Val
Cys-F3-F4a-F5c
HN
S
O
N
H
H
N
O
O
O
H
N
O
O
NH
O
Cys-F3-F4a-F5c
t 5h
0.00
2.00
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00
Minutes

Figure 19. Top: Thz removal from F3-F4a-F5c. Bottom: HPLC of the reaction.

2.5. Assembling of the C-terminal and N-terminal parts

2.5.1. NCL between F1-F2 and F3-F4a-F5c

To finally access the protein both fragments were ligated by dissolving them in a NCL buffer containing 100 mM MPOH at pH 7.5. (**Figure *20***). After overnight reaction and following purification, the Acm-protected protein was isolated in 49 %.

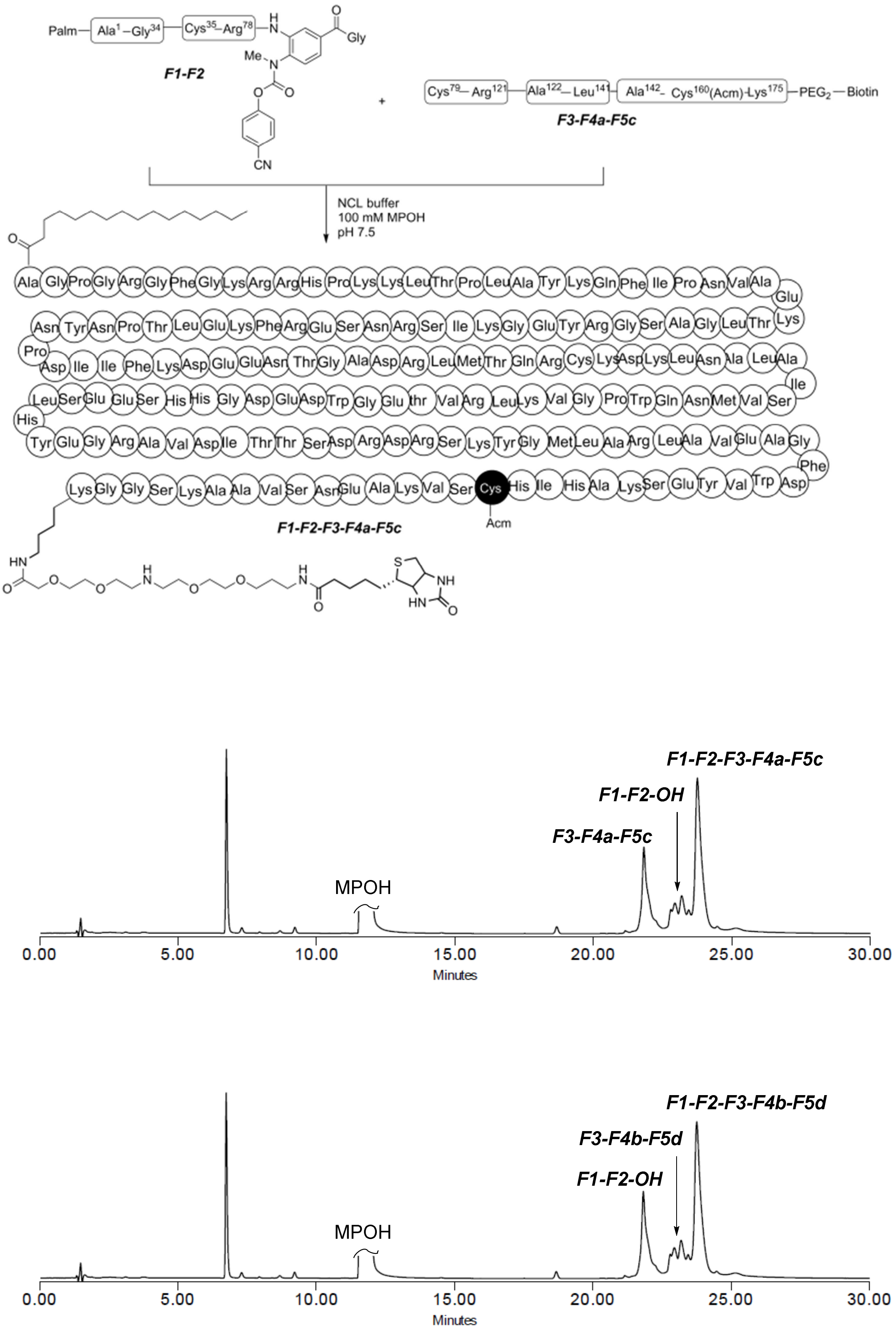
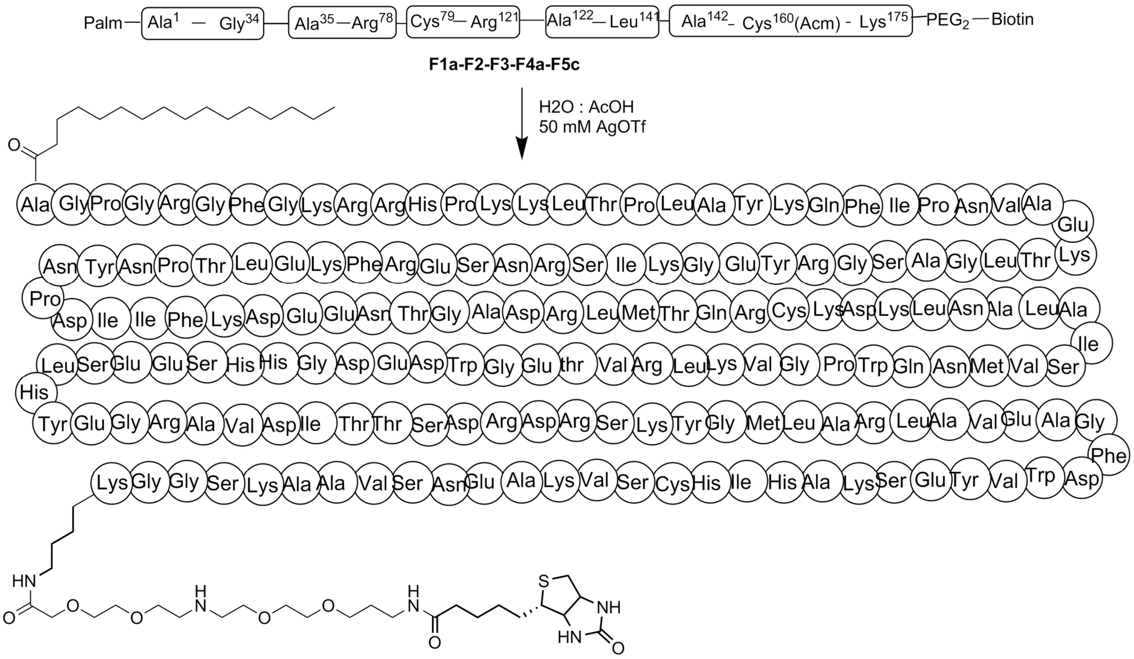

Palm
Ala1–Gly34
Cys35–Arg78
H N
O
Gly
Me N
O O
CN
F1-F2
+
Cys79–Arg121
Ala122–Leu141
Ala142–Cys160(Acm)-Lys175
PEG2
Biotin
F3-F4a-F5c
NCL buffer
100 mM MPOH
pH 7.5
Ala Gly Pro Gly Arg Gly Phe Gly Lys Arg Arg His Pro Lys Lys Leu Thr Pro Leu Ala Tyr Lys Gln Phe Ile Pro Asn Val Ala Glu
Asn Tyr Asn Pro Thr Leu Glu Lys Phe Arg Glu Ser Asn Arg Ser Ile Lys Gly Glu Tyr Arg Gly Ser Ala Gly Leu Thr Lys
Pro
Asp Ile Ile Phe Lys Asp Glu Glu Asn Thr Gly Ala Asp Arg Leu Met Thr Gln Arg Cys Lys Asp Lys Leu Asn Ala Leu Ala
Ile
Leu Ser Glu Glu Ser His His Gly Asp Glu Asp Trp Gly Glu thr Val Arg Leu Lys Val Gly Pro Trp Gln Asn Met Val Ser
His
Tyr Glu Gly Arg Ala Val Asp Ile Thr Thr Ser Asp Arg Asp Arg Ser Lys Tyr Gly Met Leu Ala Arg Leu Ala Val Glu Ala Gly
Phe
Lys Gly Gly Ser Lys Ala Ala Val Ser Asn Glu Ala Lys Val Ser Cys His Ile His Ala Lys Ser Glu Tyr Val Trp Asp
Acm
F1-F2-F3-F4a-F5c
HN O O O H N O O H N S NH HN O O
F1-F2-F3-F4a-F5c
F1-F2-OH
F3-F4a-F5c
MPOH
0.00
5.00
10.00
15.00
20.00
25.00
30.00
Minutes
F1-F2-F3-F4b-F5d
F3-F4b-F5d
F1-F2-OH
MPOH
0.00
5.00
10.00
15.00
20.00
25.00
30.00
Minutes

Figure 20. Top: scheme of the NCL between F1-F2 and F3-F4a-F5c. Bottom: HPLC of the reaction.

2.5.2. Acm removal of F1-F2-F3-F4b-F5d

In the final step the Acm was removed from Cys^{160} (**Scheme *16***). The reaction was carried out dissolving the peptide in an aqueous acetic buffer containing AgOTf (50 mM). After complete deprotection, the excess of AgOTf was eliminated with a solution of DTT and the resulting crude purified by HPLC to afford the Palm-ShhN-biotin protein in 65 % yield (**Figure *21***).


Palm—Ala1 — Gly34—Ala35—Arg78—Cys79—Arg121—Ala122—Leu141—Ala142- Cys160(Acm) - Lys175—PEG2—Biotin
F1a-F2-F3-F4a-F5c
H2O : AcOH
50 mM AgOTf
O
Ala Gly Pro Gly Arg Gly Phe Gly Lys Arg Arg His Pro Lys Lys Leu Thr Pro Leu Ala Tyr Lys Gln Phe Ile Pro Asn Val Ala Glu
Asn Tyr Asn Pro Thr Leu Glu Lys Phe Arg Glu Ser Asn Arg Ser Ile Lys Gly Glu Tyr Arg Gly Ser Ala Gly Leu Thr Lys
Pro
Asp Ile Ile Phe Lys Asp Glu Glu Asn Thr Gly Ala Asp Arg Leu Met Thr Gln Arg Cys Lys Asp Lys Leu Asn Ala Leu Ala Ile
Leu Ser Glu Glu Ser His His Gly Asp Glu Asp Trp Gly Glu thr Val Arg Leu Lys Val Gly Pro Trp Gln Asn Met Val Ser
His
Tyr Glu Gly Arg Ala Val Asp Ile Thr Thr Ser Asp Arg Asp Arg Ser Lys Tyr Gly Met Leu Ala Arg Leu Ala Val Glu Ala Gly Phe
Lys Gly Gly Ser Lys Ala Ala Val Ser Asn Glu Ala Lys Val Ser Cys His Ile His Ala Lys Ser Glu Tyr Val Trp Asp
HN
O
O
O
H N
O
O
H N
O
S
NH
HN
O

Scheme 16. Acm removal of the Palm-ShhN-biotin.

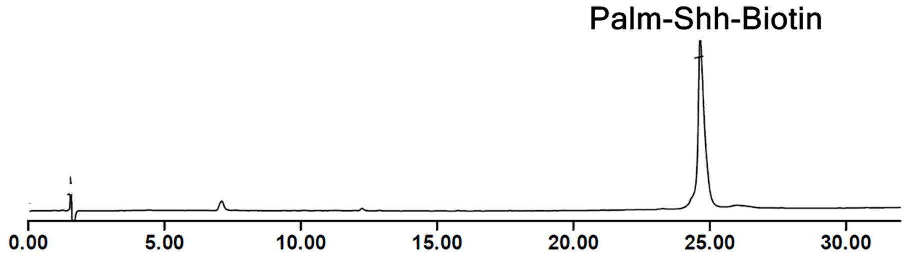


Figure 21. HPLC of the pure Palm-ShhN-Biotin.

2.6. Folding

There are two main strategies to fold or re-fold a protein: dialysis or dilution, each one with advantages and disavantages. The selection of one depends on the nature of the protein. Moreover, not just the folding strategy, but also the selection of the proper buffer and the use of additives has to be considered. Normally, either dilution or dialysis, the inclusion of additives is critic for the success of the folding.[63,64] For instance, the addition of denaturants such as urea or guanidinium at low concentration can help to properly fold the protein. Also, inhibitors of aggregation such as Pro or Arg in the folding buffer decrease the protein precipitation, similar to the behavior observed upon the addition of stabilizers like glycerol, PEG or sugars. Moreover, when working with membrane, hydrophobic or lipidated proteins, the use of detergents may also be required. To this category belong different type of detergents such as sodium-dodecyl-sulfate (SDS), triton, n-dodecyl-ß-D-maltoside (DDM), octyl-β-ᴅ-glucopyranoside or CHAPS.[65,66]

In specific cases, such as metalloproteins the addition of cationic metals is necessary for the proper folding of the protein. In our case, ShhN is known to interact with zinc.[67] In other situations, the presence of disulfide bonds demands the addition of agents that keep the proper redox environment for the correct disulfide formation.

Having in mind these considerations, it was decided first a dialysis to fold the synthetic Palm-ShhN-biotin, in where the gradual and slow removal of denaturant from the buffer could allow to fold the protein and avoid its aggregation[b].

2.6.1. . Folding of Palm-ShhN-biotin by dialysis

First attempts driven to fold ShhN analog were based on dialysis using small display systems equipped with a membrane cut off of 5000 kDa. Different buffers, temperatures and additives were screened and the results were analyzed by circular dichroism (CD).

The majority of the conditions tested led to the precipitation of the protein (**Table *3***). Unfortunately, in those assays in where CD could be performed, the spectra did not match with the CD obtained from the recombinant ShhN, used as control.

---

[b] Attempts to fold the protein without denaturing agent led to precipitation.

Buffer 1	Buffer 2	Time	Temp	Observations
10 % TFE, 150 mM NaCl, 2mMTCEP, 25 mM NaPi	150 mM NaCl, 2mMTCEP, 25 mM NaPi	24 h	4 °C	CD did not match the reference
10 % TFE, 150 mM NaCl, 2mM TCEP, 25 mM NaPi, 0,5%DDM	150 mM NaCl, 2mMTCEP, 25 mM NaPi, 0,5%DDM	18 h	4 °C	CD did not match the reference
6M GnHCl, 150 mM NaCl, 2 mM TCEP, 25 mM NaPi	150 mM NaCl, 2 mM TCEP, 25 mM NaPi	18 h	4 °C	CD did not match the reference
6M GnHCl, 150 mM NaCl, 25mM NaPi, 2mM TCEP, 0,5% DDM	4M GnHCl, 150 mM NaCl, 25 mM NaPi, 2mM TCEP, 0,5% DDM	24 h	rT	
	1M GnHCl, 150 mM NaCl, 25 mM NaPi, 2mM TCEP, 0,5% DDM	24 h	rT	
	150 mM NaCl, 25 mM NaPi, 2mM TCEP, 0,5% DDM	18 h	rT	Cloudy solution
6 M GnHCl, 150 mM NaCl, 10 mM NaPi, 0,5 mM DTT, 1% octyl	1M GnHCl, 150 mM NaCl, 10 mM NaPi, 0,5 mM DTT, 1% octyl	24 h	rT	
	150 mM NaCl, 10 mM NaPi, 0,5 mM DTT, 1% octyl	24 h	rT	Precipitate
6 M GnHCl, 150 mM NaCl, 10 mM NaPi, 0,5 mM DTT, 1% octyl	1M GnHCl, 150 mM NaCl, 10 mM NaPi, 0,5 mM DTT, 1% octyl	24 h	rT	Precipitate
6M GnHCl, 150 mM NaCl, 10 mM NaPi, 0,5 mM DTT, 1% octyl, 0,5 M sucrose	2M GnHCl, 150 mM NaCl, 10 mM NaPi, 0,5 mM DTT, 0,5 M sucrose	24 h	4 °C	
	0,5 M Arg, 150 mM NaCl, 10 mM NaPi, 0,5 mM DTT, 1% octyl, 0,5 M sucrose	24 h	4 °C	Precipitate
6M GnHCl, 150 mM NaCl, 10 mM NaPi, 0,5 mM DTT, 1% octyl, 0,5 M sucrose	150 mM NaCl, 10 mM NaPi, 1% octyl, 0,5 M sucrose, 0,5 M Arg, 0,5 mM DTT	18 h	4 °C	Precipitate

Table 3. Dialysis conditions assayed for folding of Palm-ShhN-biotin. Buffer 1: buffer in which protein is first dissolved inside the cassette. Buffer 2: buffer in which we performed the dialysis. Octyl = octyl-β-ᴅ-glucopyranoside, NaPi = phosphate. All experiments were performed at 5 µM of Palm-ShhN-biotin.

Based on these results, a quick dilution approach was attempted. In the first place, folding experiments were carried out at very low scale (5 µg of protein). The ratio of soluble fraction of the protein after dilution could be easily identified by gel electrophoresis. These conditions allowed the screening of several different parameters simultaneously and in a fast manner. Then, the best conditions would be chosen to scale-up the folding and register a CD.

2.6.2. Protein Gel Electrophoresis and silver staining

Polyacrylamide gel electrophoresis (PAGE) is one of the main techniques used for the study of proteins. Although it is a qualitative technique, it has been used by some groups to detect the relative precipitation of the protein during refolding strategies [66]. It is a very sensitive technique that allows to detect few ng of protein. Therefore, PAGE followed by silver staining is a good methodology to detect small amount of precipitate while performing screenings.

In the first place it was determined that the minimum concentration of protein that could be detected by silver staining was around 50 nM, although at 20 nM it is possible to detect faint bands of the protein. So, folding experiments were carried out at a minimum of 50 nM.

2.6.3. Quick dilution strategies

In the quick dilution technique, the protein was first dissolved in a small volume of denaturant buffer and next diluted 20-30-fold in the corresponding refolding buffer. In order to get rid of the aggregates or miss-folded protein that precipitates, the solution is centrifuged and the soluble fraction further concentrated.

Buffer	NaCl/DTT	Detergent	Other additives	pH
5 mM phosphate	150 mM / 5mM	1% octyl	-	7
5 mM phosphate	150 mM / 5mM	1% octyl	0.5 µM $ZnCl_2$	8
50 mM HEPES	150 mM / 2.5 mM	1% octyl	0.5 µM $ZnCl_2$	8
50 mM HEPES	150 mM / 2.5 mM	1% octyl	-	7
5 mM phosphate	150 mM / 5 mM DTT	1% octyl	-	5
25 mM phosphate	150 mM NaCl / 2 mM DTT	0.5 % DDM	-	7
25 mM phosphate	150 mM NaCl / 2 mM DTT	0.5 % DDM	0.5 µM $ZnCl_2$	8

Table 4. Quick dilution conditions assayed for the folding of Palm-ShhN-biotin. Protein concentration in the refolding buffer was diluted to 0.5 µM. Octyl= octyl-β-ᴅ-glucopyranoside.

Table *4* shows different experimental conditions screened in this approach. Some of them are based on the observations from previous dialysis experiments, while others include the addition of detergents, stabilizers and $ZnCl_2$. For instance, in some folding protocols, palmitoylated ShhN was dissolved in a buffer containing octyl-β-ᴅ-glucopyranoside.[28,68] However, in our hands the protein mainly crushed out. Thus, it was decided to try another detergents such as DDM. According to the relative intensity of precipitate protein and the supernatant bands, the best folding results were observed when the protein was diluted in 25 mM phosphate, 150 mM NaCl, 2 mM DTT, 0.5% DDM, 0.5 µM $ZnCl_2$ buffer (**Figure 22**)

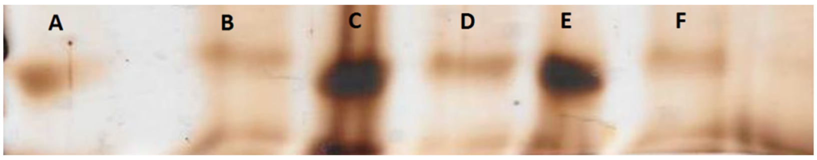


Figure 22. SDS-PAGE of the Palm-ShhN-biotin folding experiment displaying the best conditions: 25 mM NaPi, 150 mM NaCl, 2 mM DTT, 0.5 % DDM, 0.5 µM $ZnCl_2$ at pH 8.0. Lane A: supernatant from folding in buffer containing 5 mM phosphate, 150 mM NaCl, 5 mM DTT, 1% octyl. Lane B: precipitate from A after 1h. Lane C: supernatant from folding in buffer containing 25 mM phosphate, 150 mM NaCl, 2 mM DTT, 0.5% DDM, pH 7. Lane D: precipitate from D after 1h. Lane E: Left: supernatant from folding in buffer containing 25 mM phosphate, 150 mM NaCl, 2 mM DTT, 0.5% DDM, pH 8. Lane F: precipitate from E after 1h.

Next, it was estimated the maximum concentration of protein to what aggregation was not significant yet. Three different concentrations were tested: 1 µM, 0.5 µM and 0.1 µM. After analysis of the supernatant and precipitate by SDS-PAGE, it was determined that folding concentrations equal or above 0.5 µM resulted in higher precipitation. So, ideally, the folding of the protein should be performed at concentrations around 0.1 µM.

2.6.4. Circular Dichroism

Once the best folding approach was identified, it was really essential to know that these conditions resulted in the properly folding of the protein. Thus, CD is a technique that easily allows to know the main characteristics of the secondary structure of a protein. Following the folding conditions previously developed but a larger scale, it was measured the CD spectrum of the Palm-ShhN-biotin at 5 µM and compared with the recombinant non-palmitoylated protein. Although do not perfectly overlay, it is clearly seen in both a negative band at 208 nm corresponding to the formation of a α-helix in the protein. There is also a minimun around 220 nm, although this is more noticeable in the case of the recombinant ShhN. In principle, it appears that part of Palm-ShhN-biotin adopts a helical conformation, which is seen in the structure determined by X-Ray structure (**Figure *23***).

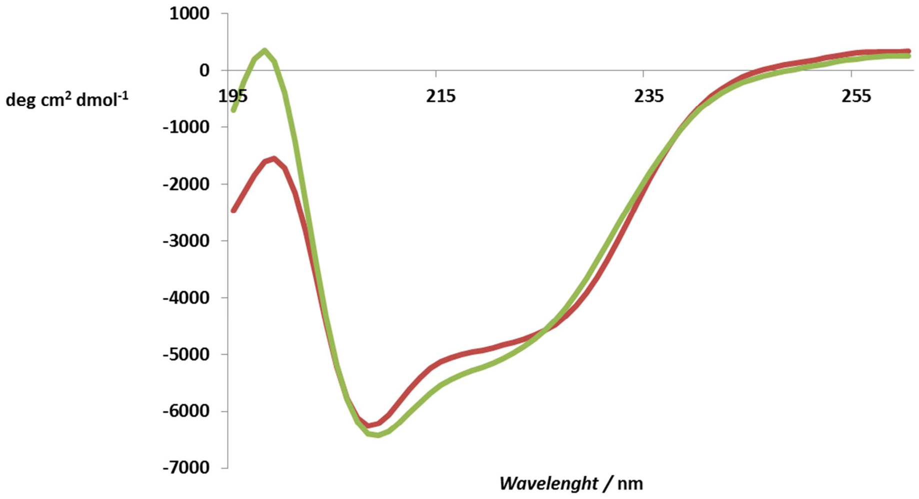


Figure 23. CD spectra of the folding of Palm-ShhN-biotin. In red, recombinant ShhN protein spectrum. In green, spectrum corresponding to the synthetic analog Palm-ShhN-biotin.

2.6.5. ELISA

Another technique that allows to figure out the correct folding of proteins is ELISA. More precisely, the monoclonal 5E1 antibody recognizes a concrete epitope of ShhN, and therefore, requires a proper folding of the protein in order to bind it.[69]

ELISA experiments show that the signal of synthetic-folded Palm-ShhN-biotin is ~40-fold greater than the denatured protein (in guanidine buffer) or BSA protein. Similarly, recombinant ShhN has a stronger response regarding to its denatured state. These results point out that both proteins are recognized by 5E1 and must have close tertiary conformations. For Palm-ShhN-biotin, it seems to indicate that the folding should be pretty much similar to recombinant ShhN (**Figure *24***).

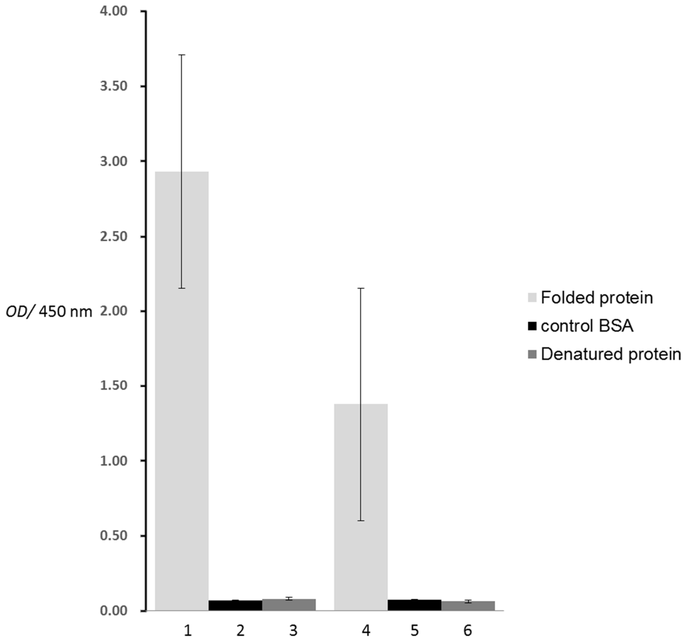


Figure 24. ELISA results from the incubation of monoclonal antibody 5E1 with the synthetic analog Palm-ShhN-biotin (lane 1), recombinant ShhN (lane 4) and their respective controls. Lanes 2 and 5 are BSA controls. Lane 3 = control with denatured Palm-ShhN-biotin. Lane 6 = control with denatured recombinant ShhN.

3. Synthesis of the IleIle-ShhN-biotin

Taking advantage of the synthetic strategy developed for Palm-ShhN-biotin, it was prepared another Sonic analog displaying the pegylated biotin residue at the C-terminus and in the N-terminus an isoleucine dipeptide instead of the hydrophobic fatty. It has been reported that this analog shows an activity 8-fold greater than the ShhN lacking the palmitic.[28]

The main idea underlying the synthesis of this variant was to compare the influence of the palmitic in the biological activity of the protein. Moreover, as one of the main goals is to develop new inhibitors of the Hh pathway that block the Sonic-Patched interaction, the presence or absence of palmitic could have significant influence in the structure of binders developed by phage display. The presence of the IleIle dipeptide should help, in principle, to avoid the aggregation of the protein while mantaining nM activity.

3.1. Retrosynthetic analysis of IleIle-ShhN-biotin

The retrosynthesis is similar to the previously planned for Palm-ShhN-biotin. The protein was divided in five fragments, and only the first (**F1a**), which bears the Ile dipeptide, is different from the others (**Figure *25***).

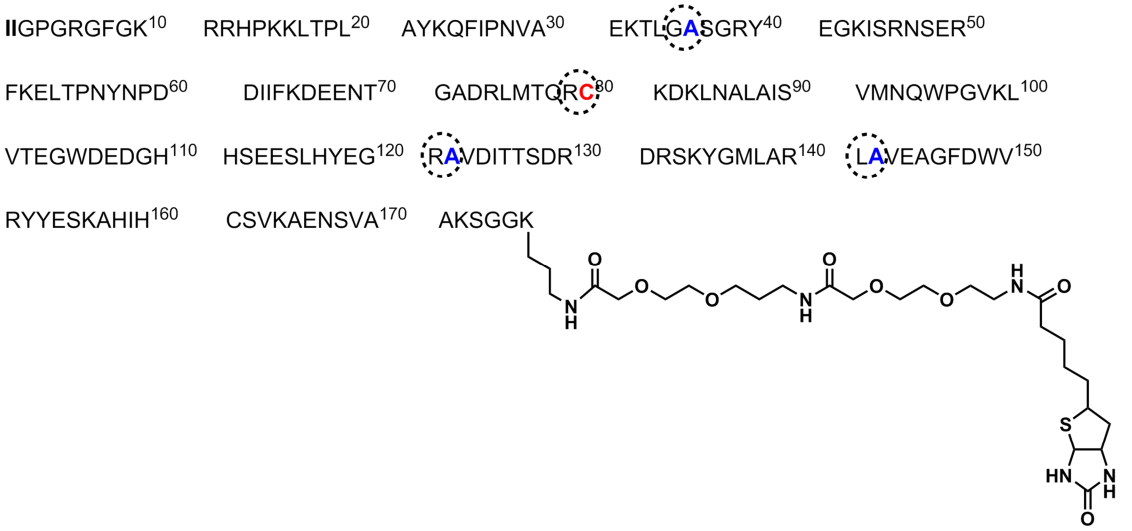

IIGPGRGFGK10 RRHPKKLTPL20 AYKQFIPNVA30 EKTLGASGRY40 EGKISRNSER50
FKELTPNYNPD60 DIIFKDEENT70 GADRLMTQRC80 KDKLNALAIS90 VMNQWPGVKL100
VTEGWDEDGH110 HSEESLHYEG120 RAVDITTSDR130 DRSKYGMLAR140 LAVEAGFDWV150
RYYESKAHIH160 CSVKAENSVA170 AKSGGK
O
N
H
O
O
N
H
O
O
O
H
N
O
S
HN
N
H
O

Figure 25. Retrosynthetic analysis toward IleIle-ShhN-biotin. Circles and letters in colors indicate ligation sites. In blue, the Ala residues which would be mutated into Cys for ligation purposes. In red, natural Cys used as ligation points.

3.2. Synthetic strategy towards IleIle-ShhN-biotin

The synthetic strategy followed the same approach described for the synthesis of the Palm-ShhN-biotin: the N-terminal half, comprising the **F1a-F2**, and the C-terminal side consisting on the **F3**, **F4b** and **F5c** peptides.

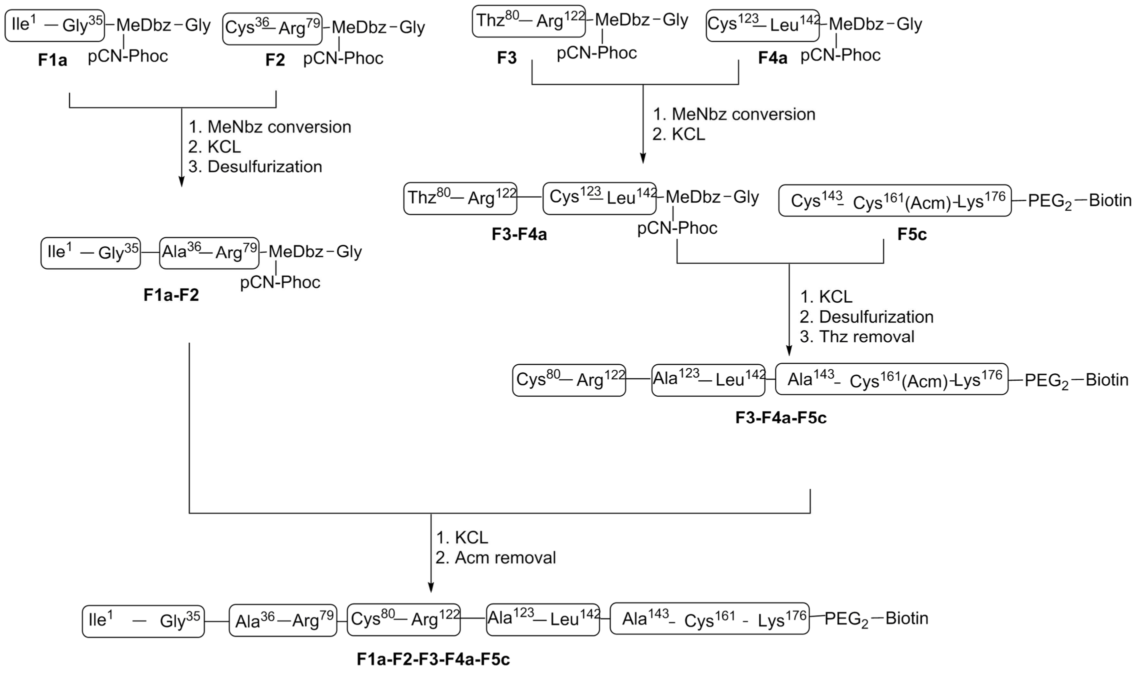

Ile1 — Gly35 -MeDbz-Gly
pCN-Phoc
F1a
Cys36— Arg79 -MeDbz-Gly
pCN-Phoc
F2
1. MeNbz conversion
2. KCL
3. Desulfurization
Ile1 — Gly35 — Ala36— Arg79 -MeDbz-Gly
pCN-Phoc
F1a-F2
Thz80— Arg122 -MeDbz-Gly
pCN-Phoc
F3
Cys123— Leu142 -MeDbz-Gly
pCN-Phoc
F4a
1. MeNbz conversion
2. KCL
Thz80— Arg122 — Cys123 — Leu142 -MeDbz-Gly
pCN-Phoc
F3-F4a
Cys143 - Cys161(Acm)-Lys176 —PEG2—Biotin
F5c
1. KCL
2. Desulfurization
3. Thz removal
Cys80— Arg122 — Ala123 — Leu142 — Ala143 - Cys161(Acm)-Lys176 —PEG2—Biotin
F3-F4a-F5c
1. KCL
2. Acm removal
Ile1 — Gly35 — Ala36—Arg79 — Cys80— Arg122 — Ala123— Leu142 — Ala143 - Cys161 - Lys176 —PEG2—Biotin
F1a-F2-F3-F4a-F5c

Figure 26. Synthetic strategy toward the IleIle-ShhN-biotin analog.

For the synthesis of the N-terminal block, the **F1a** is ligated with **F2** and the resulting product desulfurized.

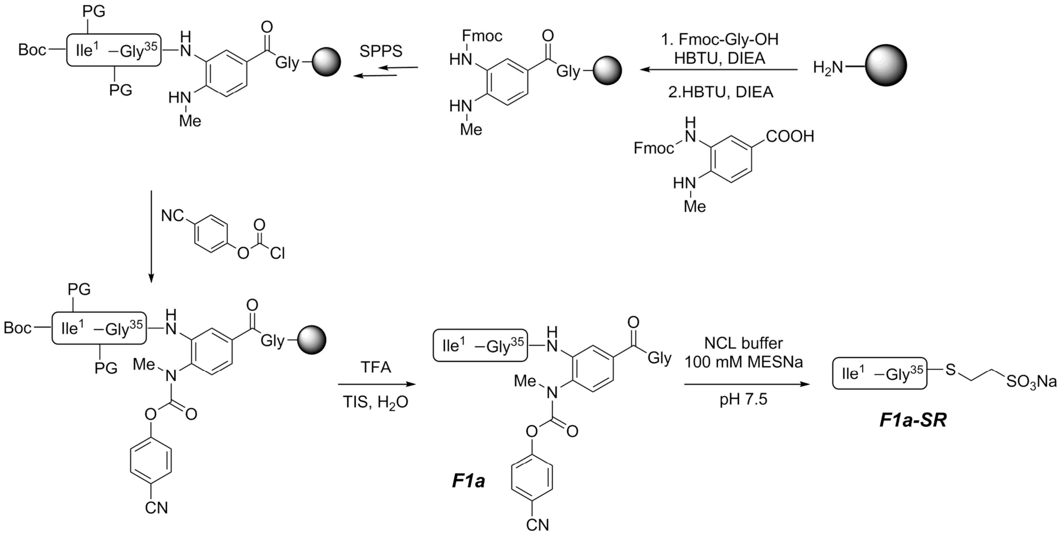
On the other hand, the synthesis of the C-terminal block is equivalent to the Palm-ShhN-biotin, and described above in 2.3.

The final step consists on the ligation between the N and the C fragments followed by Acm removal (**Figure *26***).

3.3. Synthesis of the N-terminal block: Ile^1-Arg^{79}-*p*CN-Phoc-MeDbz

3.3.1. Synthesis of the F1a: Ile^1-Gly^{35}-pCN-Phoc-MeDbz

The fragment **F1a** was prepared following the standard procedure described. The last aminoacid was coupled as Boc-Ile. Upon cleaved off the resin, it was dissolved in NCL buffer containing MESNa (100 mM), and pH adjusted to 7.5 in order to trigger the linker cyclization and the thioesterification at the C-terminus (**Scheme *17***).


PG
Boc
Ile¹ –Gly³⁵
H
N
O
Gly
HN
Me
SPPS
Fmoc
HN
1. Fmoc-Gly-OH
HBTU, DIEA
2.HBTU, DIEA
H₂N
Fmoc
COOH
NC
O
Cl
TFA
TIS, H₂O
CN
F1a
NCL buffer
100 mM MESNa
pH 7.5
S
SO₃Na
F1a-SR

Scheme 17. Synthesis of the F1a. PG = Protecting group.

3.3.2. KCL between F1a and F2

F1a was ligated to **F2** in NCL buffer containing MPOH (100 mM) at pH 7.0. The ligation gave the desired ligated peptide after 2 hours of reaction (**Figure *27***). The ligated product was obtained in 52 % yield after purification.

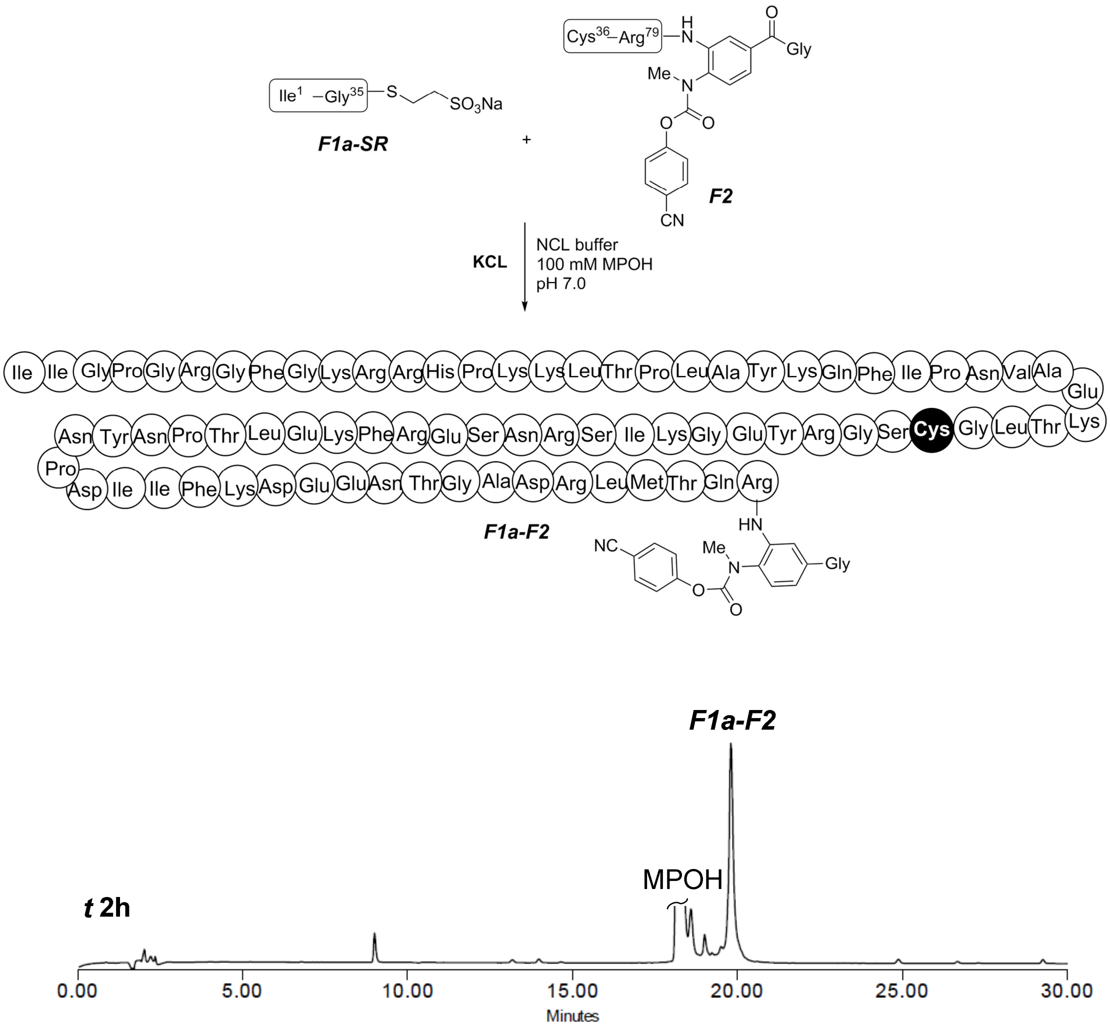

Cys36–Arg79
Ile1 –Gly35
SO3Na
F1a-SR
+
F2
Gly
Me
CN
KCL
NCL buffer
100 mM MPOH
pH 7.0
Ile Ile Gly Pro Gly Arg Gly Phe Gly Lys Arg Arg His Pro Lys Lys Leu Thr Pro Leu Ala Tyr Lys Gln Phe Ile Pro Asn Val Ala Glu
Lys Thr Leu Gly Cys Ser Gly Arg Tyr Glu Gly Lys Ile Ser Arg Asn Ser Glu Arg Phe Lys Glu Leu Thr Pro Asn Tyr Asn
Pro Asp Ile Ile Phe Lys Asp Glu Glu Asn Thr Gly Ala Asp Arg Leu Met Thr Gln Arg
F1a-F2
NC
HN
F1a-F2
MPOH
t 2h
0.00
5.00
10.00
15.00
20.00
25.00
30.00
Minutes

Figure 27. Top: KCL scheme between **F1a** and **F2**. Bottom: HPLC of the ligation after 2hours.

3.3.3. Desulfurization F1a-F2-pCN-Phoc-MeDbz

The Cys^{36} of the ligated **F1a-F2** was next reduced to Ala following the procedure reported before in section 1.6.1.2. After two hours of reaction at 37 °C, it was obtained the completed desulfurized product in 43 % yield (**Figure *28***).

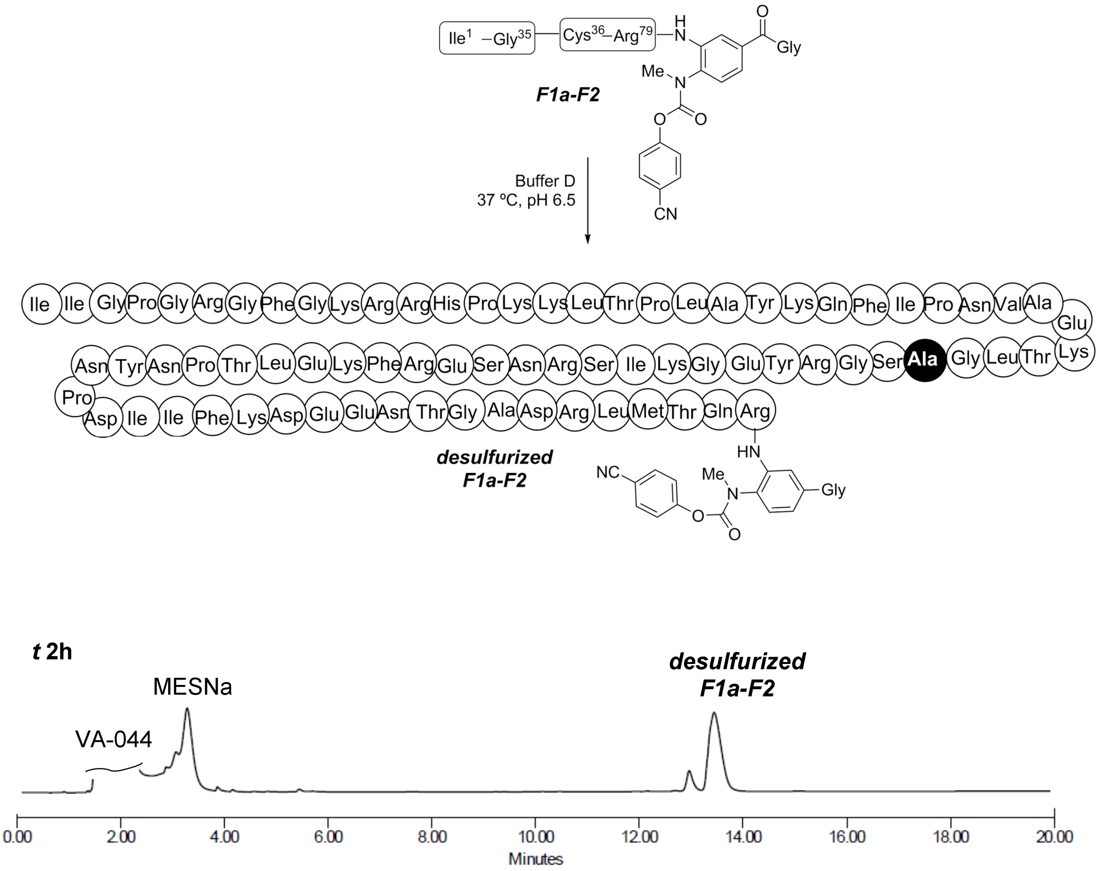

Ile1 –Gly35
Cys36–Arg79
F1a-F2
Me
N
H
O
Gly
CN
Buffer D
37 ºC, pH 6.5
Ala
desulfurized F1a-F2
NC
HN
t 2h
VA-044
MESNa
desulfurized F1a-F2
0.00
2.00
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00
Minutes

Figure 28. Top: Scheme of the desulfurization of F1a-F2 peptide. Bottom: HPLC of the reaction.

3.3.4. NCL between F1a-F2 and F3-F4a-F5c

Both fragments, **F1a-F2** and **F3-F4b-F5c**, were dissolved in NCL buffer containing MPOH (100 mM) at pH 7.5. The ligation reached completion after 18 hours of reaction (**Figure *29***).

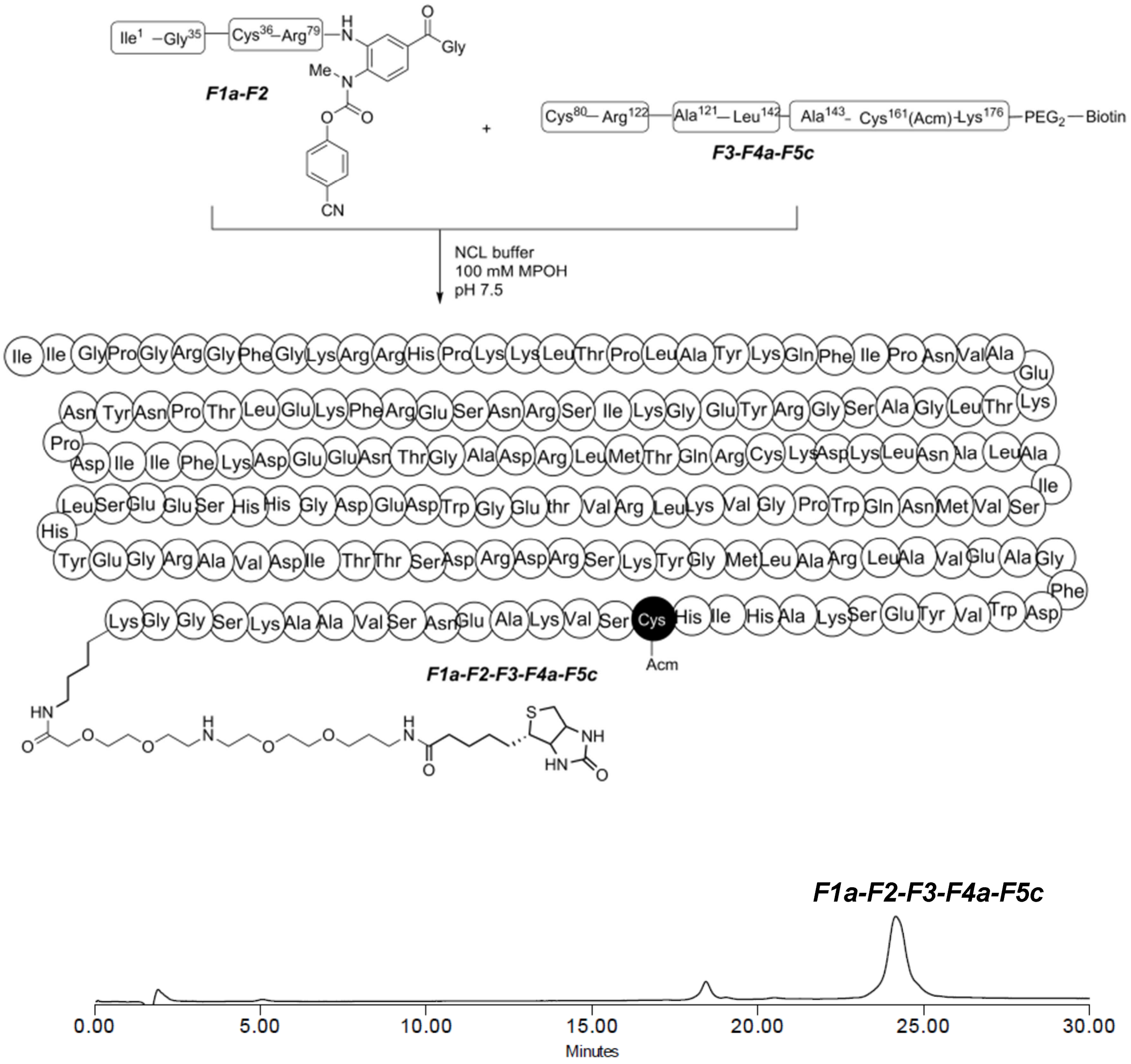

Ile¹ –Gly³⁵
Cys³⁶–Arg⁷⁹
F1a-F2
Me
CN
Gly
+
Cys⁸⁰– Arg¹²²
Ala¹²¹– Leu¹⁴²
Ala¹⁴³– Cys¹⁶¹(Acm)-Lys¹⁷⁶
PEG₂—Biotin
F3-F4a-F5c
NCL buffer
100 mM MPOH
pH 7.5
Cys
Acm
F1a-F2-F3-F4a-F5c
F1a-F2-F3-F4a-F5c
0.00
5.00
10.00
15.00
20.00
25.00
30.00
Minutes

Figure 29. Top: scheme of the KCL between F1a-F2 and F3-F4b-F5c. Bottom: HPLC of the ligation.

3.3.5. Acm removal F1a-F2-F3-F4a-F5c

The last step to finally achieve the synthesis of the IleIle-ShhN-biotin was the removal of the Acm protecting the native Cys^{161} (**Figure *30***). The reaction was undertaken using the same protocol described in section 2.3.5. As expected, the desired free Cys was observed and the protein isolated through purification by HPLC in 62 % (**Figure *31***).

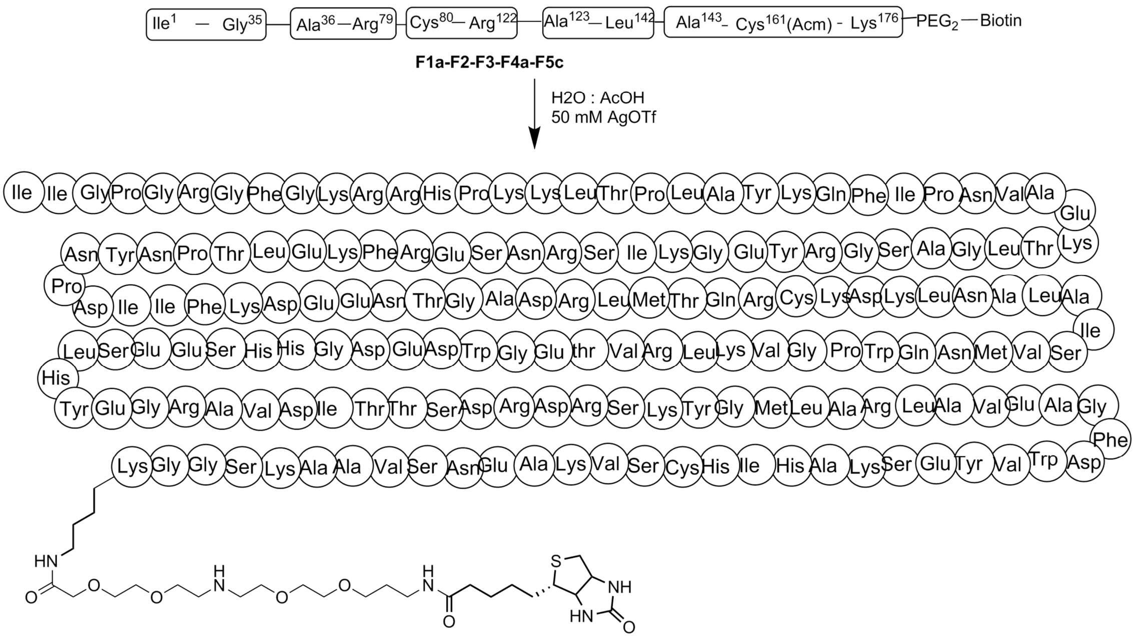

Ile¹ — Gly³⁵ — Ala³⁶—Arg⁷⁹ — Cys⁸⁰— Arg¹²² — Ala¹²³— Leu¹⁴² — Ala¹⁴³– Cys¹⁶¹(Acm) - Lys¹⁷⁶ —PEG₂—Biotin
F1a-F2-F3-F4a-F5c
H2O : AcOH
50 mM AgOTf
Ile Ile Gly Pro Gly Arg Gly Phe Gly Lys Arg Arg His Pro Lys Lys Leu Thr Pro Leu Ala Tyr Lys Gln Phe Ile Pro Asn Val Ala Glu
Asn Tyr Asn Pro Thr Leu Glu Lys Phe Arg Glu Ser Asn Arg Ser Ile Lys Gly Glu Tyr Arg Gly Ser Ala Gly Leu Thr Lys
Pro
Asp Ile Ile Phe Lys Asp Glu Glu Asn Thr Gly Ala Asp Arg Leu Met Thr Gln Arg Cys Lys Asp Lys Leu Asn Ala Leu Ala
Ile
Leu Ser Glu Glu Ser His His Gly Asp Glu Asp Trp Gly Glu thr Val Arg Leu Lys Val Gly Pro Trp Gln Asn Met Val Ser
His
Tyr Glu Gly Arg Ala Val Asp Ile Thr Thr Ser Asp Arg Asp Arg Ser Lys Tyr Gly Met Leu Ala Arg Leu Ala Val Glu Ala Gly
Phe
Lys Gly Gly Ser Lys Ala Ala Val Ser Asn Glu Ala Lys Val Ser Cys His Ile His Ala Lys Ser Glu Tyr Val Trp Asp
HN
O
O
O
H
N
O
O
H
N
O
S
NH
HN
O

Figure 30. Scheme of the Acm removal of the fully synthetized IleIle-Shh-biotin.

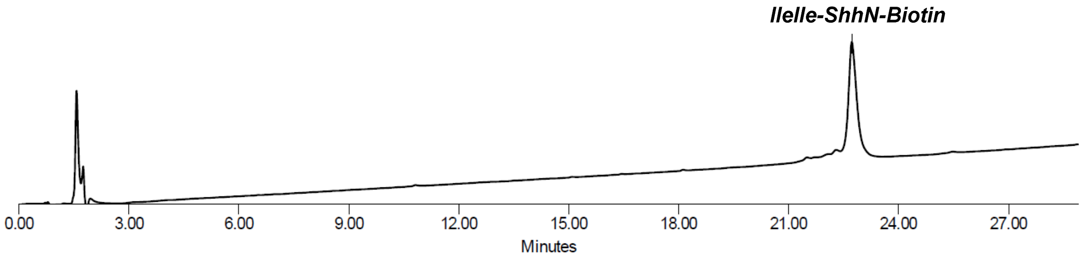


Figure 31. HPLC of the pure IleIle-ShhN-Biotin protein.

3.3.6. Folding of the IleIle-ShhN-biotin

Although this synthetic analog does not have the palmitic acid, and maybe it would not require detergent for the folding, it was decided to apply the same folding conditions in order to have a direct comparison with Palm-ShhN-biotin.

Therefore, from a denatured (in guanidine buffer) stock solution of the protein, it was quickly diluted to a final 0.1 µM concentration into the refolding buffer, incubated for 30 minutes and centrifuged to separate the insoluble aggregates. Next, the supernatant was concentrated to a final concentration of 5 µM andCD was carried out. Importantly, the spectra of recombinant and IleIle-ShhN-biotin show similar characteristics, with negative bands around at 208 and 220 nm what indicates that both proteins have an analogous structure (**Figure *32***).

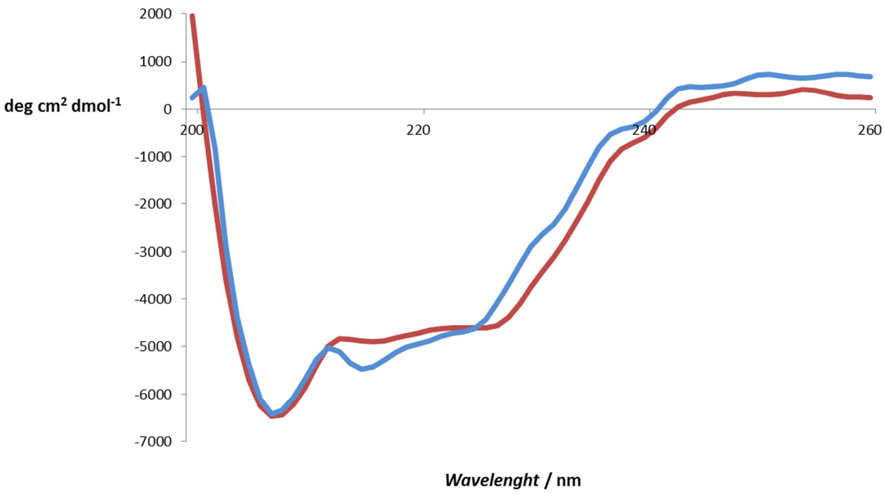


Figure 32. CD spectrum of IleIle-ShhN-biotin (blue) and recombinant ShhN (red).

ELISA experiments further confirmed that IleIle-ShhN-biotin is specifically recognized by 5E1, with a similar response to recombinant ShhN. Even when this positive result seems to confirm a proper folding of both, IleIle-ShhN-biotin and Palm-ShhN-biotin, additional experiments, i. e. NMR and cell-based assays, must validate the structure (**Figure 33**).

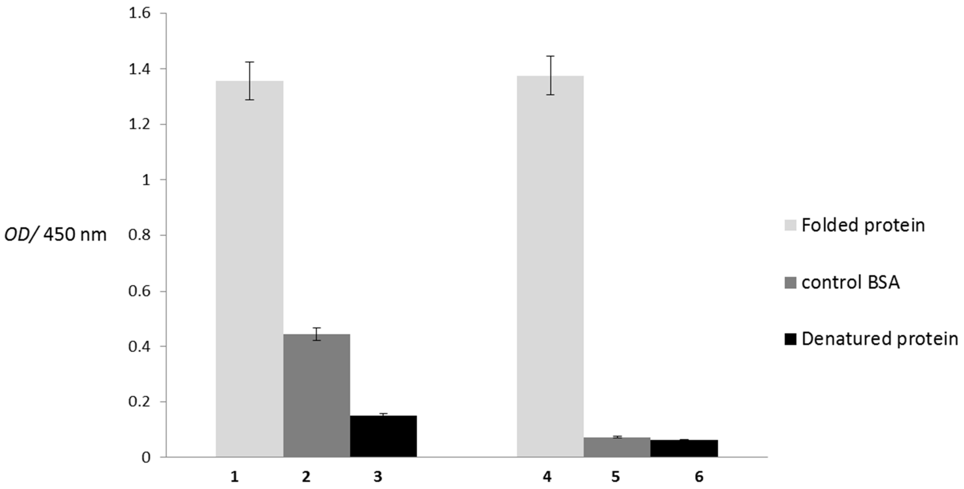


Figure 33. ELISA results from the incubation of monoclonal antibody 5E1 with the synthetic analog IleIle-ShhN-biotin (1) and recombinant ShhN (4), with their respective controls. Lanes 2 and 5, control with BSA. Lane 3, control with denatured Ile-ShhN-biotin. Lane 6, control with denatured recombinant ShhN.

CONCLUSIONS

The conclusions of this chapter focused on the synthesis of ShhN are:

1. *p*CN-Phoc-MeDbz peptides can be used for the synthesis of large protein like ShhN. In this chapter, large *p*CN-Phoc-MeDbz peptides, from 20 to 44 residues have been prepared. The fragments were synthetized in high purity by Fmoc-SPPS, and no side products from the linker or acylation with the chloroformate were detected.

2. The *p*CN-Phoc-MeDbz peptide fragments were used in KCL and NCL. The corresponding thioester peptides were successfully obtained in solution by the cyclization of the MeDbz linker *in situ* and its posterior thioesterification.

3. The *p*CN-Phoc-MeDbz peptides are stable under radical based desulfurization conditions, and thiazolidine and Acm removal standard protocols.

4. Two analogs of ShhN, a 174-mer protein, were chemically synthesized: Palm-ShhN-biotin analog, bearing a palmitic acid at the N-terminus and a pegylated biotin residue at the C-terminus, and IleIle-ShhN-biotin which shares the pegylated biotin residue at the C-terminus, but changes the lipid moiety for an Ile dipeptide at the N-terminus.

5. A fast protocol that efficiently folds the proteins by quick dilution has been developed.

6. The proper folding was ensured by circular dichroism and ELISA assays using the monoclonal antibody 5E1.

7. A cholesterol residue, present in the native ShhN, has been incorporated to the C-terminus of the Cys^{169}-Gly^{174} peptide. Although at the moment it has not been possible to assemble the protein carrying the cholesterol due to solubility issues, it represents a beginning for the full synthesis of ShhN. Currently, it is being explored the ligation adding co-solvents that increase the solubility of this cholesterol peptides.

REFERENCES

(1) Heretsch, P.; Tzagkaroulaki, L.; Giannis, A. Modulators of the Hedgehog Signaling Pathway. *Bioorg. Med. Chem. Lett.* **2010**, *18*, 6613–6624.

(2) Varjosalo, M.; Taipale, J. Hedgehog: Functions and Mechanisms. *Genes Dev* **2008**, *18*, 2454–2472.

(3) Nüsslein-Volhard, C.; Wieschaus, E. Mutations Affecting Segment Number and Polarity in Drosophila. *Nature* **1980,** *287*, 795-801.

(4) Nüsslein-Volhard, C. The Identification of Genes Controlling Development in Flies and Fishes (Nobel Lecture). *Angew. Chemie (International Ed. English)* **1996**, *35* (19), 2176–2187.

(5) Echelard, Y.; Epstein, D. J.; St-Jacques, B.; Shen, L.; Mohler, J.; McMahon, J. A.; McMahon, A. P. Sonic Hedgehog, a Member of a Family of Putative Signaling Molecules, Is Implicated in Regulation of CNS Polarity. *Cell* **1993**, *75*, 1417–1430.

(6) Chen, C. H.; Von Kessler, D. P.; Park, W.; Wang, B.; Ma, Y.; Beachy, P. A. Nuclear Trafficking of Cubitus Interruptus in the Transcriptional Regulation of Hedgehog Target Gene Expression. *Cell* **1999**, *98* (3), 305–316.

(7) Kinto, N.; Iwamoto, M.; Enomoto-Iwamoto, M.; Noji, S.; Ohuchi, H.; Yoshioka, H.; Kataoka, H.; Wada, Y.; Yuhao, G.; Takahashi, H. E.; Yoshiki, S.; Yamaguchi, A. Fibroblasts Expressing Sonic Hedgehog Induce Osteoblast Differentiation and Ectopic Bone Formation. *FEBS Lett.* **1997**, *404* (2–3), 319–323.

(8) Nakano, Y.; Guerreri, I.; Higalgo, A.; Taylor, A.; Whittle, J.; Ingham, P. W. A Protein with Several Possible Membrane-Spanning Domains Encoded by the Drosophila Segment Polarity Gene Patched. *Nature* **1989**, *341*, 508–513.

(9) Pearse, R. V.; Vogan, K. J.; Tabin, C. J. Ptc1 and Ptc2 Transcripts Provide Distinct Readouts of Hedgehog Signaling Activity during Chick Embryogenesis. *Dev. Biol.* **2001**, *239* (1), 15–29.

(10) Alcedo, J.; Ayzenzon, M.; Von Ohlen, T.; Noll, M.; Hooper, J. E. The Drosophila Smoothened Gene Encodes a Seven-Pass Membrane Protein, a Putative Receptor for the Hedgehog Signal. *Cell* **1996**, *86* (2), 221–232.

(11) Rohatgi, R.; Milenkovic, L.; Corcoran, R. B.; Scott, M. P. Hedgehog Signal Transduction by Smoothened: Pharmacologic Evidence for a 2-Step Activation Process. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (9), 3196–3201.

(12) Kasper, M.; Regl, G.; Frischauf, A. M.; Aberger, F. GLI Transcription Factors: Mediators of Oncogenic Hedgehog Signalling. *Eur. J. Cancer* **2006**, *42* (4), 437–445.

(13) Tukachinsky, H.; Petrov, K.; Watanabe, M.; Salic, A. Mechanism of Inhibition of the Tumor Suppressor Patched by Sonic Hedgehog. *Proc. Natl. Acad. Sci.* **2016**, *113* (40), E5866–E5875.

(14) Huang, P.; Nedelcu, D.; Watanabe, M.; Jao, C.; Kim, Y.; Liu, J.; Salic, A. Cellular

Cholesterol Directly Activates Smoothened in Hedgehog Signaling. *Cell* **2016**, *166*, 1176–1187.

(15) Duman-Scheel, M.; Weng, L.; Xin, S.; Du, W. Hedgehog Regulates Cell Growth and Proliferation by Inducing Cyclin D and Cyclin E. *Nature* **2002**, *417*, 299–304.

(16) Kenney, A. M. Nmyc Upregulation by Sonic Hedgehog Signaling Promotes Proliferation in Developing Cerebellar Granule Neuron Precursors. *Development* **2003**, *130* (1), 15–28.

(17) Karhadkar, S. S.; Bova, G. S.; Abdallah, N.; Dhara, S.; Gardner, D.; Maitra, A.; Isaacs, J. T.; Berman, D. M.; Beachy, P. A. Hedgehog Signalling in Prostate Regeneration, Neoplasia and Metastasis. *Nature* **2004**, *431* (7009), 707–712.

(18) Tenzen, T.; Allen, B. L.; Cole, F.; Kang, J. S.; Krauss, R. S.; McMahon, A. P. The Cell Surface Membrane Proteins Cdo and Boc Are Components and Targets of the Hedgehog Signaling Pathway and Feedback Network in Mice. *Dev. Cell* **2006**, *10* (5), 647–656.

(19) Yao, S.; Lum, L.; Beachy, P. The Ihog Cell-Surface Proteins Bind Hedgehog and Mediate Pathway Activation. *Cell* **2006**, *125* (2), 343–357.

(20) Hanna, A.; Shevde, L. A. Hedgehog Signaling: Modulation of Cancer Properies and Tumor Mircroenvironment. *Mol. Cancer* **2016**, *15* (1), 24.

(21) Varjosalo, M.; Taipale, J. Hedgehog Signaling. *J. Cell Sci.* **2006**, *120* (1), 3–6.

(22) Tukachinsky, H.; Lopez, L. V.; Salic, A. A Mechanism for Vertebrate Hedgehog Signaling: Recruitment to Cilia and Dissociation of SuFu-Gli Protein Complexes. *J. Cell Biol.* **2010**, *191* (2), 415–428.

(23) Chen, X.; Tukachinsky, H.; Huang, C. H.; Jao, C.; Chu, Y. R.; Tang, H. Y.; Mueller, B.; Schulman, S.; Rapoport, T. A.; Salic, A. Processing and Turnover of the Hedgehog Protein in the Endoplasmic Reticulum. *J. Cell Biol.* **2011**, *192* (5), 825–838.

(24) Lewis, P. M.; Dunn, M. P.; McMahon, J. A.; Logan, M.; Martin, J. F.; St-Jacques, B.; McMahon, A. P. Cholesterol Modification of Sonic Hedgehog Is Required for Long-Range Signaling Activity and Effective Modulation of Signaling by Ptc1. *Cell* **2001**, *105* (5), 599–612.

(25) Hall, T.; Porter, J.; Beachy, P. A.; Leahy, D. A Potential Catalytic Site Revealed by the 1.7A Crystal Structure of the Amino-Terminal Signaling Domain of Sonic Hedgehog. *Nature* **1995**, *378*, 212–216.

(26) Williams, K. P.; Rayhorn, P.; Chi-Rosso, G.; Garber, E. A.; Strauch, K. L.; Horan, G. S.; Reilly, J. O.; Baker, D. P.; Taylor, F. R.; Koteliansky, V.; Pepinsky, R. B. Functional Antagonists of Sonic Hedgehog Reveal the Importance of the N Terminus for Activity. *J. Cell Sci.* **1999**, *112*, 4405–4414.

(27) Pepinsky, R. B.; Rayhorn, P.; Day, E. S.; Dergay, A.; Williams, K. P.; Galdes, A.; Taylor, F. R.; Boriack-Sjodin, P. A.; Garber, E. A. Mapping Sonic Hedgehog-Receptor Interactions

by Steric Interference. *J. Biol. Chem.* **2000**, *275* (15), 10995–11001.

(28) Taylor, F. R.; Wen, D.; Garber, E. A.; Carmillo, A. N.; Baker, D. P.; Arduini, R. M.; Williams, K. P.; Weinreb, P. H.; Rayhorn, P.; Hronowski, X.; Whitty, A.; Day, E. S.; Boriack-Sjodin, A.; Shapiro, R. I.; Galdes, A.; Pepinsky, R. B. Enhanced Potency of Human Sonic Hedgehog by Hydrophobic Modification. *Biochemistry* **2001**, *40* (14), 4359–4371.

(29) McMahon, A. P.; Ingham, P. W.; Tabin, C. J. Developmental Roles and Clinical Significance of Hedgehog Signaling. *Curr. Top. Dev. Biol.* **2003**, *53*, 1–114.

(30) Bitgood, M. J.; Shen, L.; McMahon, A. P. Sertoli Cell Signaling by Desert Hedgehog Regulates the Male Germline. *Curr. Biol.* **1996**, *6* (3), 298–304.

(31) Roessler, E.; Belloni, E.; Gaudenz, K.; Jay, P.; Berta, P.; Scherer, S. W.; Tsui, L. C.; Muenke, M. Mutations in the Human Sonic Hedgehog Gene Cause Holoprosencephaly. *Nat. Genet.* **1996**, *14* (3), 357–360.

(32) Rubin, L. L.; de Sauvage, F. J. Targeting the Hedgehog Pathway in Cancer. *Nat. Rev. Drug Discov.* **2006**, *5* (12), 1026–1033.

(33) Johnson, R. L.; Rothman, A. L.; Xie, J.; Goodrich, L. V; Bare, J. W.; Bonifas, J. M.; Quinn, A. G.; Myers, R. M.; Cox, D. R.; Epstein, E. H.; Scott, M. P. Human Homolog of Patched, a Candidate Gene for the Basal Cell Nevus Syndrome. *Science* **1996**, *272* (5268), 1668–1671.

(34) Raffel, C.; Jenkins, R.; Frederick, L.; Hebrink, D.; Alderete, B.; Fults, D.; James, C. Sporadic Medulloblastomas Contain PTCH Mutations1. *Cancer Res.* **1997**, *57*, 842–845.

(35) Zibat, A.; Missiaglia, E.; Rosenberger, A.; Pritchard-Jones, K.; Shipley, J.; Hahn, H.; Fulda, S. Activation of the Hedgehog Pathway Confers a Poor Prognosis in Embryonal and Fusion Gene-Negative Alveolar Rhabdomyosarcoma. *Oncogene* **2010**, *29* (48), 6323–6330.

(36) Ju, B.; Chen, W.; Spitsbergen, J. M.; Lu, J.; Vogel, P.; Peters, J. L.; Wang, Y. D.; Orr, B. A.; Wu, J.; Henson, H. E.; Jia, S.; Parupalli, C.; Taylor, M. R. Activation of Sonic Hedgehog Signaling in Neural Progenitor Cells Promotes Glioma Development in the Zebrafish Optic Pathway. *Oncogenesis* **2014**, *3* (3).

(37) Kasper, M.; Jaks, V.; Fiaschi, M.; Toftgård, R. Hedgehog Signalling in Breast Cancer. *Carcinogenesis* **2009**, *30* (6), 903–911.

(38) Sanchez, P.; Hernández, A. M.; Stecca, B.; Kahler, A. J.; DeGueme, A. M.; Barrett, A.; Beyna, M.; Datta, M. W.; Datta, S.; Ruiz i Altaba, A. Inhibition of Prostate Cancer Proliferation by Interference with SONIC HEDGEHOG-GLI1 Signaling. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (34), 12561–12566.

(39) Ma, X.; Sheng, T.; Zhang, Y.; Zhang, X.; He, J.; Huang, S.; Chen, K.; Sultz, J.; Adegboyega, P. A.; Zhang, H.; Xie, J. Hedgehog Signaling Is Activated in Subsets of Esophageal Cancers. *Int. J. Cancer* **2006**, *118* (1), 139–148.

(40) Lee, S. Y.; Han, H. S.; Lee, K. Y.; Hwang, T. S.; Kim, J. H.; Sung, I. K.; Park, H. S.; Jin, C. J.; Choi, K. W. Sonic Hedgehog Expression in Gastric Cancer and Gastric Adenoma. *Oncol. Rep.* **2007**, *17* (5), 1051–1055.

(41) Morton, J. P.; Mongeau, M. E.; Klimstra, D. S.; Morris, J. P.; Lee, Y. C.; Kawaguchi, Y.; Wright, C. V. E.; Hebrok, M.; Lewis, B. C. Sonic Hedgehog Acts at Multiple Stages during Pancreatic Tumorigenesis. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (12), 5103–5108.

(42) Park, K.-S.; Martelotto, L. G.; Peifer, M.; Sos, M. L.; Karnezis, A. N.; Mahjoub, M. R.; Bernard, K.; Conklin, J. F.; Szczepny, A.; Yuan, J.; Guo, R.; Ospina, B.; Falzon, J.; Bennett, S.; Brown, T. J.; Markovic, A.; Devereux, W. L.; Ocasio, C. A.; Chen, J. K.; Stearns, T.; Thomas, R. K.; Dorsch, M.; Buonamici, S.; Watkins, D. N.; Peacock, C. D.; Sage, J. A Crucial Requirement for Hedgehog Signaling in Small Cell Lung Cancer. *Nat. Med.* **2011**, *17* (11), 1504–1508.

(43) Gorlin, R. J. Nevoid Basal Cell Carcinoma (Gorlin) Syndrome. *Genet. Med.* **2004**, *6* (6), 530–539.

(44) Yauch, R. L.; Gould, S. E.; Scales, S. J.; Tang, T.; Tian, H.; Ahn, C. P.; Marshall, D.; Fu, L.; Januario, T.; Kallop, D.; Nannini-Pepe, M.; Kotkow, K.; Marsters, J. C.; Rubin, L. L.; De Sauvage, F. J. A Paracrine Requirement for Hedgehog Signalling in Cancer. *Nature* **2008**, *455* (7211), 406–410.

(45) Stanton, B. Z.; Peng, L. F. Small-Molecule Modulators of the Sonic Hedgehog Signaling Pathway. *Mol. BioSyst.* **2010**, *6* (1), 44–54.

(46) Tremblay, M. R.; McGovern, K.; Read, M. A.; Castro, A. C. New Developments in the Discovery of Small Molecule Hedgehog Pathway Antagonists. *Curr. Opin. Chem. Biol.* **2010**, *14* (3), 428–435.

(47) Cooper, M. K. Teratogen-Mediated Inhibition of Target Tissue Response to Shh Signaling. *Science (80-.).* **1998**, *280* (5369), 1603–1607.

(48) Chen, J. K.; Taipale, J.; Cooper, M. K.; Beachy, P. a. Inhibition of Hedgehog Signaling by Direct Binding of Cyclopamine to Smoothened Service Inhibition of Hedgehog Signaling by Direct Binding of Cyclopamine to Smoothened. *Genes Dev.* **2002**, No. 410, 2743–2748.

(49) Dlugosz, A.; Agrawal, S.; Kirkpatrick, P. Vismodegib. *Nat. Rev. Drug Discov.* **2012**, *11* (6), 437–438.

(50) Burness, C. B. Idarucizumab: First Global Approval. *Drugs* **2015**, *75* (18), 2155–2161.

(51) Maun, H. R.; Wen, X.; Lingel, A.; De Sauvage, F. J.; Lazarus, R. A.; Scales, S. J.; Hymowitz, S. G. Hedgehog Pathway Antagonist 5E1 Binds Hedgehog at the Pseudo-Active Site. *J. Biol. Chem.* **2010**, *285* (34), 26570–26580.

(52) Ohlig, S.; Farshi, P.; Pickhinke, U.; van den Boom, J.; Höing, S.; Jakuschev, S.; Hoffmann, D.; Dreier, R.; Schöler, H. R.; Dierker, T.; Bordych, C.; Grobe, K. Sonic Hedgehog

Shedding Results in Functional Activation of the Solubilized Protein. *Dev. Cell* **2011**, *20* (6), 764–774.

(53) Stanton, B. Z.; Peng, L. F.; Maloof, N.; Nakai, K.; Wang, X.; Duffner, J. L.; Taveras, K. M.; Hyman, J. M.; Lee, S. W.; Koehler, A. N.; Chen, J. K.; Fox, J. L.; Mandinova, A.; Schreiber, S. L. A Small Molecule That Binds Hedgehog and Blocks Its Signaling in Human Cells. *Nat. Chem. Biol.* **2009**, *5* (3), 154–156.

(54) Lauth, M.; Bergström, Å. Inhibition of GLI-Mediated Transcription and Tumor Cell Growth by Small-Molecule Antagonists. *Proc. …* **2007**, *104* (20), 8455–8460.

(55) Takebe, N.; Harris, P. J.; Warren, R. Q.; Ivy, S. P. Targeting Cancer Stem Cells by Inhibiting Wnt, Notch, and Hedgehog Pathways. *Nat. Rev. Clin. Oncol.* **2011**, *8* (2), 97–106.

(56) Chen, Y. X.; Koch, S.; Uhlenbrock, K.; Weise, K.; Das, D.; Gremer, L.; Brunsveld, L.; Wittinghofer, A.; Winter, R.; Triola, G.; Waldmann, H. Synthesis of the Rheb and K-Ras4B GTPases. *Angew. Chemie - Int. Ed.* **2010**, *49* (35), 6090–6095.

(57) Yang, A.; Li, Y.; Pantoom, S.; Triola, G.; Wu, Y. W. Semisynthetic Lipidated LC3 Protein Mediates Membrane Fusion. *ChemBioChem* **2013**, *14* (11), 1296–1300.

(58) Brunsveld, L.; Kuhlmann, J.; Alexandrov, K.; Wittinghofer, A.; Goody, R. S.; Waldmann, H. Lipidated Ras and Rab Peptides and Proteins - Synthesis, Structure, and Function. *Angew. Chemie - Int. Ed.* **2006**, *45* (40), 6622–6646.

(59) Grogan, M. J.; Kaizuka, Y.; Conrad, R. M.; Groves, J. T.; Bertozzi, C. R. Synthesis of Lipidated Green Fluorescent Protein and Its Incorporation in Supported Lipid Bilayers. *J. Am. Chem. Soc.* **2005**, *127* (41), 14383–14387.

(60) Alexandrov, K.; Heinemann, I.; Durek, T.; Sidorovitch, V.; Goody, R. S.; Waldmann, H. Intein-Mediated Synthesis of Geranylgeranylated Rab7 Protein in Vitro. *J. Am. Chem. Soc.* **2002**, *124* (20), 5648–5649.

(61) Yang, A.; Zhao, L.; Wu, Y. W. Chemical Synthesis and Biological Function of Lipidated Proteins. *Top Curr Chem* **2015**, *362*, 137–182.

(62) Huang, Y. C.; Li, Y. M.; Chen, Y.; Pan, M.; Li, Y. T.; Yu, L.; Guo, Q. X.; Liu, L. Synthesis of Autophagosomal Marker Protein LC3-II under Detergent-Free Conditions. *Angew. Chemie - Int. Ed.* **2013**, *52* (18), 4858–4862.

(63) Tsumoto, K.; Ejima, D.; Kumagai, I.; Arakawa, T. Practical Considerations in Refolding Proteins from Inclusion Bodies. *Protein Expr. Purif.* **2003**, *28* (1), 1–8.

(64) Yamaguchi, H.; Miyazaki, M. Refolding Techniques for Recovering Biologically Active Recombinant Proteins from Inclusion Bodies. *Biomolecules* **2014**, *4* (1), 235–251.

(65) Durek, T.; Alexandrov, K.; Goody, R. S.; Hildebrand, A.; Heinemann, I.; Waldmann, H. Synthesis of Fluorescently Labeled Mono- and Diprenylated Rab7 GTPase. *J. Am. Chem. Soc.* **2004**, *126* (50), 16368–16378.

(66) Lahiri, S.; Brehs, M.; Olschewski, D.; Becker, C. F. W. Total Chemical Synthesis of an Integral Membrane Enzyme: Diacylglycerol Kinase from Escherichia Coli. *Angew. Chemie - Int. Ed.* **2011**, *50* (17), 3988–3992.

(67) Day, E. S.; Wen, D.; Garber, E. a; Hong, J.; Avedissian, L. S.; Rayhorn, P.; Shen, W.; Zeng, C.; Bailey, V. R.; Reilly, J. O.; Roden, J. a; Moore, C. B.; Williams, K. P.; Galdes, a; Whitty, a; Baker, D. P. Zinc-Dependent Structural Stability of Human Sonic Hedgehog. *Biochemistry* **1999**, *38* (45), 14868–14880.

(68) *Hedgehog Signaling Protocols*; JI, H., Ed.; Totowa, New Jersey, 2007.

(69) Fuse, N.; Maiti, T.; Wang, B.; Porter, J. A.; Hall, T. M.; Leahy, D. J.; Beachy, P. A. Sonic Hedgehog Protein Signals Not as a Hydrolytic Enzyme but as an Apparent Ligand for Patched. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (20), 10992–10999.

CHAPTER 3:

Phage Display against Sonic Hedgehog

INTRODUCTION

In 1985 G. Smith described the expression of an exogenous peptide on the surface of filamentous bacteriophage,[1] which later was called **phage display**. This technique was further developed by G. Winter *et al.* in Cambridge[2] and R. Lerner and C. F. Barbas at Scripps.[3] The Winter and Lerner groups reported the display of antibodies variable domains in phages.

Phage display consists on the insertion of a gene that codifies for an exogenous peptide in the bacteriophage DNA, resulting on the display of the fusion peptide on the coat protein of the virus.

The key feature of this technology relies on the linkage between the genotype of the bacteriophage and its phenotype.[1,4] This phenotype-genotype connection guarantees that the selection of a peptide means the simultaneously selection of the DNA codifying for it.[5] In other words, the sequencing of the phage recovered after selection elucidates the peptide sequence selected.

In a first approach, phage display was described for the presentation of peptides but it was soon explored for the selection of proteins, antibody fragments and other scaffolds. It was also reported for the finding of higher affinity binders, the study of protein-protein interactions, the search of better substrates for enzymes or for structural analysis of a target.[5,6]

Phage display is a powerful strategy to find new ligands that bind a specific target. It makes possible the fast screening of large libraries with a diversity of 10^9 to select specific peptide binders in a quick manner.[4] In the present thesis we used phage display technology to find new and better peptide binders against Sonic Hedgehog (ShhN).

1. Methodology

Phage display consists in normally three to four rounds of selections until significant enrichment of the library is reached (**Figure *1***).

In each round of selection, the target of interest is immobilized on a solid support, although solution biopanning have been also reported.[7] Phages expressing the peptides on their surface are then incubated with the target. Afterwards, washings are carried out in order to eliminate those phages that did not present affinity for the target, whereas the ones that bind the target remain attached to the plate. Then, recover of the binding peptides is performed using an acidic solution (0.1 M HCl solution),and after neutralization they are propagated in bacteria cells and used for the next round of selection.[8,9]

When the desired enrichment of the library is achieved, selected phages are isolated and ELISA assays are carried out in order to identify positive binders. Finally, these new binders are sequenced.

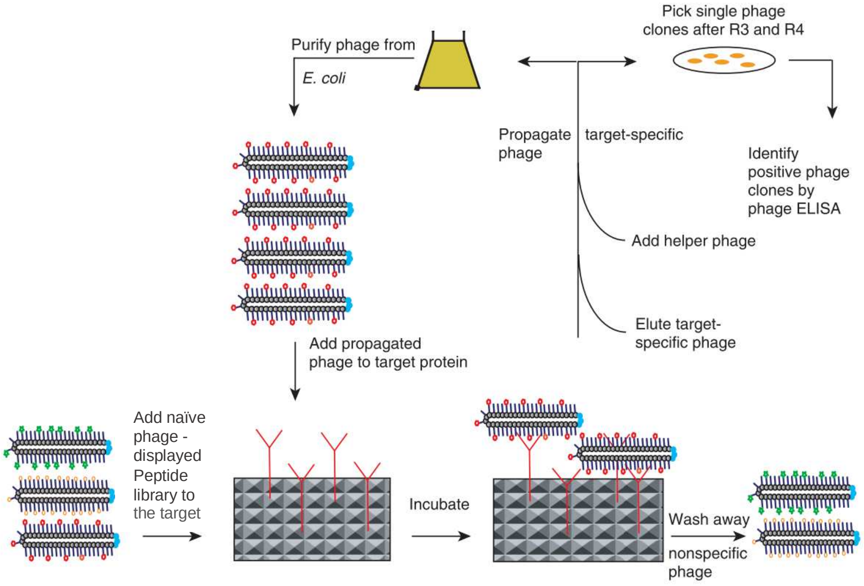

Pick single phage
clones after R3 and R4
Purify phage from
E. coli
Propagate
phage
target-specific
Identify
positive phage
clones by
phage ELISA
Add helper phage
Elute target-
specific phage
Add propagated
phage to target protein
Add naïve
phage -
displayed
Peptide
library to
the target
Incubate
Wash away
nonspecific
phage

Figure 1. Affinity selection process, image adapted from [9]. R = Round of selection.

2. Filamentous bacteriophage biology

The most used systems for phage display are the filamentous bacteriophages: **fd**, **f1** and **M13**. Filamentous bacteriophages are F-pilus dependent-viruses that infect gram-negative bacteria, in particular *E. coli.*[4,10] This type of bacteriophages are no lytic and released by secretion from the bacteria cell. The secretion process implies that, during the phage assembly phage, proteins (wild type and exogenous) will have to cross the inner membrane of the bacteria.[6,11]

These viruses have circular single stranded DNA (ssDNA) packaged into particles with a size encompassing 900 nm long and 6 nm thick.[6,10] The size of the phage particle varies along according the size of its ssDNA.[10]

The ssDNA of filamentous phages has 9 genes that codify for 11 proteins (**Table *1***). Five of the 11 proteins: pIII, pVI, pVII, pVIII and pIX are part of the phage coat and have been used for phage display.[6]

The coat of the bacteriophages consists on one major protein pVIII, and four minor coat proteins: pIII, pIV, pVII and pIX. Major protein pVIII is present in 2700 to 3000 copies per phage while minor proteins consists of five copies of each one[4,10] (**Figure *2***).

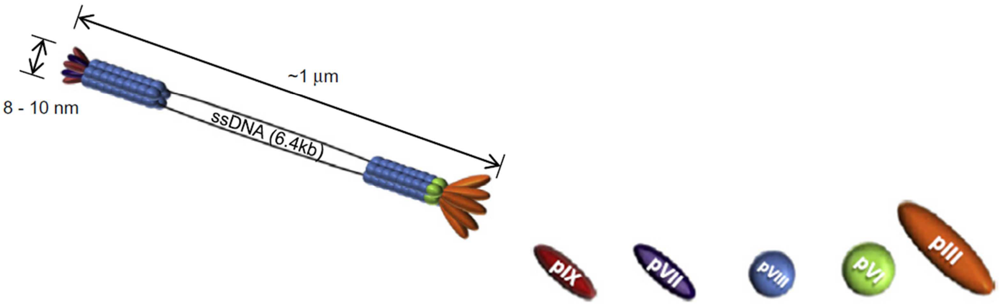
Gene	Protein	Size (aa)	Function	Phage Display
I	I	348	Phage assembly	-
	XI	108	Phage assembly	-
II	II	409	Replication	-
	X	111	Required for phage DNA synthesis	-
III	III	406	Virion component: maintains virus integrity, it is the attachment protein of the phage.	N-terminus display
IV	IV	405	Phage assembly	-
V	V	87	Replication: regulates viral DNA synthesis; protects ssDNA in bacteria cell Phage assembly: scaffold in early stage of assembly	-
VI	VI	112	Virion component: minor coat protein, implied in virus adsorption	C-terminus display
VII	VII	33	Virion component: minor coat protein	N-terminus display
VIII	VIII	50	Virion component: major coat protein. Phage assembly: replaces pV in later stages of phage assembly.	N and C-terminus display
IX	IX	32	Virion component: minor coat proteins	N-terminus display

Table 1. Bacteriophages genes, proteins and functions. Adapted from [6].

In principle, exogenous peptides can be fused through the N or C terminus of the coat proteins. Nevertheless, carboxi terminal systems reported in the literature were very poor in peptide expression when compared to N-terminal displays,[12] which fostered the use of N-terminal phage display above C-terminal. N-terminal phage display can be performed in pIII, pVII,[13] pVIII and pIX,[14] however the most used proteins are pIII and pVIII.[4,6]

In some situations in where N-terminal phage display can not be carried out, C-terminal display represents an alternative. Thus, for instance, display of intracellular proteins needs the exposure of the carboxi terminus for the folding in the bacteria cytoplasm, or study of protein-

protein interactions.[11,12] In the case of cDNA libraries, C-terminal display is mandatory because stop codons prevent N-terminal display.[12] As a consequence, in the recent years more robust C-terminal phage display systems have been developed in proteins pIII, pVI and pVIII.[11,12,15–18]



Number of amino acids:	32	33	50	112	406
Molecular weight (kDa):	3.7	3.6	5.2	12.4	42.5
Copies per virion:	5	5	2,700	5	5
Display mode:	N	N	N or C	C	N or C
N-terminal signal sequence:	No	No	Yes	No	Yes

Figure 2. Structure of filamentous phage. Image adapted from [19].

3. Phage display in protein pVIII

Gene VIII codifies for the major coat protein pVIII, which is expressed in high number of copies on the phage, from 2700 to 3000.[6] Major coat proteins are angled 20° and form an helix with little flexibility due to the hydrophobic side chains of the residues.[10] There are three domains in pVIII: the N-terminus domain is mainly formed by acidic amino acids and exposed in the virus surface.[10] The C-terminus domain consists on basic amino acids that interact with the negative phosphates of the DNA. Finally, the central region is comprised by hydrophobic residues making intertwining interactions.[10]

Fusion peptide sequences displayed on pVIII are inserted between the signal sequence and the beginning of the mature protein encoding sequence (**Figure *3*** b).[6] The high number of exogenous peptides expressed per phage represents an advantage for phage selection, the main drawback when using pVIII for peptide display is the limitation on peptide length. It has been reported that peptides no longer than 6-8 amino acids are well tolerated [20] and attempts for displaying longer peptide residues (10 to 18 amino acids) dramatically affect the structure of pVIII, preventing the proper assembly of the virus.[6,20]

To overcome this problem **phagemid vectors** (see section 5) of pVIII were developed. This alternative allows the display of peptides as long as 50 residues in pVIII.[6] By using phagemid vectors, phage display in pVIII has become the start point for most of the phage display projects because it initially allows the selection of weak binders.[6] Nevertheless, it is important to considerer possible avidity effects that can prevent the selection of tight binders with reduced expression capacity.[4,6] Normal K*d* values of the binders obtained from phage display in pVIII are around 1µM.

pVIII is the preferred display system to start a project from naïve libraries. Later on, in order to mature the peptide-leads and obtain high-affinity binders, they are transferred to pIII vectors.

4. Phage display in protein pIII

Protein pIII, codified by gene pIII, is one of the minor coat proteins. It is expressed in 3-5 copies per phage and are localized at the tip of the virus.[4,6] Protein pIII consists in three domains: two amino terminal domains that are part of the adsorption complex and facilitates the attachment of the phage to the bacteria cell,[6,10,12] and the C-terminal carboxi domain that maintains the protein attached to the phage structure.[12]

Although pIII has a major role on phage infectivity, its tolerance for the display of longer peptides is higher than pVIII.[6,10] However, infectivity can be compromised and phagemid vectors of pIII had been also developed to overcome this limitation.

Phage display using pIII protein is usually used for the selection of high affinity binders,[6] that can achieve a a K*d* on the nM or pM range.[4]

5. Phagemid vectors and helper phage

5.1. Phagemid vectors

Phagemid vectors have become an alternative to normal phage vectors to overcome infectivity problems in phages caused by the display of large peptides. They allow the expression of longer peptides without compromising phage functionality. Phagemids are very stable systems, easy to propagate and offer a better display control in comparison to phage vectors.[21]

Phagemid vectors are circular ssDNA that contain the components required for packaging of the phagemid DNA into other phage particles when the bacteria is infected with **helper phages.**[6,9] Phagemid vectors require helper phages to supply additional proteins needed for infectivity and the generation of new phages displaying the recombinant peptide.[4,6,22,23] The phagemids contain both a dsDNA and ssDNA origin of replication (ori). The dsDNA ori is necessary for the replication of the viral DNA when the phagemid infects the bacteria cell. The ssDNA origin of replication (f1 ori) contains all the DNA sequences needed for the packaging of the viral DNA.[9,24] A marker gene, like β-lactamase gene (Amp^r), is also present in phagemid vectors and confers resistance to carbenicillin and ampiciline. The final component of the phagemid vector is a DNA cassette under control of isopropyl-β-ᴅ-thiogalactopyranoside (IPTG), which encodes for the protein signal sequence fused to the coat protein (**Figure *3***), that is responsible for the secretion of the coat protein. The signal sequence is cleaved by peptidases afterwards.[9,24] The oligonucleotide encoding for the library to be displayed, is inserted in-frame with the gene that encodes for the coat protein in the phagemid plasmid.[9]

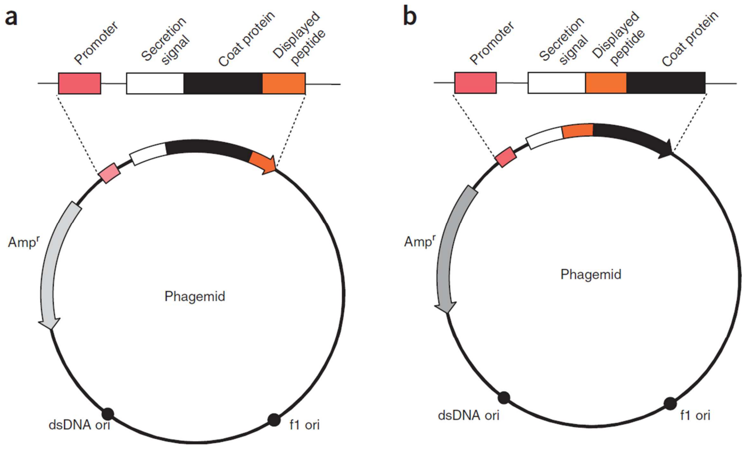


Figure 3. Phagemid vectors for phage display. Peptide library is fused to phagemid vector for the C-terminal (a), or to phagemid vector for N-terminus display (b). Image adapted from [9].

5.2. Helper phage

There are several helper phages described until the moment, although most of them are derived from M13K07, VCSM13 and R408.[25–27]

Helper phages lack the packaging signal, although they contain an additional ori in their DNA and additional genes for assembly.[5,25,26] When a bacteria cell is infected with both the phagemid vector and the helper phage, the phagemid vector will be packaged preferentially because it contains an optimal packaging signal, although packaging of helper phages results in a heterologous population of phages. The displayed peptide may be wild type (from the helper phage) or recombinant (from the phagemid vector), and the packaged DNA can also be the phagemid or the helper phage.[5] Antibiotic resistance genes facilitates the selection of those phages that incorporated the recombinant protein among those which did not.[5]

6. Design and construction of libraries for phage display

As mentioned previously, phage display technology can be used for the display of peptides of different lenghts. In the present thesis, this technology was used for the display of peptides of ~50-60 amino acids. The design and construction of the libraries is different depending on the peptide to display. This design is based on the existence of a natural binder of the target, or the knowledge of the structure of the studied protein. New libraries can be created after optimization of already existing libraries applying tehcniques suchs as nucleotide doping, biased codon selection, or small mutagenesis window.[28,29]

In theory, the variation of a phage peptide library can reach up to 10^{11}, but practical limitations, such as the number of cells transfected, and the number of possible combinations of residues for a peptide, showed that the normal variation oscillates from 10^7 to 10^9.[4] In addition, due to expression difficulties of some sequences, not all the variants are equally represented in the same library.

6.1. Random peptide libraries

Design of random peptide libraries represents an easy approach when no natural ligand of the target protein is known or there is no information about its structure. This strategy is commonly used for starting phage display projects, especially when the library to screen is small. In this case, most of the residues are randomized by the use of degenerated codons. Random libraries are also generated when there is a known natural ligand, which is used as scaffold for mutations, truncations, deletions and other modifications to improve ligand candidates derived from the parental peptide.[4]

There are two approaches for the construction of peptide libraries:

- **unconstrained peptides**: the most common peptide libraries because they are easy to construct and well tolerated for phage display. Peptides expressed do not have any constraint in their structure. Recently, unconstrained peptide libraries that include some sort of chemical modifications have been described. For instance, Derda *et al.* showed the construction of modified phage libraries in which chemical modifications, such as the incorporation of sulfonamide, biotin or manose, are performed by a chemoselective reaction carried out in the peptide displayed on the phage surface (**Figure *4***).[30,31]

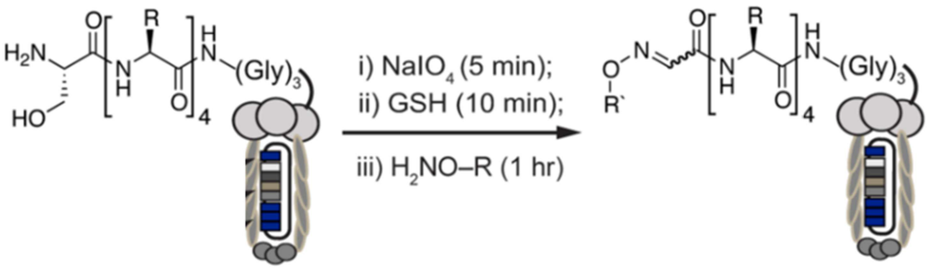


Figure 4. Chemical modification of the phage-displayed peptide library through oxime ligation to an *N*-terminal glyoxal generated by $NAIO_4$ oxidation of N-terminal serine. R = aminoacid side chain; R' = manose, biotin, sulfonamide.[30].

- constrained peptides: peptides expressed in these libraries have disulfide bridges that force them to adopt a defined structure. Constrained peptides present higher binding affinities due to their rigid conformation. Designed constrained peptide libraries contain fixed cysteines (not spaced more than 10 residues), while the remaining amino acids are randomized.[32,33] An interesting recent work has been reported by Derda, who described a methodology to functionalize libraries by macrocyclization of peptides using dichloro-oxime derivatives carrying different carbohydrates.[34] Significantly Heinis *et al* reported phage display bicyclic peptide libraries by derivatization of cysteine residues with dibromobenzyl moieties (**Figure *5***).[35,36]

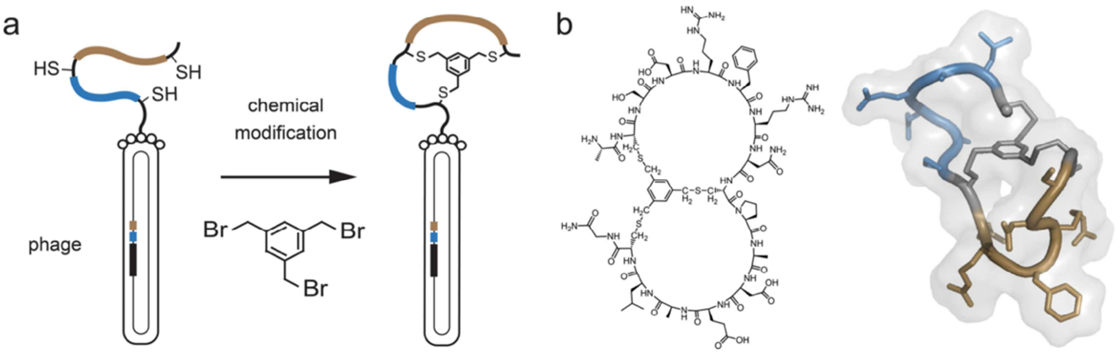


Figure 5. (a) Random peptides containing three fixed Cys residues are displayed on phage and cyclized chemically. (b) Chemical structure and 3D structural model of a byciclic peptide isolated after phage display.[36]

6.2. Rational peptide libraries

Rational design of peptide libraries is performed when there exists known natural ligands that bind the target of interest. These natural binders will provide information to start the rational design. There are two different strategies:

- reduction of protein epitopes: the objective is to identify the essential residues required for the binding to the target. The remaining residues are suppressed, thus creating a reduced epitope that just contain the necessary amino acids for the binding. This strategy was used for minimizing the Z-domain of protein A.[37]

- de novo construction of epitopes: in this case different epitopes from new peptide binders of the target are fused together, creating a new peptide library containing all the binding epitopes discovered. Wrighton *et al* used this approach to identify new peptides that activated the receptor of the erythropoietin cytokine.[29]

7. Affinity maturation

The most common strategy is to start a phage display project from naïve libraries. Then, following initial rounds of phage display screening, the refinement of the selected binders is carried out by affinity maturation in order to obtain tighter ligands. Specific mutations are introduced in selected residues using degenerated codons and site-directed mutagenesis. In addition, switching from polyvalent display (pVIII) to monovalent display (pIII) avoids avidity effects.

Phage display selection using affinity maturation normally render binders with nanomolar affinity.[21]

8. Other display systems

As discussed before, phage display is a powerful technology with a broad range of applications. However, it also faces restrictions, like the expression of post-translational modifications or the dependence of bacteria cells for phage propagation.

In order to explore more areas applying the same display idea, and also with the goal to overcome some of the phage display limitations, other technologies such **bacterial cell display**, **yeast display**, **mRNA** and **ribosome display** were developed in recent years.[4,21] The choice for one or other display system depends on many factors, but mostly on the target of study. Nevertheless, the switch from one display system to another may be useful during the affinity maturation step.

8.1. Bacterial display

In bacterial display peptides are expressed on the surface of gram negative bacteria, mainly *E. Coli.*[38] This display system present a high transformation efficiency, and can expresses 10^5 copies of the fusion peptide.[21] Bacterial display has the advantage of using FACS (Fluorescent Activated Cell Sorting) for quantitative screening, enrichment and calculation of K*d* values.[21] On the other hand, this display system also present some limitations, for instance the peptide length has to be shorter than 60 residues and they have to be fused with specific membrane proteins such surface display carrier proteins.[38] Furthermore, like phage display, it requires a bacterial host which can lead some proteins to incorrect folding or codon bias.[21]

8.2. Yeast display

As an evolution from bacterial cell display, yeast display resumes the same idea but on eukaryotic cells. Due to the fact that yeast cells are eukaryotic, and thus have similar folding pathways and codon usage to mammalian cells, yeast display becomes the choice system for the display of secreted and mammalian cell surface proteins.[21] Like bacterial display it also allows the use of FACS.

8.3. mRNA and ribosome display

Unlike the other display systems described, mRNA and ribosome display do not require the use of biological systems for the display of peptides. These free-cell display systems allow the construction of high diverse libraries, up to 10^{14}, because libraries are not limited to practical considerations like bacterial transformation or peptide secretion.[21,39] Another advantage of these display systems is the recursive mutagenesis occurring after each round of selection, which is a way of affinity maturation.[21] On counterpart, FACS can not be used, and libraries are selected following a procedure similar to phage display.

Recently, remarkable advances using *in vitro* display systems have been reported by H. Suga and co-workers. For instance, they have developed a methodology for the synthesis of libraries containing unnatural *N*-methylated residues.[40] Using this technology, the K*d* values of the binders obtained resulted on the subnanomolar to low nM range.[41] Therefore, the combination of translational machinery, which allows the incorporation of non-natural residues, together with an *in vitro* display technique called RaPID (random nonstandard peptides integrated discovery) results on a promising novel tool for the discovery of new drugs.

OBJECTIVES

Sonic Hedgehog protein is the most important ligand of the Hedgehog signaling pathway. Until the moment, only a two specific binders blocking its interaction with the receptor Ptc, the monoclonal antibody 5E1 and the small molecule robotnikinin, have been reported. In our group, we thought that the phage display technology could lead to the discover of new peptide binders for this protein.

In collaboration with Professor Sachdev Sidhu from the University of Toronto, we were given the opportunity to conduct a research visit in his group under the supervision of Gang Chen. Sachdev Sidhu is an expert in phage display and his research is focused to the generation of synthetic antibodies, structure analysis of protein structures based on phage display, and application of phage display to discover peptide binders for interesting biological targets.

The main objectives of this chapter are:

1. The use of phage display against commercial and recombinant ShhN protein in order to find new peptide binders coming from naïve peptide libraries for this target.

2. The application of phage display against synthetic analogs of ShhN protein to find new ligands for these proteins.

3. The construction and design of a new libraries for affinity maturation from the binders obtained in previous selections.

RESULTS AND DISCUSSION

The work described in this chapter has been carried out in the group of Professor Sachdev Sidhu at the University of Toronto. Professor Sidhu is an expert in phage display, especially in the generation of new antibodies from phage selection. He has also a close collaboration with Reflexion Pharmaceuticals, a company focused on using phage display technology for the development of peptides as novel drugs for the treatment of diseases.

The goal of the project was to apply phage display against ShhN to discover peptides or protein domains (50-60 amino acids) that would bind the protein. Then, synthetize them chemically and study biophysically and biologically in order to determine their capacity to modulate the Hh signaling pathway. Professor Sidhu provided us the libraries and expertise needed for the realization of the project.

1. Library selection to perform phage display against ShhN

As mentioned above, our goal was to discover new peptide binders of ShhN. Bearing this idea and taking into account the structure of our target, seven scaffold libraries from Reflexion Pharmaceuticals were selected: SCF28, SCF32, SCF40, SCF42, SCF53, SCF70 and SCF95 (**Scheme *1***).

Library SCF28
YSEEELKTHI^{10} SKGTLGKFTV^{20} PMLKEACRAY^{30} GLKSGLKKQE^{40} LLEALTKHFQ^{50} D
Library SCF32
TIDQWLLKNA^{10} KEDAIAELKK^{20} AGITSDFYFN^{30} AINKAKTVEE^{40} VNALKNEILK^{50} AHA
Library SCF40
GVEVIAETNV^{10} PLDPNVHQAI^{20} AMVESDDVAP^{30} GNVLGIMQKG^{40} YTLNGRTIRA^{50} AMVTVAKAKA^{60}
Library SCF42
RPKTLFEPGE^{10} MVRVNDGPFA^{20} DFNGVVEEVD^{30} YEKSRLKVSV^{40} SIFGRATPVE^{50} LDFSQVEKA
Library SCF53
GGSEGGEPCA^{10} CPHALHRVCG^{20} SDGETYSNPC^{30} TLNCAKFNG^{40} KPELVKVHDG^{50} PCEPD
Library SCF70
HDNYADLSDT^{10} ELTTLLRRYN^{20} IPHGPVVGST^{30} RRLYEKKIFE^{40} YETQRRR
Library SCF95
VDNKFNKEQQ^{10} NAFYEILHLP^{20} NLNEEQRNAF^{30} IQSLKDDPSQS^{40} ANLLAEAKKL^{50} NDAQAPK

Scheme 1. Amino acid sequence of peptides encoded in SCF libraries selected.

Library SCF28 derived from the chain A of the C-terminal DNA binding domain of human Ku70, a 151-mer protein which folded structure contains three alfa helixes. This library was constructed from the residue 47 to the 97 of the original protein, however, this amino acid sequence was the region that displays the three alpha-helixes. Thereby, the peptides displayed on this library would essentially maintain this folded structure

Library SCF32 was designed from the GA module from *F. magna*. The structure of this 61-mer protein consisted on 3 alpha-helixes and 2 beta turns. Therefore, the peptides from this library would also show this 3D structure.

The nucleotide exchange factor GrpE from *E. Coli* was the scaffold protein selected to construct the **library SCF40**. The whole protein has 197 residues, and its final structure was composed by four alpha-helixes and 6 beta-sheets. To design the library, the residues selected where those displaying the beta-sheets, those from position 138 to 197. Thus, all the peptides from this library would adopt essentially beta-sheets on their tertiary structure.

Library SCF42 came from the transcription termination/antitermination protein NusG from *E. Coli*. This protein consists of 123 amino acids, however, just the residues from position 123 to 181 displayed an structure composed by turns and 5 beta-sheets. The amino acids contained in these positions were the ones used to construct the library SCF42.

The design of **library SCF53** was inspired from Domain-I of the Kazal-type Thrombin inhibitor, a 55-mer protein. In this case, the scaffold for the construction of the library differed in some residues from the original protein. Despite of that, the peptides displayed from this library would have similar folded state like the reference protein. Thereby, in their tertiary structure they would combine an alpha-helix, two small beta sheets and also beta turns.

Library SCF70 was derived from the LEM-domain of the nuclear envelope protein emerin. It was a 47 residue protein which contained 2 alfa-helixes. So, the peptides from this library would also have the helixes on the folded state.

And finally, the scaffold for the construction of the **library SCF95** was protein A, a 59-mer protein which formed three alfa-helixes. Once again, the peptides displayed from this library would be essentially alpha-helixes.

To sum-up, we screened seven libraries in total: four libraries adopted essentially alpha-helixes structure (SCF28, SCF32, SCF70 and SCF95), two of them displayed beta-sheet structure (SCF40 and SCF42) and one was a mixed displaying both an alpha helix and a beta-turn on its structure (SCF53).

These scaffold libraries are encoded in phagemid vectors, with a diversity up to 10^9, and express peptides, from 47 to 60 residues. Although in principle, different aspects such as the structure and the residues from both the target and the peptide scaffold were taken into account, these libraries were completely naïve for display against ShhN. The selected libraries were fused in a pVIII phagemid vector, thereby the peptides were expressed in high number of copies. So, in the initial stages of the selection, we would be able to obtain tight and even weak binders for ShhN.

In the pVIII phagemid vector used[9] the library gene was fused before the major coat protein gene, meaning that the display of the peptide was N-terminal. The gene encoding for the libraries was flanked with a flag tag (SDYKDDDDKGGS) at the N-terminus, necessary for the recognition of the phage by the anti M13 antibody. On the C-terminus it was introduced a linker a spacer between the fusion protein and the phagemid in order to decrease the disruption of the phage assembly due to the expression of bulky peptides or proteins[42] (**Figure *1***). Spacer linkers are flexible and also resistant to proteolysis, and usually their sequence consists on a Gly and Ser rich sequence.

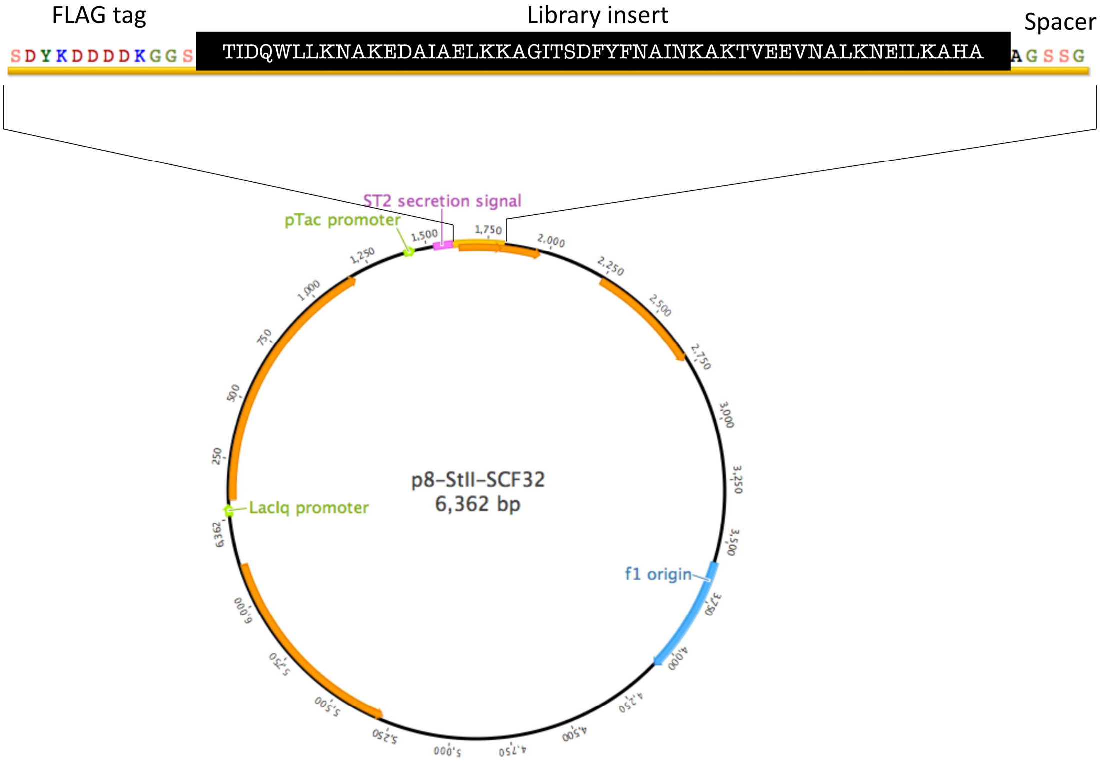

FLAG tag
Library insert
Spacer
SDYKDDDDKGGS
TIDQWLLKNAKEDAIAELKKAGITSDFYFNAINKAKTVEEVNALKNEILKAHA
AGSSG
ST2 secretion signal
pTac promoter
LacIq promoter
f1 origin
p8-StII-SCF32
6,362 bp
250
500
750
1,000
1,250
1,500
1,750
2,000
2,250
2,500
2,750
3,000
3,250
3,500
3,750
4,000
4,250
4,500
4,750
5,000
5,250
5,500
5,750
6,000
6,362

Figure 1. pVIII phagemid vector used from Reflexion Pharmaceuticals. It shows the insertion corresponding to library SCF32, although any library insert would occupy the same position in the plasmid (highlighted in black). Image created with Geneious software.

2. Phage display against commercial ShhN

We decided to start the project using commercial ShhN from Peprotech and then move on to synthetic analog depending on the results obtained.

Commercial ShhN (**Scheme** 2 ***2***) does not contain any lipidic residue attached. They are replaced by hydrophobic residues at the N-terminus in order to mimic the behavior expected of a palmitic. Unlike synthetic analogs commercial ShhN does not have neither biotin residue at the C-terminus.

IVIGPGRGFG^{10} KRRHPKKLTP^{20} LAYKQFIPNV^{30} AEKTLGASGR^{40} YEGKISRNSE^{50}
RFKELTPNYN^{60} PDIIFKDEEN^{70} TGADRLMTQR^{80} CKDKLNALAI^{90} SVMNQWPGVK^{100}
LRVTEGWDED^{110} GHHSEESLHY^{120} EGRALDITTS^{130} DRDRSKYGML^{140} ARLAVEAGFD^{150}
WVYYESKAHI^{160} HCSVKAENSV^{170} AAKSGG^{176}

Scheme 2. Amino acid sequence of commercial ShhN.

2.1. Biopanning against commercial ShhN

Commercial ShhN was reconstituted using water to a concentration of 1mg/mL. Direct coating of the protein onto the microtiter plates was performed at a concentration of 5µg/mL in PBS buffer. After blocking the plate with BSA (Bovine Serum Albumin), the target was incubated with the libraries SCF28, SCF32, SCF40, SCF42, SCF53, SCF70 and SCF95. For the first round 8 wells per library were used, and the concentration of the antigen for coating was 5 µg/mL. In the second round of selection the number of wells used remained the same but the coating concentration of the antigen was decreased to 2 µg/mL. In the following rounds the concentration of antigen for coating was maintained at 2 µg/mL, but the number of wells used were 4 for 3rd round and 2 for the last round. The objective of reducing both the number of wells and the target concentration was to select the tighter binders.

After first round of selection we did not observe enrichment of any library. This result could be expected, because after first rounds of phage selections with naïve libraries most of the phages will not bind the target, therefore very few would be selected. However, this trend would change as one move on with the selections, thus observing more enrichment on the last rounds due to the propagation of phages that bind to the target. Nevertheless, the important goal on projects starting from naïve libraries is to obtain at least few real binders which could undergo affinity maturation.

Fortunately, after the second round, smooth enrichments of the libraries were detected in titration plates. Finally, higher enrichments were visible for the last rounds in most of the libraries when compared to the control plates.

After plating selected dilutions[c] from each library, picked clones were amplified and tested by ELISA. For ELISA assays, the concentration of the antigen directly coated was 2 µg/mL, and the supernatants containing the amplified phages were diluted 4-fold. In ELISA clones that showed a ratio value (Abs at 450 nm/ control well at 450 nm) equal or superior to 5 are considered positive. With this cut-off value, we obtained 18 positive clones from library SCF32, 14 from SCF40 and 86 from library SCF42. On the contrary, libraries SCF28, SCF53, SCF70 or SCF95 did not give any positive clone (**Table** 1 ***1***), suggesting that the peptide sequence that they displayed cannot bind to ShhN. In order to identify real binders for the target, all the positive clones obtained in this first ELISA assay were tested again in a second ELISA. After this second test 17 positive clones out of the first 18 positives from SCF32 were obtained; all 14 clones from SCF40 and the 86 ones from SCF42 showed positive signal again. All the positive clones after the second ELISA assay were sequenced.

Library	Positives clones	Confirmed positives	Unique clones
SCF28	0	-	-
SCF32	18	17	7
SCF40	14	14	Aligned with SCF42
SCF42	86	86	24
SCF53	0	-	-
SCF70	0	-	-
SCF95	0	-	-

Table 1. ELISA results from phage display against commercial ShhN.

With the sequencing results in hand, unique clones could be identified. From library SCF32 7 unique clones were obtained; whereas for library SCF40 we could not obtain unique clones since they did not align with the reference sequence of SCF40 library but do with the reference sequence from library SCF42. This cross-information could be due to phage contamination from one library to the other at some point during phage selections[d]. It can be assumed that clones from SCF42 had higher affinity for ShhN than those from SCF40, and so when contamination occurred the ones selected were those binding tighter to the target. Therefore, all the clones coming from SCF40 were considered from SCF42. Thus, for library SCF42 we identified 24 unique clones after merging the ones from SCF42 and SCF40 (**Table** 1 ***1***).

---

[c] *The selected dilutions for plating would be those which ideally show 5 to 15 colonies on the titration plates after the last biopanning round.*

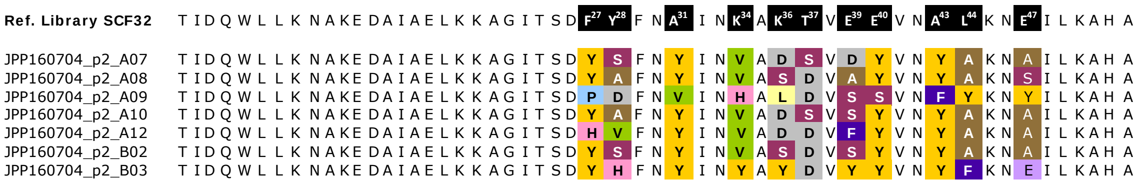
[d] *It is important to remark that it is really common to have undesired contaminations when working with phages. Ideally, biopanning rounds for different libraries should be carried out on different plates. However, in practice, to avoid working with several plates at the same time, the biopanning selections are performed on the same plate. Nevertheless, this practice increases the risk of contaminations from one library to the other. Since the clones displaying best affinities for the target are the ones selected and amplified, when a contamination coming from a library with tighter binders occur, these clones are going to be preferentially selected among those from the library with poor affinity binders.*

2.2. Analysis of unique clones obtained

The goal of analyzing the sequences obtained for unique clones is to find similarities and differences between the residues at the same positions. This could provide information about the positions that play an important role on binding, because would maintain similar residues characteristics, whereas the non-critical positions would present very random mutations. Besides that, it is possible to envisage which type of interaction take place between the binder and the target when a certain category of amino acids prevails in a specific position. Importantly, by comparsion of the variability between the parental sequence and the binders it is possible to draw an idea of the mutations required for binding.

2.2.1. Analysis of unique clones obtained from library SCF32 for commercial ShhN

Only seven unique clones were obtained from library SCF32 (**Figure *2***).



Ref. Library SCF32 T I D Q W L L K N A K E D A I A E L K K A G I T S D F^{27} Y^{28} F N A^{31} I N K^{34} A K^{36} T^{37} V E^{39} E^{40} V N A^{43} L^{44} K N E^{47} I L K A H A

JPP160704_p2_A07 T I D Q W L L K N A K E D A I A E L K K A G I T S D Y S F N Y I N V A D S V D Y V N Y A K N A I L K A H A
JPP160704_p2_A08 T I D Q W L L K N A K E D A I A E L K K A G I T S D Y A F N Y I N V A S D V A Y V N Y A K N S I L K A H A
JPP160704_p2_A09 T I D Q W L L K N A K E D A I A E L K K A G I T S D P D F N V I N H A L D V S S V N F Y K N Y I L K A H A
JPP160704_p2_A10 T I D Q W L L K N A K E D A I A E L K K A G I T S D Y A F N Y I N V A D S V S Y V N Y A K N A I L K A H A
JPP160704_p2_A12 T I D Q W L L K N A K E D A I A E L K K A G I T S D H V F N Y I N V A D D V F Y V N Y A K N A I L K A H A
JPP160704_p2_B02 T I D Q W L L K N A K E D A I A E L K K A G I T S D Y S F N Y I N V A S D V S Y V N Y A K N A I L K A H A
JPP160704_p2_B03 T I D Q W L L K N A K E D A I A E L K K A G I T S D Y H F N Y I N Y A Y D V Y Y V N Y F K N E I L K A H A

Figure 2. Unique clones sequences from SCF32 after phage display to commercial ShhN.

With the sequences from the unique clones we analyzed them to find common mutations among them or a trend of a specific type of residues in concrete positions (**Figure *3***). To start with, in position 27 where a Phe is present in the parental sequence most of the mutations resulted on a Tyr, which is also an aromatic amino acid. Contrary, in position 28 where a Tyr is the wild type residue, we could not determine any similarity between the different clones. This position was mutated into a wide variety of amino acids.

Unlike previously position, Ala^{31} was mutated into a Tyr residue in almost all clones sequenced, which indicated an important role of this aromatic amino acid in binding to ShhN.

Positions 34, 36 and 37 showed a conserved variability. Wild type Arg^{34} was mutated into Val in the majority of the binders sequenced. Arg^{36} was changed by an Asp in almost all peptides. Thr^{37} was principally substituted by an Asp. The area grouping all these residues seems to play an important role in binding, as consensus mutations in these positions were observed among all the ligands obtained. Likely, ionic interactions predominate between positive residues in the ShhN binding region and negatively charged.

In contrast, the Glu[39] was mutated into various amino acids, with no similarity.

Glu[40] and Ala[43] were mostly mutated into Tyr, which seems to indicate that may be required for π-stacking or Van der Waals interactions with the target.

Finally, Leu[44] and Glu[47] were mutated into Ala in most of the binders suggesting the need for a small and hydrophobic amino acid in this region.

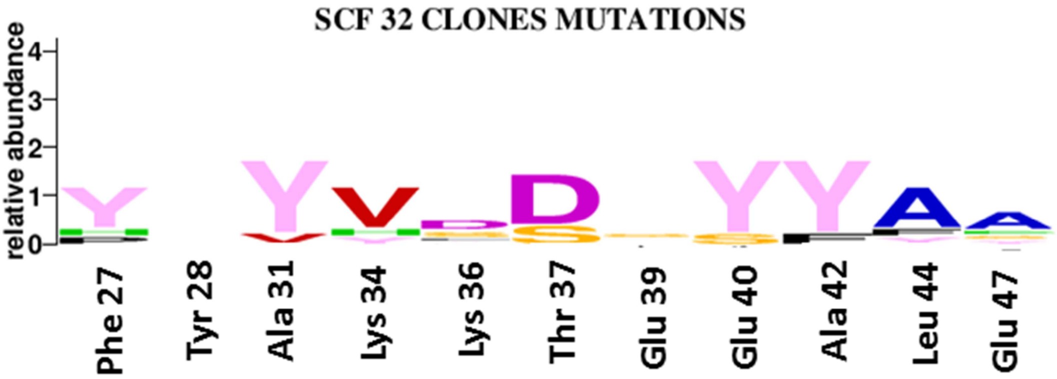


Figure 3. Amino acid composition of the unique clones from SCF32 after phage display to commercial ShhN. Graphic created with WebLogo.

2.2.2. *Analysis of the unique clones obtained from library SCF42 for commercial ShhN*

Asp[17] was mutated mostly to Ser and His residues. Gly[18] was principally changed by Ser, pointing out the importance of having a polar residue at this position. Next to it, Pro[19] was mutated to Tyr and Phe, being both of them aromatic residues, suggesting that the type of interactions with the target carried out by this sort of amino acids, π-stacking and Van der Waals, are important for the binding.

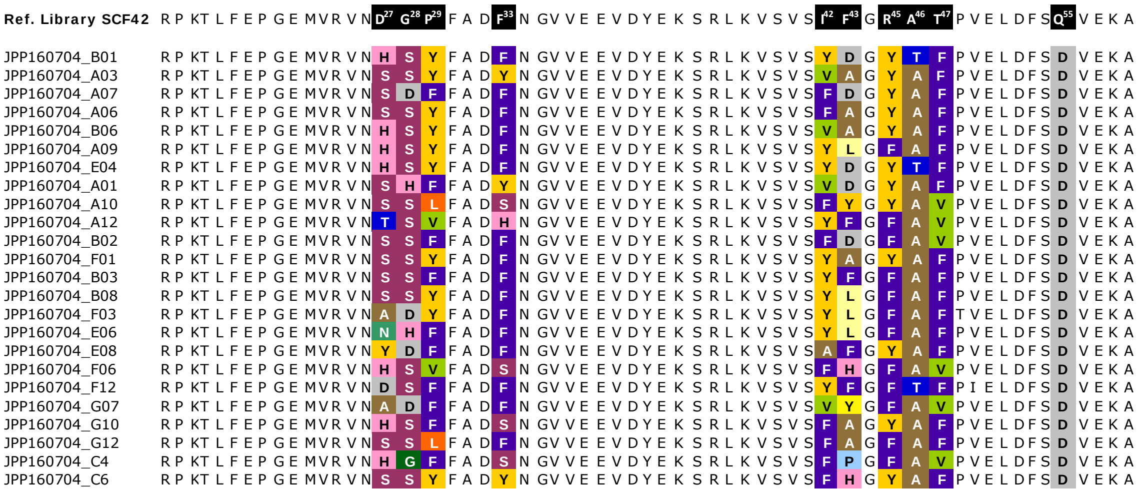
Position 23 (Phe[23]) maintains the wild type Phe or is substituted by Tyr and Ser. Similarly, Ile[42] was predominantly mutated to the aromatic Tyr and Phe.

Mutations in Phe[43] did not show consensus in the different residues. Thus, this position does not appear to play an important role in the binding.

Arg[45] was changed to either a Tyr or a Phe, highlighting the importance of aromatic interactions with the target in this specific position.

In the last region, wild type Ala^{46} was mostly maintained, and just in a few clones was substituted by Thr. The final position Thr^{47}, was changed into either a Phe or Val, both bulky and hydrophobic amino acids.

Analysis of the binders obtained from SCF42 against ShhN (**Figure *4*** and **Figure**), identified a conserved mutation in all the unique clones. Wild type Gln^{55} mutated into an Asp, indicating a critical residue for binding to ShhN.



Ref. Library SCF42	RPKTLFEPGEMVRVN D^{27} G^{28} P^{29} FAD F^{33} NGVVEEVDYEKSRLKVSVS I^{42} F^{43} G R^{45} A^{46} T^{47} PVELDFS Q^{55} VEKA
JPP160704_B01	RPKTLFEPGEMVRVN H S Y FAD F NGVVEEVDYEKSRLKVSVS Y D G Y T F PVELDFS D VEKA
JPP160704_A03	RPKTLFEPGEMVRVN S S Y FAD Y NGVVEEVDYEKSRLKVSVS V A G Y A F PVELDFS D VEKA
JPP160704_A07	RPKTLFEPGEMVRVN S D F FAD F NGVVEEVDYEKSRLKVSVS F D G Y A F PVELDFS D VEKA
JPP160704_A06	RPKTLFEPGEMVRVN S S Y FAD F NGVVEEVDYEKSRLKVSVS F A G Y A F PVELDFS D VEKA
JPP160704_B06	RPKTLFEPGEMVRVN H S Y FAD F NGVVEEVDYEKSRLKVSVS V A G Y A F PVELDFS D VEKA
JPP160704_A09	RPKTLFEPGEMVRVN H S Y FAD F NGVVEEVDYEKSRLKVSVS Y L G F A F PVELDFS D VEKA
JPP160704_E04	RPKTLFEPGEMVRVN H S Y FAD F NGVVEEVDYEKSRLKVSVS Y D G Y T F PVELDFS D VEKA
JPP160704_A01	RPKTLFEPGEMVRVN S H F FAD Y NGVVEEVDYEKSRLKVSVS V D G Y A F PVELDFS D VEKA
JPP160704_A10	RPKTLFEPGEMVRVN S S L FAD S NGVVEEVDYEKSRLKVSVS F Y G Y A V PVELDFS D VEKA
JPP160704_A12	RPKTLFEPGEMVRVN T S V FAD H NGVVEEVDYEKSRLKVSVS Y F G F A V PVELDFS D VEKA
JPP160704_B02	RPKTLFEPGEMVRVN S S F FAD F NGVVEEVDYEKSRLKVSVS F D G F A V PVELDFS D VEKA
JPP160704_F01	RPKTLFEPGEMVRVN S S Y FAD F NGVVEEVDYEKSRLKVSVS Y A G Y A F PVELDFS D VEKA
JPP160704_B03	RPKTLFEPGEMVRVN S S F FAD F NGVVEEVDYEKSRLKVSVS Y F G F A F PVELDFS D VEKA
JPP160704_B08	RPKTLFEPGEMVRVN S S Y FAD F NGVVEEVDYEKSRLKVSVS Y L G F A F PVELDFS D VEKA
JPP160704_F03	RPKTLFEPGEMVRVN A D Y FAD F NGVVEEVDYEKSRLKVSVS Y L G F A F TVELDFS D VEKA
JPP160704_E06	RPKTLFEPGEMVRVN N H F FAD F NGVVEEVDYEKSRLKVSVS Y L G F A F PVELDFS D VEKA
JPP160704_E08	RPKTLFEPGEMVRVN Y D F FAD F NGVVEEVDYEKSRLKVSVS A F G Y A F PVELDFS D VEKA
JPP160704_F06	RPKTLFEPGEMVRVN H S V FAD S NGVVEEVDYEKSRLKVSVS F H G F A V PVELDFS D VEKA
JPP160704_F12	RPKTLFEPGEMVRVN D S F FAD F NGVVEEVDYEKSRLKVSVS Y F G F T F PIELDFS D VEKA
JPP160704_G07	RPKTLFEPGEMVRVN A D F FAD F NGVVEEVDYEKSRLKVSVS V Y G F A V PVELDFS D VEKA
JPP160704_G10	RPKTLFEPGEMVRVN H S F FAD S NGVVEEVDYEKSRLKVSVS F A G Y A F PVELDFS D VEKA
JPP160704_G12	RPKTLFEPGEMVRVN S S L FAD F NGVVEEVDYEKSRLKVSVS F A G F A F PVELDFS D VEKA
JPP160704_C4	RPKTLFEPGEMVRVN H G F FAD S NGVVEEVDYEKSRLKVSVS F P G F A V PVELDFS D VEKA
JPP160704_C6	RPKTLFEPGEMVRVN S S Y FAD Y NGVVEEVDYEKSRLKVSVS F H G Y A F PVELDFS D VEKA

Figure 4. Unique clones sequences from SCF42 obtained after phage display to commercial ShhN.

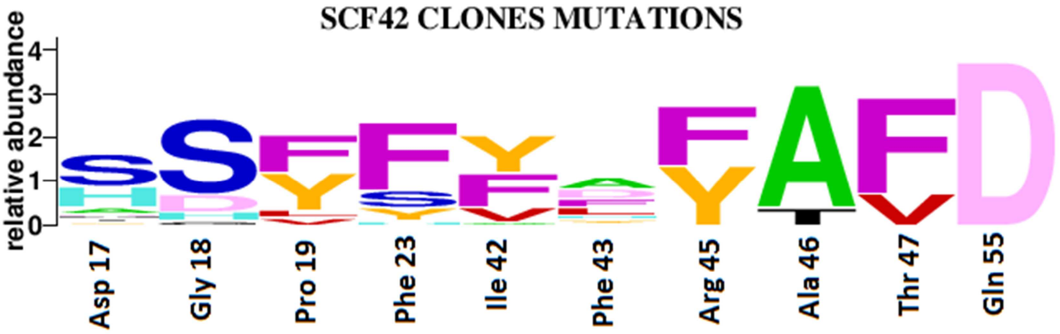


Figure5. Amino acid composition of the unique clones from SCF42 after phage display against commercial ShhN. Graphic created with WebLogo.

3. Phage display against N-terminal biotinylated ShhN

Recombinant ShhN was obtained from expression in *E. Coli.* This analog of ShhN contains a sortase tag (LPSTG sequence) at the C-terminus for site-specific labelling. It is known that, the c-terminal side of ShhN is not involved in the interaction to its natural receptor Patched. Thus, this small tag at the C-terminus is not expected to affect its activity or to interfere with the binding of the peptides expressed on phage surface.[43] In addition, it does not have a palmitic at the N-terminus. Instead of the palmitoylation, the protein was biotinylated for site-specific immobilization on the plate (**Scheme *3***).

Biotin-PEG-AGPGRGFGKR^{10} RHPKKLTPLA^{20} YKQFIPNVAE^{30} KTLGASGRYE^{40}
GKISRNSERF^{50} KELTPNYNPD^{60} IIFKDEENTG^{70} ADRLMTQRCK^{80}
DKLNALAISV^{90} MNQWPGVKLR^{100} VTEGWDEDGH^{110} HSEESLHYEG^{120}
RALDITTSDR^{130} DRSKYGMLAR^{140} LAVEAGFDWV^{150} YYESKAHIHC^{160}
SVKAENSVAA^{170} KSGGLPSTGG^{180}

Scheme 3. Aminoacid sequence of N-terminus biotinylated ShhN.

3.1. Biopanning against N-terminal biotinylated ShhN

We decided to screen the seven libraries previously selected: SCF28, SCF32, SCF40, SCF42, SCF53, SCF70 and SCF95.

The protein was reconstituted in water to a concentration of 1 µg/mL. Unlike the commercial ShhN, immobilization of this protein requires pre-coating of the plates with neutravidin at 5 µg/mL for either phage selections or ELISA assays. After blocking the plates with BSA, the N-terminal biotinylated protein was attached to the plate through a neutravidin-biotin complex at a concentration of 300 nM.

Four rounds of phage selection were performed. For the first and second round eight wells were used per library, whereas in the last two rounds it was decreased to four wells. However, the coating concentration was maintained at 300 nM of target protein during all the biopanning rounds. After the selections, high enrichment was observed specially from libraries SCF32, SCF40 and SCF42, while the other libraries showed less enrichment. Next, the selected dilutions were plated, phages grown and tested in ELISA in order to identify positive clones. The first ELISA assay showed 5 positive clones from SCF28, 36 positive binders from SCF32, 33 from SCF40, 62 from SCF42, 3 from SCF53 and 2 positive binders from SCF70. Unfortunately, no positive binder was recovered from SCF95.

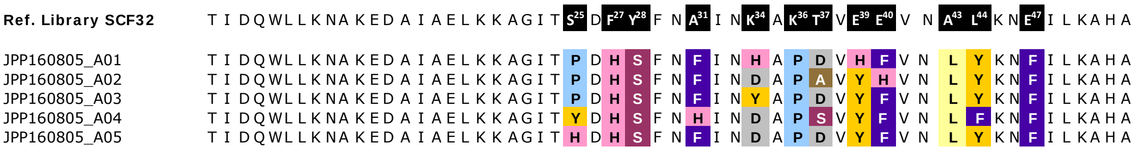
Following a second ELISA, the results confirmed 4 positive clones from SCF28, all 36 from SCF32, 31 from SCF40, 51 positives from SCF42, 3 from SCF53 and only 1 positive from SCF95. The confirmed positive clones were sequenced (**Table *2***).

Careful analysis showed that clones from SCF28, SCF40, SCF42 and SCF53 could not be aligned with their respective library of reference, but did with the reference sequence SCF32.

We again suspected from a cross-contamination, this time coming from library SCF32. We proceeded to align all the sequences obtained from these libraries and also those from SCF32 with the reference sequence from SCF32. After sorting all the sequences 5 unique clones were identified (**Figure *6***). We figured that binders from SCF32 had higher affinity for the antigen than peptides displayed from other libraries, so, when binders from SCF32 contaminated other libraries they were rapidly selected above the ones from the original library. Besides, because contamination may be due to few clones, when the outputs were amplified there existed very low diversity which may explain the few unique clones finally identified.

Library	Positive clones	Confirmed positives	Unique clones
SCF28	5	4	Moved to SCF32
SCF32	36	36	5
SCF40	33	31	Moved to SCF32
SCF42	62	51	Moved to SCF32
SCF53	3	3	Moved to SCF32
SCF70	2	1	Moved to SCF32
SCF95	0	-	-

Table 2. ELISA results from phage display against N-terminal biotinylated ShhN.



Ref. Library SCF32 T I D Q W L L K N A K E D A I A E L K K A G I T S^{25} D F^{27} Y^{28} F N A^{31} I N K^{34} A K^{36} T^{37} V E^{39} E^{40} V N A^{43} L^{44} K N E^{47} I L K A H A

JPP160805_A01 T I D Q W L L K N A K E D A I A E L K K A G I T P D H S F N F I N H A P D V H F V N L Y K N F I L K A H A
JPP160805_A02 T I D Q W L L K N A K E D A I A E L K K A G I T P D H S F N F I N D A P A V Y H V N L Y K N F I L K A H A
JPP160805_A03 T I D Q W L L K N A K E D A I A E L K K A G I T P D H S F N F I N Y A P D V Y F V N L Y K N F I L K A H A
JPP160805_A04 T I D Q W L L K N A K E D A I A E L K K A G I T Y D H S F N H I N D A P S V Y F V N L F K N F I L K A H A
JPP160805_A05 T I D Q W L L K N A K E D A I A E L K K A G I T H D H S F N F I N D A P D V Y F V N L Y K N F I L K A H A

Figure 6. Unique clones sequences from SCF32 obtained for N-terminal biotinylated ShhN.

3.1.1. Analysis of the unique clones obtained from library SCF32 for N-terminal biotinylated ShhN

Despite of having few clones, we can see at first sight that most of the positions susceptible to change showed the same mutation in most of the clones sequenced (**Figure *6* and Figure *7***).

To start with, Ser^{25}, showed different mutations, but Pro is the predominant. It may suggest rigidity in this region that guides the binding with the target.

Phe^{27} was substituted for a His in all selected clones. The mutation into His may not elucidate the type of interaction occurring between the peptide and the target because His can play different roles in protein-protein interaction: aromatic π-motif or as a cation when it is protonated, morevover it can be found coordinating with metals.[44] However, because the mutation is replacing a wild type Phe which could also interact by π-stacking, it is envisaged that the main role of His in this position is cationic, and therefore stablishing ionic interactions with the target.

Another consistent mutation in all the unique clones is observed for Tyr^{28}, which is substituted for a Ser. Ser may interact with other polar residues present on the target.

Following with the sequence analysis, Ala^{31} and Glu^{40} were mostly mutated into Phe and His. Similar behavior was observed in Glu^{39} which was predominantly changed for a Tyr or His. These observations suggested that the presence of aromatic residues at these locations may help to bind to the antigen.

Lys^{34} was changed mainly for Asp, similar to the behavior observed in Thr^{37}.

In contrast, Lys^{36} was mutated exclusively into Pro. Introduction of Pro constrains the peptide conformation and points the residues in the correct orientation for binding.

Ala^{43} was completely mutated into Leu, thus increasing the hydrophobicity in this region.

Leu^{44} was substituted for aromatic residues, either Phe or Tyr. Also, Glu^{47} was exchanged for Phe in all clones, indicating like in previous residue, π-stacking type interactions.

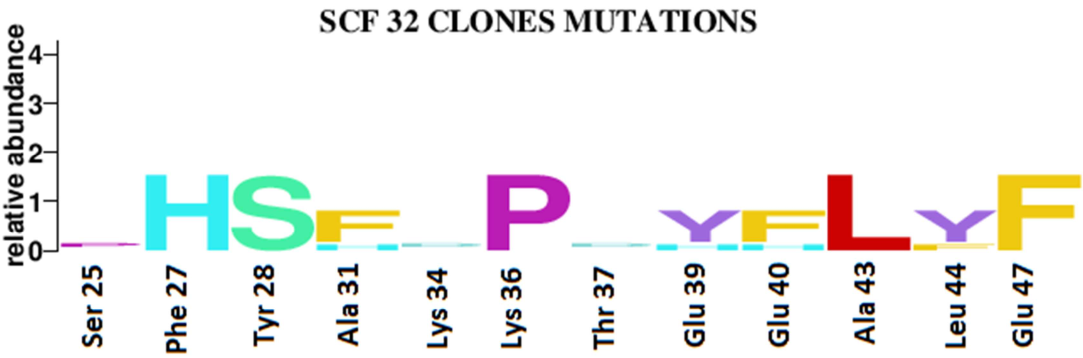


Figure 7. Amino acid composition of the unique clones from SCF32 after phage display against N-terminal biotinylated ShhN. Graphic created with WebLogo.

4. Phage display against synthetic Palm-ShhN-Biotin analog

Our main objective was to perform phage display against the synthetic analog Palm-ShhN-Biotin, which bears a palmitic residue attached at the N-terminus and also contained a biotin residue at the C-terminus (**Scheme *4***). In principle, this synthetic analog has a more similar structure to the natural ShhN, which also harbors a palmitic acid at the N-terminus part. So, we hypothesized that the possible binders for the synthetic ShhN analog would also recognize the natural ShhN.

Palm-AGPGRGFGKR^{10} RHPKKLTPLA^{20} YKQFIPNVAE^{30} KTLGASGRYE^{40} GKISRNSERF^{50}
KELTPNYNPD^{60} IIFKDEENTG^{70} ADRLMTQRCK^{80} DKLNALAISV^{90} MNQWPGVKLR^{100}
VTEGWDEDGH^{110} HSEESLHYEG^{120} RALDITTSDR^{130} DRSKYGMLAR^{140} LAVEAGFDWV^{150}
YYESKAHIHC^{160} SVKAENSVAA^{170} KSGGK-PEG-Biotin

Scheme 4. Amino acid sequence of synthetic Palm-ShhN-Biotin

4.1. Biopanning against Palm-ShhN-Biotin

For biopanning against synthetic Palm-ShhN-Biotin, the plates were pre-coated with neutravidin at 5 µg/mL and next the antigen was coated at 300 nM. The libraries used for screening were SCF32 and SCF42, selected based on the previous results with the commercial ShhN and the N-terminal biotinlylated ShhN.

Four total rounds of phage selection were carried out. For the first and second rounds the coating of the protein was performed at 300 nM, whereas for the third round it was decreased to 100 nM and to 40 nM for the last round. Eight wells were used for each library during all rounds of selections.

After the first round of biopanning no remarkable enrichment of any of the libraries was observed. As expected, in next rounds gradually enrichment was observed. Very importantly, the goal was to obtain at least a few real binders that could undergo next affinity maturation step.

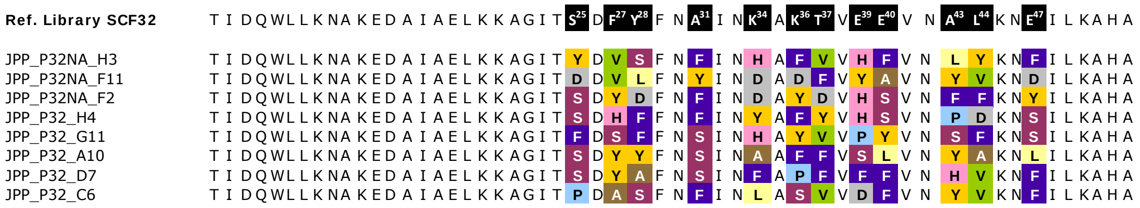
In our case, after the second round of biopanning, enrichment was seen for both libraries SCF32 and SCF42. It was expected even higher enrichment on the next third and fourth rounds, but when comparing with the control plates, it was not possible to determine any enrichment. This failure could be due to the selection of unspecific binders, i. e., ligands that bind either to the neutravidin or the biotin. Therefore, in order to avoid this unwilled selection, the incubation of the phages in the presence of free neutravidin could help on avoiding the selection of unspecific binders.

We decided to adopt this strategy and repeat the third and last round of selections. Thus, neutravidin at a concentration of 5 µg/mL was incubated with the phages for one hour upon their incubation with the target. Following this protocol, enrichment was observed when comparing to the control plates. Next, the selected dilutions[1] were the colonies grown. Those positive clone binders were identified by ELISA: from SCF32 were obtained 27, while 45 came from SCF42. A second ELISA was run giving 20 positives out of the 27 from SCF32 and 8 out of the 45 from SCF42 (**Table** ***3***).

Eight unique clones were identified from SCF32, whereas those from SCF42 displayed stop codons in the middle of their sequence, meaning that they were not specific binders (**Figure *8***).

Library	Positive clones	Confirmed positives	Unique clones
SCF32	27	20	8
SCF42	45	8	0

Table 3. ELISA results from phage display against folded Palm-ShhN-Biotin.



Ref. Library SCF32	T I D Q W L L K N A K E D A I A E L K K A G I T S^{25} D F^{27} Y^{28} F N A^{31} I N K^{34} A K^{36} T^{37} V E^{39} E^{40} V N A^{43} L^{44} K N E^{47} I L K A H A
JPP_P32NA_H3	T I D Q W L L K N A K E D A I A E L K K A G I T Y D V S F N F I N H A F V V H F V N L Y K N F I L K A H A
JPP_P32NA_F11	T I D Q W L L K N A K E D A I A E L K K A G I T D D V L F N Y I N D A D F V Y A V N Y V K N D I L K A H A
JPP_P32NA_F2	T I D Q W L L K N A K E D A I A E L K K A G I T S D Y D F N F I N D A Y D V H S V N F F K N Y I L K A H A
JPP_P32_H4	T I D Q W L L K N A K E D A I A E L K K A G I T S D H F F N F I N Y A F Y V H S V N P D K N S I L K A H A
JPP_P32_G11	T I D Q W L L K N A K E D A I A E L K K A G I T F D S F F N S I N H A Y V V P Y V N S F K N S I L K A H A
JPP_P32_A10	T I D Q W L L K N A K E D A I A E L K K A G I T S D Y Y F N S I N A A F F V S L V N Y A K N L I L K A H A
JPP_P32_D7	T I D Q W L L K N A K E D A I A E L K K A G I T S D Y A F N S I N F A P F V F F V N H V K N F I L K A H A
JPP_P32_C6	T I D Q W L L K N A K E D A I A E L K K A G I T P D A S F N F I N L A S V V D F V N Y V K N F I L K A H A

Figure 8. Unique clone sequences from SCF32 for Palm-ShhN-Biotin.

4.1.1. Analysis of the unique clones obtained from library SCF32 for Palm-ShhN-Biotin

Analysis of the sequences of the unique clones showed no conserved. Unlike the unique clones obtained for N-terminal biotinylated ShhN or those for the commercial ShhN, the binders obtained for Palm-ShhN-Biotin presented more variability between them (**Figure *9***).

To start with, wild type Ser^{25} was mostly conserved. Following on the analysis, Phe^{27} was mutated mostly into Tyr or Val, but also into His, Ser or Ala. Next to it, Tyr^{28} was changed into a wide range of residues: from aromatics to hydrophobic or even acidic. The same trend was detected in wild type Lys^{34}. Due to the wide variety on the residues at these positions among the unique clones, we could not envisage any constant type of interaction with the target.

Ala^{31} was mostly mutated into aromatic residues, suggesting a need for hydrophobic residues to interact with the target.

In Lys^{36}, most mutations codified for aromatic residues like Tyr and Phe. This trend was also observed in Thr^{37}, where mutations into aromatic or hydrophobic residues were the most abundant. This fact suggested that this sort of residues stablished a similar type of interactions with the protein.

Glu^{39} and Glu^{40} did not have a predominant mutation replacing the original residue. However, in position 40 the presence of both hydrophobic and aromatic amino acids, suggested van der Waals or π-stacking interactions with the target.

Ala^{43} was mutated into a variety of different amino acid, mostly aromatic.

In contrast, Leu[44] was substituted by both aromatic and hydrophobic residues. Finally, Glu[47] suffered mutations leading to different kind of amino acid (Figure 9).

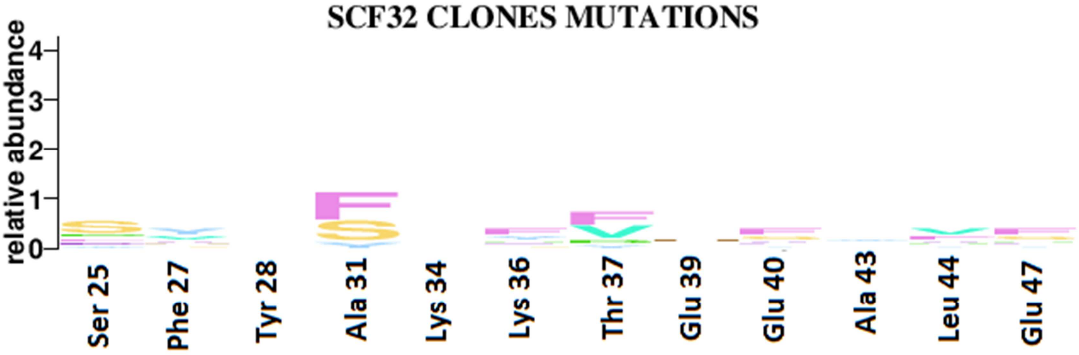


Figure 9. Amino acid compostition of the unique clones from SCF32 after phage display against Palm-ShhN-Biotin. Graphic created with WebLogo.

4.2. Cross reactivity ELISA

A very important point was to figure out if the binders obtained for Palm-ShhN-Biotin would also recognize the recombinant ShhN. This would indicate that both proteins have similar folding conformations. In order to confirm this hypothesis, it was carried out cross reactivity ELISA.

In the cross-reactivity ELISA assay, recombinant ShhN was coated on the microtiter plate and next, it was incubated with the phages that showed affinity for the synthetic analog.

After reading the plates at 450 nm all the selected phages for Palm-ShhN-Biotin also showed a positive signal. Therefore, they also recognized the recombinant ShhN. The positive results from this cross-reactivity ELISA also confirmed a similar folding of both proteins.

5. Phage display against IleIle-ShhN-Biotin analog

The IleIle-ShhN-Biotin analog lacks the palmitic residue, but instead it has a hydrophobic isoleucine dipeptide at the N-terminal region. Like the Palm-ShhN-Biotin analog, it harbors a pegylated biotin residue at the C-terminus (**Scheme *5***).

The interest of performing phage display against this synthetic analog relied on its similarity with the commercial ShhN. Because they just differed on the N-terminal hydrophobic residues mimicking the palmitic acid. The libraries screened were SCF32 and SCF42.

IIAGPGRGFG^{10} KRRHPKKLTP^{20} LAYKQFIPNV^{30} AEKTLGASGR^{40} YEGKISRNSE^{50}
RFKELTPNYN^{60} PDIIFKDEEN^{70} TGADRLMTQR^{80} CKDKLNALAI^{90} SVMNQWPGVK^{100}
LRVTEGWDED^{110} GHHSEESLHY^{120} EGRALDITTS^{130} DRDRSKYGML^{140} ARLAVEAGFD^{150}
WVYYESKAHI^{160} HCSVKAENSV^{170} AAKSGG-PEG-Biotin

Scheme 5. Amino acid sequence of IleIle-ShhN-Biotin

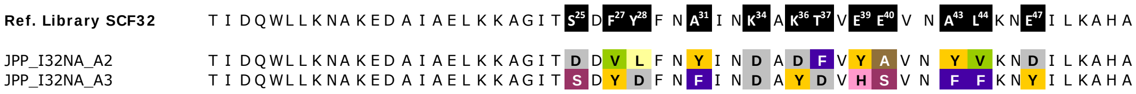
5.1. Biopanning against IleIle-ShhN-Biotin

Four rounds of biopanning were performed using plates pre-coated with neutravidin previous. With the exception of the last round, in which the coating concentration of the protein was 50 nM, for the rest of the rounds the coating contentration was maintained at 300 nM. The number of wells was kept constant at eight, for all the biopanning rounds. It was observed remarkable enrichment after the third round of selections.

Library	Positive clones	Confirmed positives	Unique clones
SCF32	14	2	2
SCF42	0	-	-

Table 4. ELISA results from phage display against IleIle-ShhN-Biotin.

Then, after plating and growing the selected colonies, ELISA assays identified just 14 binders from SCF32, whereas no one was confirmed from SCF42. After second ELISA, 2 unique binders out of the 14 first positives from SCF32 were confirmed after sequencing (**Table *4*** and **Figure *10***)



Ref. Library SCF32 T I D Q W L L K N A K E D A I A E L K K A G I T S^{25} D F^{27} Y^{28} F N A^{31} I N K^{34} A K^{36} T^{37} V E^{39} E^{40} V N A^{43} L^{44} K N E^{47} I L K A H A

JPP_I32NA_A2 T I D Q W L L K N A K E D A I A E L K K A G I T D D V L F N Y I N D A D F V Y A V N Y V K N D I L K A H A
JPP_I32NA_A3 T I D Q W L L K N A K E D A I A E L K K A G I T S D Y D F N F I N D A Y D V H S V N F F K N Y I L K A H A

Figure 10. Unique clones sequences from SCF32 obtained for synthetic analog IleIle-ShhN-Biotin.

6.1.2. Analysis of the unique clones obtained from SCF32 for IleIle-ShhN-Biotin

Interestingly, both unique clones from SCF32 obtained for the synthetic analog IleIle- ShhN-Biotin were also obtained after phage display for Palm-ShhN-Biotin. Thereby, we can hypothesize that these clones bind to the same region of the antigen and also that binders found for Palm-ShhN-Biotin may also bind the analog lacking the palmitic.

6. Competitive elisa

Once specific clones binding Palm-ShhN-biotin were identified, the next step was to determine their approximated affinity. These peptide libraries were fused to the major coat protein pVIII, thereby, expressing thousands of copies of the peptides. Although the measure of the affinity when using this pVIII display is not accurate because it can be biased due to avidity effects, it allows to determine an estimation. Forward in the project, it was planned to switch the display system from pVIII to pIII, and then the measurement of the affinity would be more accurate.

One of the easiest and most used assay to estimate the affinity of the binders for a specific target is competitive ELISA. In this type of ELISA, the target is immobilized on a microtiter plate. Then, it is incubated with a fixed concentration of the unique phages mixed with serial dilutions of the target.[4] The control well is incubated with the phage supernatant without any dilution of the target, so all the absorbance measured in this well will be due to the phages binding to the immobilized target and it will be the maximum value of absorbance. The IC_{50} will be the concentration of target in solution at which the half of the maximum absorbance value is reached.

6.1. Competitive ELISA with binders from phage display against Palm-ShhN-Biotin

For competitive ELISA, the plates were pre-coated with neutravidin at 5 µg/mL, and after blocking, the antigen Palm-ShhN-Biotin was coated at 100 nM concentration. In parallel, phage supernatants were diluted 4-fold and incubated for 1 hour with different serial concentrations of the target: 1 µM, 0.5 µM, 0.25 µM, 0.125 µM, 0.0625 µM and 0.03125 µM. Then, these mixed dilutions were incubated with the immobilized target for 15 minutes, washed right away and incubated with secondary antibody for 30 minutes. The final incubation with the TMB substrate was allowed for 10 straight minutes. The reaction was stopped with phosphoric acid and read at 450 nm.

Sadly, the analysis of the results did not show the expected pattern because there was no relation between the absorbance values and the concentration of antigen in solution. Two possible scenarios were considered: either the phages were not sufficient diluted, meaning that the number of phages present in the supernatant was high enough to bind both the antigen in solution and the immobilized, or the concentration of antigen when coating was insufficient thus leaving free neutravidin molecules that later could bind to the complex protein-phage present in the mixed solution and attach them on the plate surface.

In order to overcome these inconsistent results, it was decided to carry out a first experiment with a single clone in which the antigen was directly coated on the microtiter plate (at 100 nM) instead of attaching it through the biotin-neutravidin interaction. In addition, phage supernatants were diluted 6-fold instead of 4-fold. The pre-incubation of the phages with the antigen at different concentrations in solution and the incubation of the mixed solution with the immobilized target were performed following the same protocol as previous experiment.

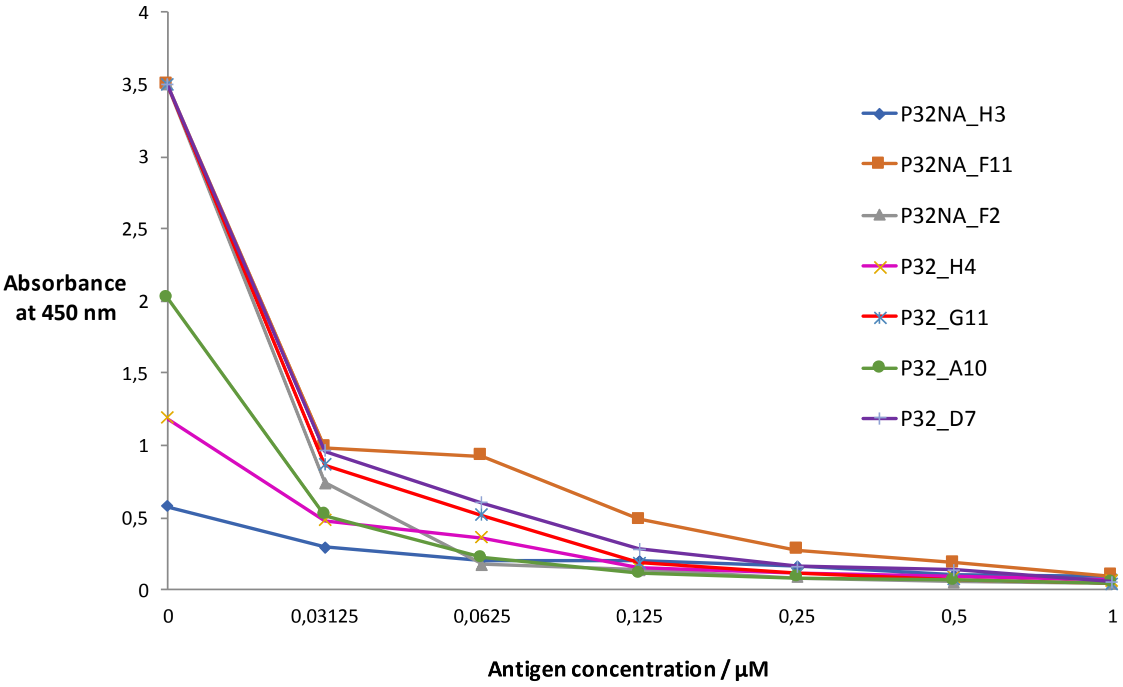


Figure 11. Graphic from competitive ELISA results for selected clones binding Palm-ShhN-biotin. Clone P36_C6 was not represented due to deviations observed on its absorbance values.

Following this new protocol, consistent results were observed: the absorbance was inversely proportional to the antigen concentration in solution. With this satisfactory result in hand, this protocol was further adopted to carry out competitive ELISA assays with the rest of the unique clones (**Figure** ***11***). Very nicely, it was possible to determine that most of the binders' affinities (measured as relative affinity) are in the range of 0.0315 µM.

In the next step, these unique clones were submitted for affinity maturation.

7. Affinity maturation for Palm-ShhN-Biotin

The main goal of the affinity maturation step is to design new libraries from the unique clones in order to select binders with improved affinity for the target.

Although there exist multiple strategies to design libraries for affinity maturation, one of the first common steps is to transfer the libraries from polyvalent (pVIII) to monovalent (pIII) display system. This first change would increase the selection of tight binders.

7.1. Affinity maturation with binders from SCF32 obtained for Palm-ShhN-Biotin

In our project, we started phage display with naïve libraries contained in pVIII phagemid vectors. Therefore, the unique clones obtained after phage display, had to be switched into pIII phagemid vector to move on affinity maturation. Phagemid pIII vector, supplied by Reflexion Pharmaceuticals, was similar to pVIII phagemid vector, but the library gene was fused to the pIII coat protein gene (**Figure *12***).

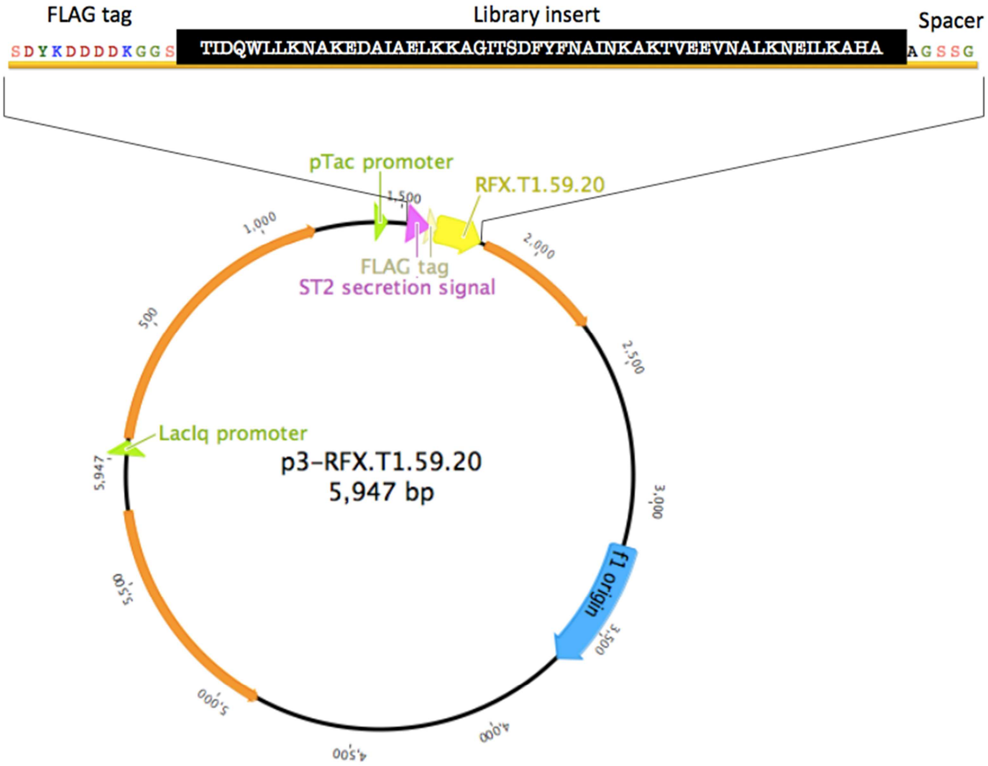

FLAG tag
Library insert
Spacer
SDYKDDDDKGGS
TIDQWLLKNAKEDAIAELKKAGITSDFYFNAINKAKTVEEVNALKNEILKAHA
AGSSG
pTac promoter
1,500
RFX.T1.59.20
FLAG tag
ST2 secretion signal
1,000
2,000
500
2,500
LacIq promoter
p3-RFX.T1.59.20
5,947 bp
5,947
3,000
f1 origin
5,500
3,500
5,000
4,000
4,500

Figure 12. pIII phagemid vector from Reflexion Pharmaceuticals. It shows the insertion corresponding to library SCF32, although any library insert can occupy the same position in the plasmid (highlighted in black). Image created with Geneious software.

1.1.1. Swtich from pVII*I* to pIII phagemid vectors

The first step to transfer the genes from pVIII phagemid vector to pIII, is the choice of the suitable restriction enzymes. These enzymes should have suitable restriction sites on both the pVIII and pIII vector: on pVIII vector they have to cut the insert length containing the unique clone gene, and on the receptor phagemid pIII they should cut on a suitable position to fuse the gene. pVIII and pIII have homologous restriction sites at NsiI and SacI: NsiI recognizes and cut the sequence 5’-ATGCAˈT-3’, whereas SacI cuts on 5’-GAGCTˈC-3’.

7.1.1. *Amplification of the gene containing the unique clones*

There are two main strategies to amplify the insert containing the desired gene: one consists on the PCR amplification of the whole plasmid, posterior cut with restriction enzymes and gel purification; and the other is the PCR amplification of the desired gene sequence using appropriate primers. In this case, it was chosen the PCR amplification of the gene.

In order to amplify the region of interest it was necessary the design of new. In our case the restriction site was not added to the primer because it was already present on the sequence. Moreover, primers should be at least 21 nucleotides long to ensure specific alignment and also pair to 5 bases up and down flanking the desired sequence to amplify. The sequence of the primers designed are shown below:

FORWARD primer 5’-GGA TAT GCA TCC GAT TAT AAA GATG-3’
REVERSE primer 5’-TCC AGA GCT CCC GGC GTG AGC TTTC-3’

PCR amplification of the unique clones required some additional optimization of the PCR program (**Table *5***). First, a scale up of the PCR reaction was done until 100 µL of reaction volume and the amount of phage supernatant added changed from 0.1 to 0.4 µL. Next, three extra cycles were incorporated (twenty-eight instead of previously twenty-five cycles). And finally, the annealing temperature was decreased it to 50 °C instead of 60 °C because of the different content on GC bases between the primers, there was a high difference in the annealing temperature. After these changes, PCR yielded bright bands, meaning that the PCR and the primers worked out (**Figure *13***).

	Temperature (ºC)	Time	Cycles
Initial denaturation	94	5 min	1x
Denaturation	94	30 sec	28x
Annealing	50	30 sec	
Elongation	72	1 min	
Final elongation	72	7 min	1x
Cooling	4	Unlimited time	

Table 5. Optimized PCR program for amplification of the unique clones genes.

However, there was a clone, P32_D7, that did not show amplification of the desired region (the gel showed a smaller fragment), which could be due to the presence of stop codons. So, for this specific clone, it was decided to amplify the whole plasmid, then cut with restriction enzymes and purify the band of the gel that contains the desired fragment.

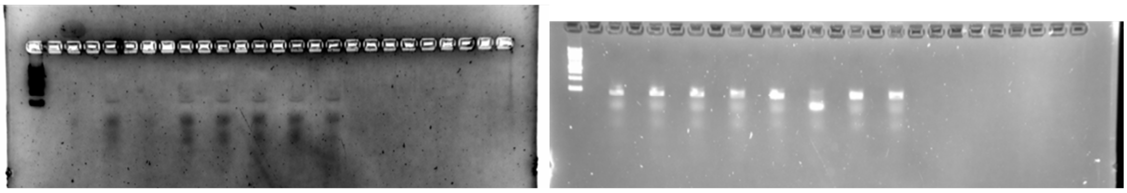
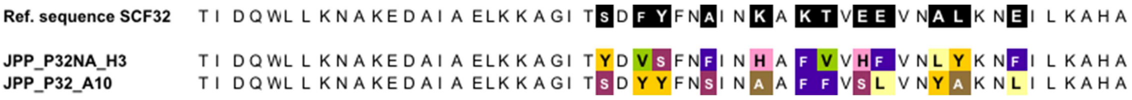


Figure 13. Image of the PCR product on 2% agarose gel run for 8 min at 220V. On the left PCR products before PCR optimization. On the right PCR product bands after PCR optimization. Clones per well: well 1 P32_NA_H3, well 2 P32_NA_F11, well 3 P32_F2, well 4 P32_H4, well 5 P32_G11, well 6 P32_D7, well 7 P32_A10, well 8 P32_C6.

7.1.2. Digestion and ligation of the amplified inserts into pIII paghemid vector

The clones with inserts successfully amplified by PCR were digested with NsiI and SacI at 37 °C for 2 hours. In parallel minipreps of the pIII phagemid vector were carried out and later were digested with the same restriction enzymes and purified. The next step consisted in the ligation of the inserts into the digested plasmid. The ligation was conducted at room temperature for 2 hours, and once finished, the ligated plasmids were transformed into competent *E. Coli* CJ236 cells and grown in agar plates containing carbenicillin. A control ligation without any insert was carried out.

As expected, more number of colonies were observed on plates from ligation compared to the control. However, after sequencing only 2 phagemids contained the desired insert (**Figure *14***). The other six did not have neither the insert in frame or contained stop codons.



Ref. sequence SCF32	TIDQWLLKNAKEDAIAELKKAGITSDFYFNAINKAKTVEEVNALKNEILKAHA
JPP_P32NA_H3	TIDQWLLKNAKEDAIAELKKAGITYDVSFNFINHAFVVHFVNLYKNFILKAHA
JPP_P32_A10	TIDQWLLKNAKEDAIAELKKAGITSDYYFNSINAAFFVSLVNYAKNLILKAHA

Figure 14. Amino acid sequence after ligation into pIII phagemid vector

7.2. Design of the library for affinity maturation

The following stage in affinity maturation consists in the design and construction of a library derived from the clones obtained.

We started the design of the libraries derived from the two clones successfully transferred to the monovalent vector. In order to do so, we took into consideration the information from the analysis of all the unique clones obtained, and tried to put together that data with the clones contained in pIII vectors. The objective was to design primers for the two clones in pIII vector that could introduce desired mutations in specific residues by PCR mutagenesis. The primers designed would contain degenerated codons (**Table *6***) to introduce random residues or more specific ones, in accordance with the previous analysis of the clones obtained (**Figure *15***). Due to the fact that ideally the length of the primers for PCR mutagenesis should be between 25 to 45 bases we had to split the area we wished to run the PCR mutagenesis into three primers to cover all the residues to mutate. Thereby, the PCR mutagenesis would be sequential, one after the other using one of the primers designed every time.

The primers designed will be used on the construction of the libraries in the future.

Codon[a]	Description	Aminoacids	Stop codons	Unique codons[b]
NNK	All 20 amino acids	All 20	TAG	32
NNC	15 amino acids	A,C,D,F,G,H,I,L,N,P,R,S,T,V,Y	None	16
NVT	Charged, hydrophilic	C,D,G,H,N,P,R,S,T,Y	None	12
NNT	Mixed	A,D,G,H,I,L,N,P,R,S,T,V	None	16
TDK	Hydrophobic	C,F,L,W,Y	TAG	6
RST	Small side chains	A,G,S,T	None	4

Table 6. Degenerated codons. Table adapted from [9]. [a]Equimolar DNA degeneracies are represented in the IUB code (D, A/G/T; K, G/T; N, A/C/G/T; R, A/G; S, G/C; V, A/C/G; W, A/T). [b]The number of unique codons contained in the degenerate codon. Owing to the nature of the genetic code, many degenerate codons are redundant. That is, the number of unique codons exceeds the number of unique amino acids encoded.

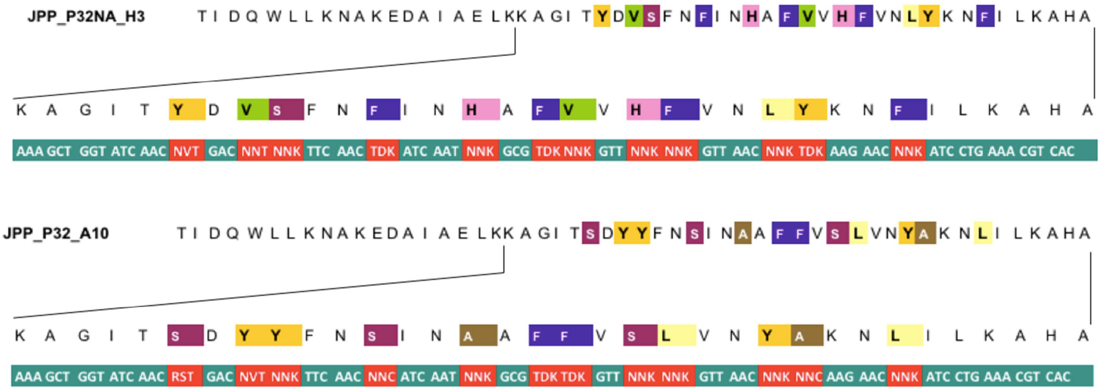

JPP_P32NA_H3
T I D Q W L L K N A K E D A I A E L K K A G I T Y D V S F N F I N H A F V V H F V N L Y K N F I L K A H A
K A G I T Y D V S F N F I N H A F V V H F V N L Y K N F I L K A H A
AAA GCT GGT ATC AAC NVT GAC NNT NNK TTC AAC TDK ATC AAT NNK GCG TDK NNK GTT NNK NNK GTT AAC NNK TDK AAG AAC NNK ATC CTG AAA CGT CAC
JPP_P32_A10
T I D Q W L L K N A K E D A I A E L K K A G I T S D Y Y F N S I N A A F F V S L V N Y A K N L I L K A H A
K A G I T S D Y Y F N S I N A A F F V S L V N Y A K N L I L K A H A
AAA GCT GGT ATC AAC RST GAC NVT NNK TTC AAC NNC ATC AAT NNK GCG TDK TDK GTT NNK NNK GTT AAC NNK NNC AAG AAC NNK ATC CTG AAA CGT CAC

Figure 15. Primers designed for PCR mutagenesis. Residues to be mutated were identified (bases in red) and most suitable degenerate codons were introduced in those positions substituting the original bases. To cover all the mutations 3 primers were designed.

In addition to the construction of the libraries, future work will be also focused on the validation of the binders encountered. Actually, to determine the *off phage* affinity of the clones obtained, we are currently working on the chemical synthesis of some of the peptide binders. The objective, once they will be synthetized, is to test their affinity for the synthetic

Palm-ShhN-Biotin and also the recombinant ShhN using Biacore system. Moreover, it would be ideal to determine if the clones selected bind the region of ShhN that recognizes its receptor, and therefore, if they could be used as antagonists. To further confirm or deny this hypothesis, assays with cells and also competitive ELISA using the binders and the 5E1 antibody, would have to be performed.

CONCLUSIONS

The conclusions of this chapter are:

1. After phage display using seven naïve scaffold libraries against commercial ShhN (SCF28, SCF32, SCF40, SCF42, SCF53, SCF70 and SCF95), a total of thirty-four unique clones were obtained from libraries SCF32 and SCF42.

2. Phage display using these seven naïve libraries, against N-terminal biotinylated ShhN yielded five unique clones for this analog. These were different from the clones obtained for commercial ShhN.

3. When using SCF32 and SCF42 libraries for phage display against IleIle-Shh-Biotin two unique clones from SCF32 were selected, one of them common with the Palm-ShhN-Biotin analog.

4. Phage display using libraries SCF32 and SCF42 against Palm-ShhN-Biotin resulted in eight unique clones from SCF32. All these unique clones displayed cross-reactivity against recombinant ShhN, thus confirming a similar folding of the synthetic analog.

5. In order to estimate the affinity of the binders obtained for Palm-Shh-biotin, competitive ELISA assays were carried out. The results showed that the peptide binders displayed in a polyvalent system had an IC_{50} around 0.04 µM. Encouraged by the results we decided to move forward to perform affinity maturation with all these unique clones.

6. For affinity maturation, the first requirement was to display the selected peptides in a monovalent system. Thus, the selected clones had to be switched from the pVIII system to the pIII phagemid vector. We tried to transfer all the eight binders, but due to problems during ligation step just two were successfully switched. Since our objective was the design of new libraries for affinity maturation we decided to move on with these two binders.

7. We started the design of new primers for PCR mutagenesis which would introduce random or more specific residues at selected positions on the new libraries for affinity maturation. Future work will be addressed to the construction of the libraries.

8. In addition, validation of the binders' affinity *off phage* will be studied in the near future. We are working on the synthesis of the ligands and will carry out biophysically studies to determine their affinity for the target.

REFERENCES

(1) Smith, G. Filamentous Fusion Phage: Novel Expression Vectors That Display Cloned Antigens on the Virion Surface. *Science (80).* **1985**, *228* (4705), 1315–1317.

(2) McCafferty, J.; Griffiths, A. D.; Winter, G.; Chiswell, D. J. Phage Antibodies: Filamentous Phage Displaying Antibody Variable Domains. *Nature*. 1990, pp 552–554.

(3) Barbas, C. F.; Kang, A. S.; Lerner, R. A.; Benkovic, S. J. Assembly of Combinatorial Antibody Libraries on Phage Surfaces: The Gene III Site. *Proc. Natl. Acad. Sci.* **1991**, *88* (18), 7978–7982.

(4) Lowman, H. B. Bacteriophage Display and Discovery of Peptide Leads for Drug Development. *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 401–424.

(5) Chasteen, L.; Ayriss, J.; Pavlik, P.; Bradbury, A. R. M. Eliminating Helper Phage from Phage Display. *Nucleic Acids Res.* **2006**, *34* (21), 1–11.

(6) Clackson, T.; Lowman, H. B.; Russel, M. Introduction to Phage Biology and Phage Display. In *Phage Display*; Clackson, T., Lowman, H. B., Eds.; Oxford, **2004**; pp 1–26.

(7) Haque, A.; Tonks, N. K. The Use of Phage Display to Generate Conformation-Sensor Recombinant Antibodies. *Nat. Protoc.* **2012**, *7* (12), 2127–2143.

(8) Bazan, J.; Całkosiński, I.; Gamian, A. Phage Display - A Powerful Technique for Immunotherapy: 1. Introduction and Potential of Therapeutic Applications. *Hum. Vaccines Immunother.* **2012**, *8* (12), 1817–1828.

(9) Tonikian, R.; Zhang, Y.; Boone, C.; Sidhu, S. S. Identifying Specificity Profiles for Peptide Recognition Modules from Phage-Displayed Peptide Libraries. *Nat. Protoc.* **2007**, *2* (6), 1368–1386.

(10) Rasched, I.; Oberer, E. Ff Coliphages: Structural and Functional Relationships. *Microbiol. Mol. Biol. Rev.* **1986**, *50* (4), 401–427.

(11) Velappan, N.; Fisher, H. E.; Pesavento, E.; Chasteen, L.; Angelo, S. D.; Kiss, C.; Longmire, M.; Pavlik, P.; Bradbury, A. R. M. A Comprehensive Analysis of Filamentous Phage Display Vectors for Cytoplasmic Proteins : An Analysis with Different Fluorescent Proteins. **2010**, *38* (4).

(12) Fuh, G.; Sidhu, S. S. Efficient Phage Display of Polypeptides Fused to the Carboxy-Terminus of the M13 Gene-3 Minor Coat Protein. *FEBS Lett.* **2000**, *480* (2–3), 231–234.

(13) Gao, C.; Mao, S.; Lo, C.-H. L.; Wirsching, P.; Lerner, R. A.; Janda, K. D. Making Artificial Antibodies: A Format for Phage Display of Combinatorial Heterodimeric Arrays. *Chemistry (Easton).* **1999**, *96* (May), 6025–6030.

(14) Høydahl, L. S.; Nilssen, N. R.; Gunnarsen, K. S.; Pré, M. F. du; Iversen, R.; Roos, N.; Chen, X.; Michaelsen, T. E.; Sollid, L. M.; Sandlie, I.; Løset, G. Å. Multivalent pIX Phage Display Selects for Distinct and Improved Antibody Properties. *Sci. Rep.* **2016**, *6* (1), 39066.

(15) Held, H. A.; Sidhu, S. S. Comprehensive Mutational Analysis of the M13 Major Coat

Protein : Improved Scaffolds for C-Terminal Phage Display. **2004**, 587–597.

(16) Sidhu, S. S.; Feld, B. K.; Weiss, G. a. M13 Bacteriophage Coat Proteins Engineered for Improved Phage Display. *Methods Mol. Biol.* **2007**, *352* (5), 205–219.

(17) Hufton, S. E.; Moerkerk, P. T.; Meulemans, E. V.; De Bru??ne, A.; Arends, J. W.; Hoogenboom, H. R. Phage Display of cDNA Repertoires: The pVI Display System and Its Applications for the Selection of Immunogenic Ligands. *J. Immunol. Methods* **1999**, *231* (1–2), 39–51.

(18) Fransen, M.; Van Veldhoven, P. P.; Subramani, S. Identification of Peroxisomal Proteins by Using M13 Phage Protein VI Phage Display: Molecular Evidence That Mammalian Peroxisomes Contain a 2,4-Dienoyl-CoA Reductase. *Biochem. J.* **1999**, *340*, 561–568.

(19) Løset, G. Å.; Sandlie, I. Next Generation Phage Display by Use of pVII and pIX as Display Scaffolds. *Methods* **2012**, *58* (1), 40–46.

(20) Iannolo, G.; Minenkova, O.; Petruzzelli, R.; Cesareni, G. Modifying Filamentous Phage Capsid: Limits in the Size of the Major Capsid Protein. *J. Mol. Biol.* **1995**, *248* (4), 835–844.

(21) Levin, A. M.; Weiss, G. A. Optimizing the Affinity and Specificity of Proteins with Molecular Display. *Mol. BioSyst.* **2006**, *2* (1), 49–57.

(22) Smith, G. P.; Petrenko, V. A. Phage Display. *Chem. Rev.* **1997**, *97* (2), 391–410.

(23) Lowman, H. B. Phage Display of Peptide Libraries on Protein Scaffolds. *Methods Mol. Biol.* **1998**, *87*, 249–264.

(24) Wells, J. A. 333 [21] Phage Display for selectin of Novel Binding Peptides. **2000**, *328*.

(25) Rakonjac, J.; Jovanovic, G.; Model, P. Filamentous Phage Infection-Mediated Gene Expression: Construction and Propagation of the gIII Deletion Mutant Helper Phage R408d3. *Gene* **1997**, *198* (1–2), 99–103.

(26) Yanisch-Perron, C.; Vieira, J.; Messing, J. Improved M13 Phage Cloning Vectors and Host Strains: Nucleotide Sequences of the M13mpl8 and pUC19 Vectors. *Gene* **1985**, *33* (1), 103–119.

(27) Vieira, J.; Messing, J. Production of Single-Stranded Plasmid DNA. *Methods Enzymol.* **1987**, *153* (1982), 3–11.

(28) Martens, C. L.; Cwirla, S. E.; Lee, R. Y.; Whitehorn, E.; Chen, Y.; Bakker, A.; Martin, E. L.; Wagstrom, C.; Gopalan, P.; Smith, W.; Tate, E.; Koller, K. J.; Schatz, P. J.; Dower, W. J.; Barrett, W.; Chen, E. Y.; Smith, C. W.; Barrett, R. W. Protein Chemistry and Structure : Peptides Which Bind to E-Selectin and Block Neutrophil Adhesion Peptides Which Bind to E-Selectin and Block Neutrophil Adhesion . **1995**, *270* (36), 21129–21136.

(29) Wrighton, N. C.; Farrell, F. X.; Chang, R.; Kashyap, a K.; Barbone, F. P.; Mulcahy, L. S.; Johnson, D. L.; Barrett, R. W.; Jolliffe, L. K.; Dower, W. J. Small Peptides as Potent

Mimetics of the Protein Hormone Erythropoietin. *Science* **1996**, *273* (1 7), 458–464.

(30) Tjhung, K. F.; Kitov, P. I.; Ng, S.; Kitova, E. N.; Deng, L.; Klassen, J. S.; Derda, R. Silent Encoding of Chemical Post-Translational Modifications in Phage-Displayed Libraries. *J. Am. Chem. Soc.* **2016**, *138* (1), 32–35.

(31) Ng, S.; Lin, E.; Kitov, P. I.; Tjhung, K. F.; Gerlits, O.; Deng, L.; Kasper, B.; Sood, A.; Paschal, B.; Zhang, P.; Ling, C.; Klassen, J. S.; Noren, C.; Mahal, L.; Woods, R.; Coates, L.; Derda, R. Genetically Encoded Fragment-Based Discovery of Glycopeptide Ligands for Carbohydrate-Binding Proteins. *J. Am. Chem. Soc.* **2015**, *137*, 5248–5251.

(32) Palacios-Rodríguez, Y.; Gazarian, T.; Huerta, L.; Gazarian, K. Constrained Peptide Models from Phage Display Libraries Highlighting the Cognate Epitope-Specific Potential of the Anti-HIV-1 mAb 2F5. *Immunol. Lett.* **2011**, *136* (1), 80–89.

(33) Giebel, L. B.; Cass, R. T.; Milligan, D. L.; Young, D. C.; Arze, R.; Johnson, C. R. Screening of Cyclic Peptide Phage Libraries Identifies Ligands That Bind Streptavidin with High Affinities. *Biochemistry* **1995**, *34* (47), 15430–15435.

(34) Ng, S.; Derda, R. Phage-Displayed Macrocyclic Glycopeptide Libraries. *Org. Biomol. Chem.* **2016**, *14* (24), 5539–5545.

(35) Chen, S.; Rebollo, I. R.; Buth, S. A.; Morales-sanfrutos, J.; Touati, J.; Leiman, P. G.; Heinis, C. Bicyclic Peptide Ligands Pulled out of Cysteine-Rich Peptide Libraries *J. Am. Chem. Soc.*. **2013,** *135,* 6562-6569.

(36) Deyle, K.; Kong, X.; Heinis, C. Phage Selection of Cyclic Peptides for Application in Research and Drug Development. *J. Am. Chem. Soc.* **2017**, *50*, 1866–1874.

(37) Braisted, A. C.; Wells, J. A. Minimizing a Binding Domain from Protein A. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93* (12), 5688–5692.

(38) Francisco, J. A.; Campbell, R.; Iverson, B. L.; Georgiou, G. Production and Fluorescence-Activated Cell Sorting of Escherichia Coli Expressing a Functional Antibody Fragment on the External Surface. *Proc. Natl. Acad. Sci.* **1993**, *90* (22), 10444–10448.

(39) Mattheakis, L. C.; Bhattt, R. R.; Dower, W. J. An in Vitro Polysome Display System for Identifying Ligands from Very Large Peptide Libraries. *Biochemistry* **1994**, *91*, 9022–9026.

(40) Kawakami, T.; Murakami, H.; Suga, H. Messenger RNA-Programmed Incorporation of Multiple N-Methyl-Amino Acids into Linear and Cyclic Peptides. *Chem. Biol.* **2008**, *15* (1), 32–42.

(41) Yamagishi, Y.; Shoji, I.; Miyagawa, S.; Kawakami, T.; Katoh, T.; Goto, Y.; Suga, H. Article Natural Product-Like Macrocyclic N -Methyl-Peptide Inhibitors against a Ubiquitin Ligase Uncovered from a Ribosome-Expressed De Novo Library. *Chem. Biol.* **2011**, *18* (12), 1562–1570.

(42) Weiss, G. A.; Wells, J. A.; Sidhu, S. S. Mutational Analysis of the Major Coat Protein of

M14 Identifies Residues That Control Protein Display. *Protein Sci.* **2000**, *9*, 647–654.

(43) Taylor, F. R.; Wen, D.; Garber, E. A.; Carmillo, A. N.; Baker, D. P.; Arduini, R. M.; Williams, K. P.; Weinreb, P. H.; Rayhorn, P.; Hronowski, X.; Whitty, A.; Day, E. S.; Boriack-Sjodin, A.; Shapiro, R. I.; Galdes, A.; Pepinsky, R. B. Enhanced Potency of Human Sonic Hedgehog by Hydrophobic Modification. *Biochemistry* **2001**, *40* (14), 4359–4371.

(44) Liao, S.-M.; Du, Q.-S.; Meng, J.-Z.; Pang, Z.-W.; Huang, R.-B. The Multiple Roles of Histidine in Protein Interactions. *Chem. Cent. J.* **2013**, *7* (1), 44.

MATERIALS AND METHODS

1. Reagents and solvents

Reagents of the purest grade available were purchased from commercial sources and used without further purification: DMF (synthesis grade) from Panreac. CH_2Cl_2, acetic acid glacial, piperidine and CH_3CN (HPLC grade) from Carlo Erba. TIS (Triisopropylsilane), TCEP·HCl (Tris(2-carboxyethyl)phosphine hydrochloride), DIEA (N,N-diisopropylethylamine), DIC (N,N'-diisopropylcarbodiimide), MPAA (4-Mercaptophenylacetic acid, 97%), Sodium 2-mercaptoethanesulfonate (MESNa), silver triflate, palmitic acid, methoxylamine hydrochloride, triphosgene, pyridine, magnesium sulfate anhydrous, *p*-cyanophenol, *p*-fluorophenylchloroformate, phenylchloroformate, and 2,2'-Azobis[2-(2-imidazolin-2-yl)propane] dihydrochloride (VA-044) were purchased from Sigma-Aldrich. TFA (peptide grade) and 4-Fluoro-3-nitrobenzoic acid were obtained from Fluorochem. $MeNH_2$ (40% in MeOH), and MPOH (4-Mercaptophenol, 95%) from TCI. HBTU (2-(1H-Benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate) and Fmoc-amino acids from Iris-Biotech. Side-chain protecting groups were: Arg(Pbf), Asp(tBu), Asn(Trt), Cys(Trt), Glu(tBu), Gln(Trt), His(Trt), Lys(Boc), Ser(tBu), Thr(tBu), Trp(tBu), Tyr(tBu). HATU (2-(1H-7-Azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate) was from Medalchemy (Spain). OxymaPure (Ethyl cyano(hydroxyimino)acetate) was obtained from Louxemburg Laboratories. For ligation experiments, Guanidine Hydrochloride Ultra-Pure (MP Biomedicals) was used. Water was purified using a Millipore MilliQ water purification system. Salts (from Sigma): 0.1 M $MgCl_2$ (cat. no. M8266), 1.0 M Tris (pH 8.0), KCl (cat. no. P9541), NaCl (cat. no. S3014), EDTA (cat. no. E9884), Tris-acetate (cat. no. 93339), Na_2HPO_4 (cat. no 255793), $ZnCl_2$ (cat. no. 229997), K_2HPO_4 (cat. no. 450200).

1.0 M H_3PO_4 (Fisher, cat. no. AC29570); 1.0 M HEPES, pH 7.4 (Fisher, cat. no. BP299); PEG-8000 (Fisher, cat. no. BP233); 1.0 M Tris, pH 11.0 (Invitrogen, cat. no. 15568-025); 10 mM ATP (Invitrogen, cat. no. 18330-019); SYBR™ Safe DNA Gel Stain (Thermofischer, cat. no. S33102); 100 mM dithiothreitol (Fisher, cat. no. BP172; Sigma); 1 kb DNA ladder (New England Biolabs, cat. no. N3232L); 100 mM HCl (Fisher, cat. no. A144), 10mM dNTP mix (10mM of each dATP,dCTP,dGTPanddTTP) (Fermentas, cat. no. R0192); Agarose (Fisher, cat. no. BP160); Tetracycline (Sigma, cat. no. T7660), 5 mg/mL in water, filter sterilize; Kanamycin (Fisher, cat. no. AC61129), 5 mg/mL in water, filter-sterilize; Carbenicillin (Sigma, cat. no. C1389), 5 mg/mL in water, filter-sterilize; Tryptone (Fisher, cat. no. 0123-07-5); Tween 20 (Sigma, cat. no. P2287); BSA (Sigma, cat. no. 05477); Granulated agar (Fisher, cat. no. DF-0140-07-04);

Ultrapure irrigation USP water (Braun Medical Inc., cat. No. R5007); yeast extract (VWR, cat. No. DF0127-08); Sonic Hedgehog from Peptrotech (cat. no. 100-45); Pierce™ TMB Substrate Kit H_2O_2 peroxidase substrate (Kirkegaard & Perry Laboratories Inc., cat. no. 50-65-02); HRP/anti-M13 antibody conjugate (GE Healthcare, cat. no. 27-9421-01); PCR cleanup mix (exonuclease I and shrimp alkaline phosphatase from United States Biochemical, cat. no. 7820); QIAquick Gel Extraction Kit (Qiagen, cat. no. 28706); QIAquick PCR Purification Kit (Qiagen, cat. no. 28104); n-Dodecyl-beta-maltoside (DDM) from Cube Biotech.

Restriction endonucleases (all from New England Biolabs): NsciI-HF (cat. no. R3127S), SacI-HF (cat. no. 3156S), Cutsmart buffer (cat. no. B7204S); T4 DNA ligase (Invitrogen, cat. no. 15224-041), T4 DNA polymerase (New England Biolabs, cat. no. M0203S), AmpliTaq DNA polymerase (Applied Biosystems, cat. no. N8080157).

Bacterial strains: *E. Coli* CJ236 (New England Biolabs), One Shot^{TM} OmniMAX^{TM} 2 T1R Chemically Competent *E. Coli*.

2. Materials

For SPPS polypropylene syringes and Teflon valves were used.96-well Maxisorp immunoplates (Nunc, cat. no. 430341), baffled flasks, 1.5 mL microcentrifuge tubes, 15 mL cell culture tubes from Fischer Scientific (cat. No. 14-959-11B), 50 mL falcon tubes from VWR (cat. no. 89039-658), Nunc® MaxiSorp™ 384 well plates from Sigma Aldrich (cat. no. P6366-1CS), Axygen® 96 Well Polypropylene Cluster Tubes (Blue Box) from Thomas Scientific (cat. no. MTS-11-12-C).

3. Instrumentation

3.1. General instrumentation

Balances: Mettler Toledo model MS 303-S, AB 204-S and AT-261; Uv-VIS spectrometer shimadzu mini 1240, orbital shakers Heidolph, vortex mixers MERCK® eurolab, plate reader Biotek Synergy H1 Hybrid Reader, plate washer ELx50 Autostrip Washer, benchtop centrifuge Eppendorf Microcentifuge model 5424 and 5424R, shaker incubator Innova 26 Incubator Shaker (from New Brunswick Scientific), Sorvall^{TM} XT Centrifuge, titer plate shaker from EDM Millipore, 12–25ºC water bath, BIO-RAD PowerPAC Basic Electrophoresis Power Supply, Biorad electrophoresis kit gel kit, B-Box^{TM} Blue Light LED Epi-Iluminator from Smobio Technology Inc., S1000^{TM} Thermal Cycler from Biorad, ChemiDoc^{TM} XRS+ System from Biorad.

3.2. RP-HPLC.

3.2.1. Analytitcal RP-HPLC

Analytical HPLC were carried out in a Waters e2695 separations module equipped with a Waters 2998 photodiode array detector. Separations were performed using the following columns: XBridge C18 (5 µm, 150x4.6 mm), Phenomenex Aeris XB-C18 (3.6 µm, 150x4.6 mm) and Agilent Zorbax SB-C3 (5 µm, 4.6x150 mm) at a flow rate of 1 mL/min using H_2O (buffer A) and CH_3CN (buffer B) containing 0.045% and 0.036% TFA, respectively, in the buffer system.

3.2.2. Semi-preparative RP-HPLC

Preparative HPLC was performed in a Waters 2545 connected to a Waters 2489 dual λ absorbance detector using a Jupiter Proteo 90Å (10 µm, 100x21.2 mm) at a flow rate of 16 mL/min or XBridge OBD PrepC18 column (5 µm, 100x10 mm) or a Jupiter C18 (10 µm, 300 Å, 10x150 mm) at a flow rate of 6 mL/min operating in a linear mode with H_2O and CH_3CN containing 0.1% and 0.05% TFA, respectively. Preparative injections were monitored at 220 and 280 nm.

3.3. Mass spectrometry.

Peptide masses were determined on a Waters 2695 separations module connected to a Waters 2998 photodiode array detector, and a Waters Micromass ZQ with a quadropole mass analyzer (m/z) using electrospray ionization operating in the positive ion mode.

Peptide masses were also determined by UPLC-MS Waters Acquity H class UPLC (separation using Acquity UPLC BEH C18 Column, 130 Å, 1.7 µm, 2.1 mm x 100 mm) system equipped with Acquity photodiode array detector and a UPLC SQ Detector2 (Zspray^{TM}), at a flow rate of 0.610 mL/min using H_2O with 0.1% formic acid (solvent A) and acetonitrile with 0.07% formic acid (solvent B).

3.4. MALDI-TOF

Mass spectra were recorded on MALDI-TOF Applied Biosystem 4700 with N_2 laser of 337 nm using ACH matrix (10 mg/mL of ACH in CAN-H_2O-TFA (1:1:0.1)).

Sample preparation: a mixture of sample solution (1 µL) and matrix (1 µL) is prepared, placed on a MALDI-TOF plate and dried by air.

3.5. Nuclear magnetic resonance

^{1}H and ^{13}C NMR spectra were recorded on a Varian MERCURY 400 (400 MHz for ^{1}H NMR, 100 MHz for ^{13}C NMR) spectrometer for organic small molecules. Chemical shifts (δ) are expressed in parts per million downfield from tetramethylsilyl chloride. Coupling constants are expressed in Herts.

3.6. Microwave synthesis.

Microwave assisted peptide synthesis was carried out in a CEM Microwave Peptide Synthesizer.

3.7. Circular dichroism.

The circular dichroism experiments were carried out in spectro-polarimeter (Jasco Inc., Easton, MD USA) and Spectra Manager version 1.53.01 (Jasco). The following parameters were used: sensitivity (standard (100 mdeg)), start (260 nm), end (195 nm), data pitch (0.1 nm), scanning mode (immediately), scanning speed (20 nm/min), response (1 sec), band width (1.0 nm), and accumulation (5).

4. Buffers

- Buffer C or NCL buffer: 6 M Gn·HCl, 0.2 M phosphate, 20 mM TCEP.
- Buffer D: 6M Gn·HCl, 0.2 M phosphate, 250 mM MESNa, 250 mM TCEP and 100 mM VA-044.
- Loading sample buffer 4x: Tris-HCl 0.4 M, pH 6.8, glycerol 69.6 %, SDS 8 % (p/v), bromophenol blue 0.05 mg/mL, DDT 100 mM.
- Fixation buffer: 50 ml MeOH, 12 ml AcOH, milliQ H2O up to 100 ml. Add 50 µl 37 % formaldehyde at the last moment (just before soaking the gel into this buffer).
- Pretreatment buffer: 19.12 mg sodium thiosulfate (Na2S2O3) in 150 ml milliQ H2O.
- Impregnation buffer: 0.2 g silver nitrate (AgNO3), milliQ H2O up to 100 ml. Keep it wrapped in aluminium foil! Add 74.8 µl 37 % formaldehyde at the last moment (just before soaking the gel into this buffer).
- Revealing solution: 6 g sodium carbonate (Na2CO3), 2 ml Pretreatment solution, milliQ H2O up to 100 ml. Add 50 µl 37 % formaldehyde at the last moment (just before soaking the gel into this buffer).
- Stop solution:. 50 ml MeOH, 15 ml AcOH, milliQ H2O up to 100 ml.
- PBS (1x): 137 mM NaCl, 3 mM KCl, 8 mM $NaHPO_4$ and 1.5 mM $KHPO_4$, pH was adjusted to 7.2 with HCl, and autoclaved.

- PT buffer: PBS, 0.05% Tween 20, filter-sterilized.
- PBT buffer: PBS, 0.05% Tween 20, 0.5% BSA (w/v), filter-sterilized.
- KCM buffer: 0.5 M KCl, 0.15 M $CaCl_2$, 0.25 M $MgCl_2$
- Blocking buffer: PBS, 0.5% BSA (w/v), filter-sterilized.
- PEG/NaCl: 20% PEG-8000 (w/v), 2.5 M NaCl. Mix and autoclave.
- TAE buffer: 40 mM tris-acetate, 1.0 mM EDTA, adjust pH to 8.0; autoclave.
- 1% TAE/agarose gel: TAE buffer, 1.0% (w/v) agarose, 1:5000 (v/v) 10% SYBR DNA stain.
- 2% TAE/agarose gel: TAE buffer, 2.0% (w/v) agarose, 1:5000 (v/v) 10% SYBR DNA stain.
- TMB substrate: mix equal volumes of TMB and H_2O_2 peroxidase substrate.
- LB medium: 5 g yeast extract, 10 g tryptone, 10 g NaCl. Add water to make up the volume to 1.0 liter, adjust pH to 7.0 with NaOH, autoclave. For LB agar, add 20 g of granulated agar.
- LB/carb plates LB agar, 50 mg/ml carbenicillin.
- LB/tet plates LB agar, 5 mg/ml tetracycline.
- LB/kan plates LB agar, 50 mg/ml kanamycin.
- 10x PCR buffer: 600 mM Tris-HCl pH 8.3, 250 mM KCl, 15 mM $MgCl_2$.
- 2YT medium: 10 g yeast extract, 16 g tryptone, 5 g NaCl. Add water to make up the volume to 1.0 liter, adjust pH to 7.0 with NaOH, autoclave.
- 2YT/carb/KO7 medium: 2YT, 50 µg/mL carbenicillin, 10^{10} phage/mL of M13KO7.
- 2YT/carb/tet medium: 2YT, 50 µg/mL carbenicillin, 10 µg/mL tetracycline.
- 2YT/carb/tet/KO7 medium: 2YT, 50 mg/mL carbenicillin, 10 mg/mL tetracycline, 10^{10} phage per ml of M13KO7.
- 2YT/kan medium: 2YT, 25 mg/mL kanamycin.
- 2YT/kan/tet medium: 2YT, 25 mg/mL kanamycin, 5 mg/mL tetracycline.
- 2YT/tet medium 2YT, 5 µg/mL tetracycline.
- Folding buffer: 150 mM NaCl, 25 mM NaPi, 25 mM DTT, 0.5% DDM, 1 µM $ZnCl_2$, pH 8.0.

5. Peptide synthesis and characterization

5.1. General considerations

Peptides were synthetized by Fmoc-SPPS. Peptide elongation, linker acylation and cyclization on resin were performed in polypropylene syringes equipped with a polypropylene porous filter. All solvents and soluble reagents were removed by suction. Washings between deprotection and coupling steps were carried out with DMF (3 x 1 min) and DCM (3 x 1 min) using 10 mL solvent/g resin for each wash. During coupling, linker acylation and cyclization steps, the mixture was stirred in orbital shakers.

5.2. Colorimetric tests

5.2.1. Kaiser test

The Kaiser test is a colorimetric test that enables qualitative detection of free primary amines. The peptide-resin is washed with DCM and the solvent removed by suction. A small amount of resin beads are transferred to a glass tube and 6 drops of solution A and 2 drops of solution B are added. Then, the tube is incubated at 110 °C for three minutes. A dark blue color in the supernatant and/or on the resin beads reveals the presence of free primary amines. This indicates that the test is positive, and thereby, the coupling is incomplete. Contrary, when the supernatant is yellow, it means the absence of free primary amines, and the test is negative, ensuring 99.5 % of coupling.

Solution A was prepared by adding 40 g of phenol to 10 mL of ethanol and heated until complete dissolution. Another solution of 65 mg of KCN in 100 mL of water was prepared and mixed with 100 mL of pyridine (freshly distilled over ninhydrin). Both solutions were stirred for 45 minutes with 4g of Amberlite MB-3 resin and finally filtered and mixed.

Solution B was prepared by dissolving 205 mg of ninhydrin in 50 mL of ethanol. This solution was protected from light.

5.2.2. Chloranil test

The chloranil test is a qualitative test for the detection of free secondary amines on resin. This test is based on the formation of dialkylaminovinyl derivatives of chloranil. In this assay a few washed resin beads are transferred to a glass tube containing 200 µL of acetone. Subsequently, 20 µL of a stock solution of chloranil is added. After shaking briefly, the ixture is left for 5 minutes at rT. Blue to green bead color means free primary amines (positive test), whereas yellowish beads indicates the absence of amines (negative test).

Chloranil stock solution is prepared by dissolving the chloranil compound in DMF at 2% (w/v). This solution should be stored at 4 °C protected from the light for maximum a month.

5.3. Conditioning of the resin

Peptides shorter than 15 residues were synthetized on Fmoc-Rink Amide (Aminomethyl)polystyrene resin with a substitution of 0.4 -0.6 mmol/g. Peptides longer than 15 residues were synthetized on ChemMatrix® resin with a loading of 0.52 mmol/g. Peptides cleaved from these resins harbor an amide function at the C-terminus.

Acid terminal peptides were synthetized on 2-chlorotrytil chloride resin (2-CTC).

In all cases, prior to the first aminoacid coupling, the resin was swelled in DCM for 30 minutes and subsequently washed with DMF (5 x 30 sec). When using Fmoc-Rink Amide (Aminomethyl) polystyrene resin, a first deprotection of the Fmoc group with 20% piperidine in DMF was required.

5.4. Loading of the first aminoacid

Depending on the type of resin used, the following procedures were followed:

	Fmoc-Rink Amide (Aminomethyl)polystyrene resin	ChemMatrix ® Rink Amide	2- chlorotrytil chloride
Coupling	4 eq Fmoc-aa-OH 4 eq HBTU 8 eq DIEA In DMF (30 min)	4 eq Fmoc-aa-OH 4 eq HBTU 8 eq DIEA In DMF (30 min)	1 eq Fmoc-aa-OH 10 eq DIEA DCM (40 min)
Capping	--	--	MeOH (0.8 mL/g resin) 10 eq DIEA DCM (15 min)

5.5. Fmoc removal

The Fmoc group was removed before each coupling by treating the resin with 20 % (v/v) piperidine in DMF (10 mL/g resin), 2 x 1 min and 1 x 5 min). After deprotection the resin was washed with DMF (3 x 1 min) and DCM (3 x 1 min).

5.6. Quantification of resin loading capacity

In order to measure the loading capacity of the resin, after coupling of the first aminoacid, piperidine washes were collected and measured by UV spectroscopy. Loading was calculated using the following equation:

$$X = (A \cdot V) / (\varepsilon \cdot m \cdot b)$$

Where *X* is the loading of the resin, *A* is the Fmoc absorbance at 290 nm, *V* is the dilution volume of solvent (mL), ε is the molar extinction coefficient of Fmoc at 290 nm (5800 M^{-1} cm^{-1}), *m* is the mass of the resin (mg) and *b* is the optical bath (typically 1 cm).

5.7. Chain elongation in manual synthesis

There are several procedures to perform aminoacid coupling. In the present thesis we used the following protocols:

Protocol A.

Step	Treatment	Conditions
1	Washes	DCM (3 x 1 min), DMF (3 x 1 min), DCM (3 x 1 min)
2	Deprotection	20 % piperidine in DMF (2 x 1 min and 1 x 5 min)
3	Washes	DMF (3 x 1 min), DCM (3 x 1 min)
4	Coupling	5 eq Fmoc-aa-OH, 5 eq HBTU, 10 eq DIEA, in DMF, 30 min
5	Washes	DMF (3 x 1 min), DCM (3 x 1 min)
6	Colorimetric test	Kaiser or chloranil test as needed

Protocol B.

Step	Treatment	Conditions
1	Washes	DCM (3 x 1 min), DMF (3 x 1 min), DCM (3 x 1 min)
2	Deprotection	20 % piperidine in DMF (2 x 1 min and 1 x 5 min)
3	Washes	DMF (3 x 1 min), DCM (3 x 1 min)
4	Coupling	5 eq Fmoc-aa-OH, 5 eq HATU, 10 eq DIEA, in DMF, 30 min
5	Washes	DMF (3 x 1 min), DCM (3 x 1 min)
6	Colorimetric test	Kaiser or chloranil test as needed

Protocol C.

Step	Treatment	Conditions
1	Washes	DCM (3 x 1 min), DMF (3 x 1 min), DCM (3 x 1 min)
2	Deprotection	20 % piperidine in DMF (2 x 1 min and 1 x 5 min)
3	Washes	DMF (3 x 1 min), DCM (3 x 1 min)
4	Coupling	5 eq Fmoc-aa-OH, 5 eq DIC, 5 eq oxyma, in DMF, 30 min
5	Washes	DMF (3 x 1 min), DCM (3 x 1 min)
6	Colorimetric test	Kaiser or chloranil test as needed

All the coupling treatments were performed at 25 °C and lasted 30 minutes. After 3 min of manual stirring, they were kept in an orbital shaker. Coupling conversion was checked by means of the appropriate colorimetric test. When the colorimetric tests do not show complete coupling, a recoupling is then required and the procedure was repeated starting at step 3. The volume of DMF used in each coupling was the required to have the aminoacid at 0.4 M concentration during coupling.

5.8. Coupling of the linker Fmoc-MeDbz

The MeDbz linker was coupled dissolving 5 equivalents in DMF to a final concentration of 0.4 M, then 5 equivalents of HBTU and 10 equivalents of DIEA were added and mixed for 30 seconds prior to the addition to the resin. The coupling was left for 30 minutes at room temperature.

5.9. Coupling of the first aminoacid after the MeDbz linker.

The first aminoacid after the linker was coupled using 5 equiv of Fmoc-AA-OH dissolved in DMF to a 0.4 M concentration, then 5 equivalents of HATU and 10 equivalents of DIEA were added. The solution was mixed briefly before addition to the resin. The coupling was left for 1h at room temperature. Some aminoacids use to require recoupling steps.

5.10. Chain elongation in automated microwave synthesis

A CEM Liberty Blue microwave peptide synthetizer was used. Drain washes were performed with DMF (2 x 5 mL) with nitrogen gas agitation. Fmoc deprotection was carried out using 10 % (w/v) piperazine and 0.1 M oxyma in a 9:1 mixture of NMP:EtOH. For the coupling step 5 equivalents of Fmoc-AA-OH to a final concetration of 0.2 M in DMF, 5 equivalents of oxyma and 5 equivalents of DIC were used. The couplings were performed at 90 °C for 3 min, expect for cysteine, histidine and arginine, which were coupled at 50 °C for 10 min. After completion of the peptide chain elongation, the peptidyl resin was washed twice with 10 mL of DMF before being removed from the reactor.

5.11. Cleavage and deprotection of side chains

Cleavage from Fmoc Rink-Amide Aminomethyl(polypropylene) and Chemmatrix® resins:

Cleavage from the resin requires harsh acidic conditions and furnishes the completely unprotected peptide. The cleavage cockatil is chosen according to the sensitiveness of the residues present in the peptide sequence and their protecting groups. For all the syntheses performed in the present thesis using this resin, after washes with DCM (3 x 1 min), the polymeric support was treated with a solution containing 95% TFA, 2.5% TIS and 2.5% H_2O (1 mL/40 mg peptidyl-resin) for 2 h. The TFA solution containing the peptide was concentrated under vacuum to a minimal volume, added over cold ether and precipitated by centrifugation. The supernatant was removed and the residue was dissolved in 50% B (HPLC buffer) and lyophilized. The resulting solid was purified on a semi-preparative column to obtain the desired peptide as a white powder.

Cleavage from 2-chlorotrytil resin chloride: cleavage from this resin is achieved by acidolytic treatment in very mild TFA conditions. After washing the resin with DCM (3 x 1 min), it is treated with a cockatil consisting in 98.1 % DCM, 1.5 % TFA (5 x 2 min) and then washed again with DCM until the resin color changes from red to the original yellow. All washes are collected in a flask containing H_2O (15 mL H_2O/200 g resin). Tfa and DCM are removed under vaccum, CAN is added and the resulting solution is liophylized. The peptide is obtained protected on the side chain of the aminoacids.

Further deprotection of the side chain after lyophilizatio could be carried out by dissolving the peptide in the cocktail: 95% TFA, 2.5% TIS and 2.5% H_2O for 2h.

5.12. Purification

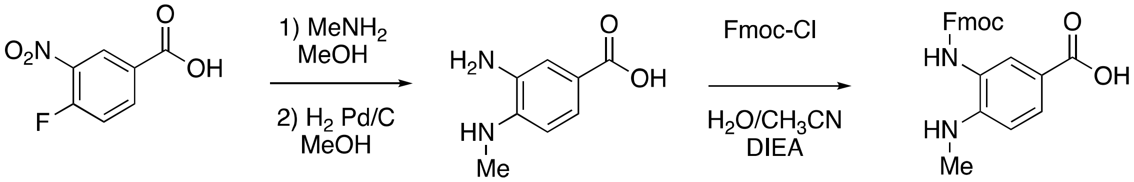
Crudes were dissolved in a solution of 0.05% TFA in H_2O/ACN (keeping the percentage of CAN to the minimum) and filtered through a 0.45 µM filter. The solution containing the crude is then purified by semi-preparative RP-HPLC. Fractions containing the peptide of interest were analyzed by HPLC, pooled and lyophilized.

5.13. Peptide characterization

Peptides identity was confirmed by HPLC-MS and their purity was checked by HPLC-PDA.

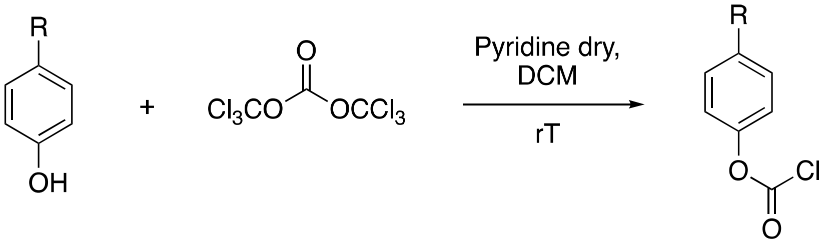
6. Synthesis of *p*CN-Phoc-MeDbz peptides and derivatives

6.1. Synthesis of Fmoc-MeDbz linker



10 g (54.0 mmol) of 3-fluoro-4-nitrobenzoic acid were dissolved in 100mL of MeOH, then 40 mL in EtOH solution (40%) of methylamine were added. After 5 h, the mixture was acidified with HCL conc. until pH 4, forming a yellow precipitate that was filtered and washed with H_2O, yielding the resulting product, 3-nitro-4-(methylamino)- benzoic acid. Next, the nitro group was reduced through catalytic hydrogenation over Pd/C. After, to selectively protect the primary amine, 8.2 g (49.3 mmol) of Fmoc-C and 9 mL (51.7 mmol) of DIEA were added to a solution of the starting material in ACN:H_2O 50 mL. The reaction was stirred for 1 h 30 min at room temperature. Once the protection was completed the mixture was acidified with HCl 1M until pH 3. The product precipitated into a white powder which was filtered and washed with ACN and Et_2O. Then, it was dried in the oven.

6.2. General synthesis of phenylchloroformates



To a solution of the corresponding phenol (1 g, 1 equivalent) in dry DCM was added triphosgene (1.3 equivalent) and the reaction was cooled in an ice bath. Dry pyridine (1 equivalent) was added dropwise and the reaction was left at room temperature for 18h. Then, water was added until all the excess of triphosgene reacted and no foam was observed. Next, 1M HCl was added until acid pH was reached. After extraction, the aqueous layer was washed with DCM (2 times x 20 mL). The organic layers were joined, dried over $MgSO_4$, filtered and concentrated under vaccum. The resulting crude (aprox. 80 % yield) was used without further purification.

6.3. General procedure for the synthesis of *p/m*X-Phoc-MeDbz peptides

Syntheses of *p/m*X-Phoc-MeDbz peptides were carried out on a 0.25 or 0.5 mmol scale using Fmoc-Rink Amide Resin (loading 0.69 mmol/g).

First aminoacid coupled on the resin was Fmoc-Gly-OH, using 5 equivalents of the aminoacid, 5 equivalents of HBTU and 10 excess equivalents of DIEA. After Fmoc deprotection with 20% piperidine in DMF for 5 minutes, we coupled the MeDbz linker following the same coupling conditions as for the first aminoacid. After Fmoc deprotection, the first aminoacid was coupled using 5 equiv of the Fmoc-AA-OH, 5 equiv of HATU and 10 eq of DIEA, the coupling was left for 1h at rT. After, peptide elongation was carried out using a 5-fold excess of the corresponding aminoacid in DMF at 0.4 M, 5-fold excess of HBTU and 10-fold excess of DIEA. The aminoacids were preactivated during 30 seconds prior to their addition to the resin, and the couplings were left for 30 min. Kaiser test was carried out after each coupling to check coupling completion. Fmoc removal was performed using 20% piperidine solution in DMF for 5 minutes. The corresponding *N*-terminal aminoacids were coupled as Boc-aminoacids following the same coupling conditions described above.

Following peptide elongation, the resin was washed with DCM (3 x 1 min) and CH_2Cl_2 (3 x 1 min) and a solution of 5 to 10 equiv of the corresponding phenylchloroformate in CH_2Cl_2 was added and gently stirred for 1h. Next, the resin was washed with CH_2Cl_2 (3 x 1 min), DMF (3 x 1 min) and finally CH_2Cl_2 (3 x 30 sec) and dried under vacuum. Finally, the peptide was cleaved from the resin with a cocktail consisting on: 95% TFA, 2.5% TIS and 2.5% H_2O (1 mL/40 mg peptidyl-resin) for 1 h. The TFA solution containing the peptide was concentrated under vacuum to a minimal volume, added over cold ether and precipitated by centrifugation. The supernatant was removed and the residue was dissolved in 50% B (HPLC buffer) and lyophilized. The resulting solid was purified on a semi-preparative column to obtain the desired peptide as a white powder.

6.4. General procedure for the synthesis of MeNbz peptides

Synthesis of MeNbz peptides was carried out on a 0.25 mmol scale using Fmoc-Rink Amide Resin (loading 0.69 mmol/mg). Following the same protocol used for the synthesis of *p/m*X-

Phoc-MeDbz peptides (section 6.3), peptides were elongated and acylated with *p*NO_2-phenyl chloroformate. After acylation, the resin was washed with CH_2Cl_2 (3 x 1 min) and DMF (3 x 1 min). Then a solution of 0.5 M DIEA in DMF was added and left stirring at room temperature for 30 minutes. Next, the resin was washed with DMF (3 x 1 min) and DCM (3 x 1 min) and dried under vacuum. The following cleavage and purification were performed following the same protocol described for the synthesis of *p/m*X-Phoc-MeDbz peptides (section 6.3).

6.5. General activation of *p*X-Phoc-MeDbz peptides in solution

*p/m*X-Phoc-MeDbz peptides were dissolved to 2 mM concentration in NCL buffer consisting on 6 M Gn·HCl, 0.2 M phosphate, 20 mM TCEP and 100 mM MPOH, pH was adjusted to 7.5 with a solution of 1 M NaOH $_{(aq)}$. The reaction was left at room temperature and monitored by HPLC and ESI/MS.

7. NCL and KCL

7.1. General protocol for NCL:

NCL reactions are carried out by dissolving the *p*CN-Phoc-MeDbz or MeNbz peptide to a 2 mM and adding 1.1 equiv of the cysteinyl peptide in a NCL buffer containing 6 M Gn·HCl, 0.2 M phosphate, 20 mM TCEP and 100 mM MPOH. When performing the ligation with MeNbz peptides the pH of the reaction was adjusted to 6.5-7, whereas when the ligation is performed with *p*CN-Phoc-MeDbz the pH was raised to 7.5. The ligation is monitored by HPLC.

Workup: when both starting fragments were consumed, the reaction mixture was acidified with HCl until pH 1-2. Then, the MPOH thiol was extracted with Et_2O (1 mL reaction/10 mL Et_2O), and the aqueous layer containing the peptides was purified directly on the semi-preparative HPLC. Pool fractions containing the ligated product were combined and lyophilized.

7.2. General procedure for the synthesis of MESNa thioester peptides from *p*CN-Phoc-MeDbz

Lyophilized *p*CN-Phoc-MeDbz peptides were dissolved to a 2 mM concentration in a buffer containing 6 M Gn·HCl, 0.2 M phosphate, 20 mM TCEP, and 100 mM MESNa. The pH is raised to 7.5 with a solution of 1 M $NaOH_{(aq)}$, and the reaction left at room temperature and monitored by HPLC and ESI-MS.

7.3. Racemization experiments:

Model peptide LSYRAY-*p*CN-Phoc-MeDbz (0.4 mg, 0.35 µmol) and LSYRAy-*p*CN-Phoc-MeDbz peptide (0.4 mg, 0.35 µmol) were dissolved respectively in 500 µL of degassed NCL buffer containing 100 mM MPAA, then model CRAFS peptide (0.8 mg, 1.37 µmol) was added into each reaction mixture and pH was adjusted to 7.5. The reaction mixtures was left at room temperature and monitored by HPLC using a gradient from 0% ACN to 46% ACN in 20 minutes.

7.4. KCL experiment between MeNbz, *p*CN-Phoc-MeDbz and cysteinyl peptide:

LSYRA-MeNbz peptide (1.6 mg, 1.7 µmol) and CSYRAF-*p*CN-Phoc-MeDbz peptide (1.9 mg, 1.7 µmol) were dissolved in 850 µL of previously degassed NCL buffer containing 100 mM MPOH, the pH of the reaction was adjusted to 7.0. The mixture was stirred at room temperature and monitored by HPLC. After 3 h 30 minutes all fragments were consumed and the desired product formed. In order to perform the following NCL, cysteine peptide CRAFS (1.3 mg, 2.2 µmol) was added and pH was adjusted to 7.5 with NaOH 1M. The reaction mixture was stirred at room temperature and monitored by HPLC. The ligated product was obtained after 10 h.

8. Common reactions on chemical protein synthesis

8.1. General procedure for Acm removal:

Cys(Acm)-Peptide was dissolved to a 2mM concentration in a degassed buffer containing 1:1 AcOH:H_2O and 50 mM of silver triflate. The reaction mixture was wrapped with aluminum foil and left overnight at room temperature. Once the cysteine was completely deprotected, the reaction was quenched by transferring it in a 5 mL solution of 6M GnHCl, 0.2 M NaPi, 0.25M DTT and pH was adjusted to 7.0. The mixture was left stirring in ice bath for 30min, a white precipitate was observed. Then it was centrifuged at 3500 rpm, the supernatant containing the peptide was kept and the precipitate was washed 3 times with the work-up buffer, and centrifuged every time. All supernatants were joined and directly purified on semi-preparative HPLC.

8.2. General procedure for radical-based desulfurization:

Cys-Peptide was dissolved to 0.5 mM concentration in desulfurization buffer containing 6M Gn·HCl, 0.2 M phosphate, 250 mM MESNa, 250 mM TCEP and 100 mM VA-044. The buffer was degassed and pH adjusted to 5 with triethylamine. The reaction was left at 37 °C for 2h and monitored by UHPLC/ESI-MS. Once desulfurization was completed the reaction was quenched by acidification with TFA:H_2O (50:50) until pH 1 was reached. The mixture was directly purified by semi-preparative HPLC.

Desulfurization of peptide fragments containing more than one Cys, were carried out at 60 °C. Moreover, when a thiazolidine is present, care must be taken to ensure not undesired removal.

8.3. General procedure for thiazolidine removal:

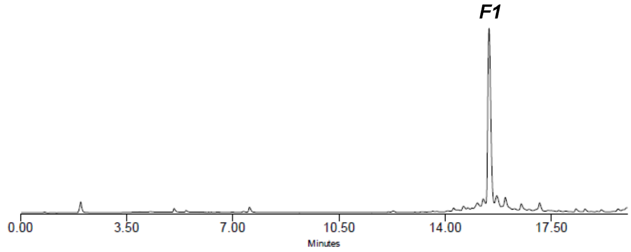
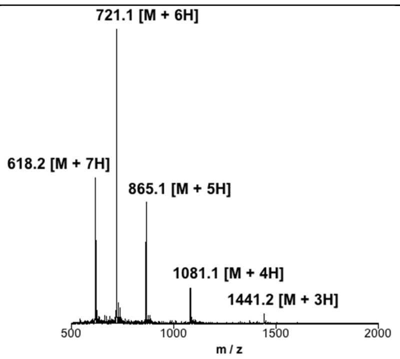
Thz-peptide was dissolved to 2 mM concentration in a degassed buffer containing 6 M Gn·HCl, 0.2 M NaPi, 20 mM TCEP and 0.2 M methoxylamine. The pH was adjusted to 4.0 and left stirring at rT. The reaction was monitored by HPLC and ESI-MS. Once the Thz was completely removed the reaction mixture was directly purified on the semi-preparative HPLC.

9. Synthesis of Palm-ShhN-biotin

9.1. Peptide fragment syntheses

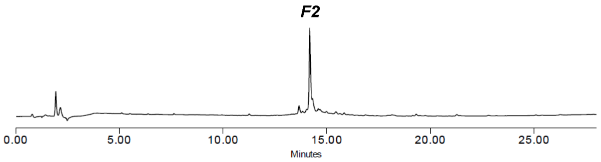
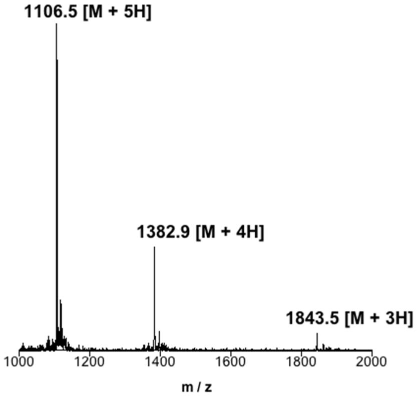
9.1.1. Synthesis of F1: Palm-Ala^{1}-Gly^{34}-MeDbz-Gly

Starting from 230 mg (0.12 mmol) of Chemmatrix® resin and following the general procedure for the synthesis of *p*CN-Phoc-MeDbz peptides, 62 mg (46%) of crude were recovered after cleavage. Palmitic acid was coupled following standard coupling procedure, but dissolved in a mixture of DCM:DMF (1:1).

<table>
<tr><td colspan="2">Palm – AGPGRGFGKRRHPKKLTPLAYKQFIPNVAEKTLG-<i>p</i>CN-Phoc-MeDbz-Gly</td></tr>
<tr><td>Exact mass: 4320.5</td><td>Yield: 46 % crude</td></tr>
<tr><td>
</td><td>
</td></tr>
</table>

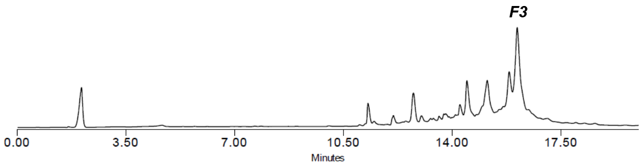
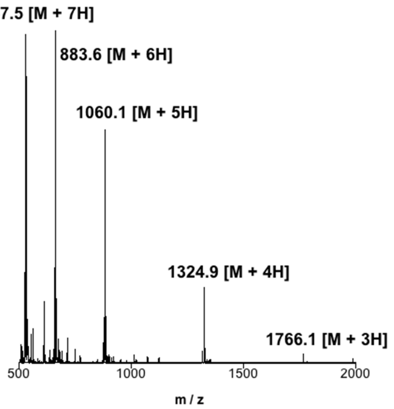
9.1.2. Synthesis of F2: Cys^{35}-Arg^{78}-pCN-Phoc-MeDbz-Gly

Starting from 2 g resin (1.0 mmol) of Chemmatrix® resin and following the general procedure for the synthesis of *p*CN-Phoc-MeDbz peptides, 750 mg (63 %) of crude were recovered after cleavage. After purification, 250 mg of pure peptide (17 %) yield were obtained.

CSGRYEGKISRNSERFKELTPNYNPDIIFKDEENTGADRLMTQR-*p*CN-Phoc-MeDbz-Gly	
Exact mass: 5527.6	**Yield:** 17 %
	

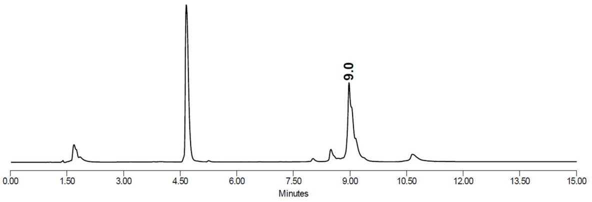
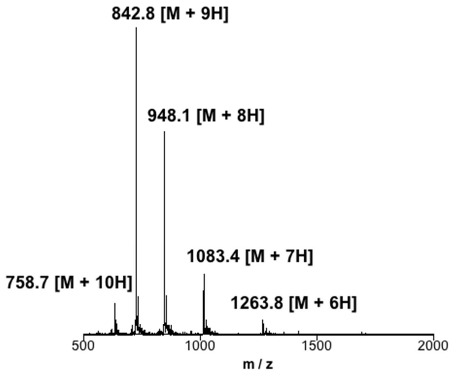
9.1.3. *Synthesis of F3: Thz^{79}-Arg^{121}-MeDbz-Gly*

Starting from 4 g resin (2.1 mmol) of Chemmatrix® resin and following the general procedure for the synthesis of *p*CN-Phoc-MeDbz peptides, 1.95 g (89 %) of crude were recovered after cleavage.

ThzKDKLNALAISVMNQWPGVKLRVTEGWDEDGHHSEESLHYEGR-*p*CN-Phoc-MeDbz-Gly	
Exact mass: 5295.5	**Yield:** 89 % crude
	

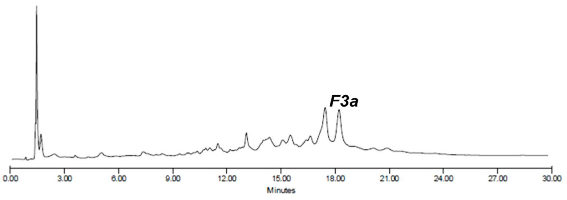
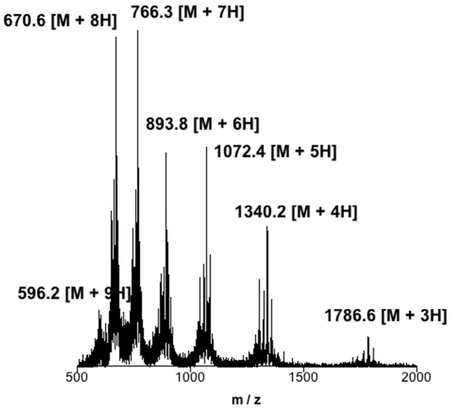
9.1.4. *Synthesis of F3 thioester: Thz^{79}-Arg^{121}-S(CH$_2$)$_2$SO$_3$Na*

87 mg (14.8 mmol) of crude F3 were dissolved in 10 mL of buffer containing 6 M Gn·HCl, 0.2 M phosphate, 20 mM TCEP and 100 mM MESNa. The pH was adjusted to 7.5 using a solution of 1 M NaOH$_{(aq)}$. The reaction was left at rT. After purification, 45 mg (54 % yield) of the product were obtained.

ThzKDKLNALAISVMNQWPGVKLRVTEGWDEDGHHSEESLHYEGR-S(CH$_2$)$_2$-SO$_3$Na	
Exact mass: 5070.3	**Yield:** 54 %
	

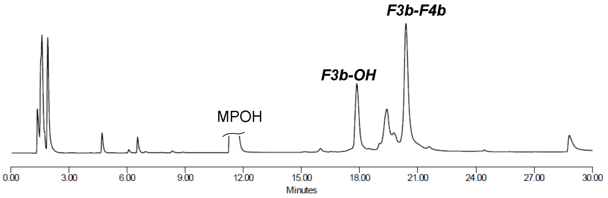
9.1.5. Synthesis of F3a: Cys^{79}(Acm)-Arg^{121}-pCN-Phoc-MeDbz-Gly

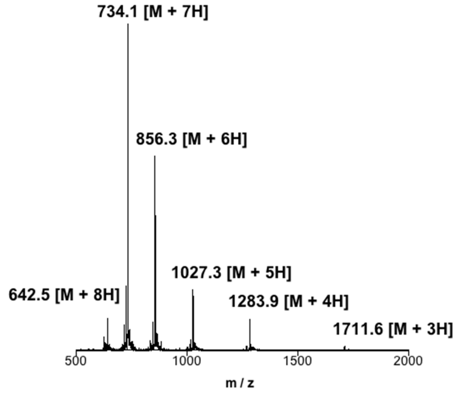
Starting from 1.7 g resin (0.88 mmol) of Chemmatrix® resin and following the general procedure for the synthesis of *p*CN-Phoc-MeDbz peptides, 500 mg (70 %) of crude were recovered after cleavage.

Cys(Acm)KDKLNALAISVMNQWPGVKLRVTEGWDEDGHHSEESLHYEGR-*p*CN-Phoc-MeDbz-Gly	
Exact mass: 5356.9	**Yield:** 70 % crude
	

9.1.6. Synthesis of F3a thioester: Cys^{79}(Acm)-Arg^{121}-S(CH$_2$)$_2$SO$_3$H

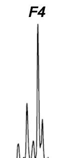
150 mg (0.03 mmol) of crude F3a were dissolved in 15 mL of buffer containing 6 M Gn·HCl, 0.2 M phosphate, 20 mM TCEP and 100 mM MESNa. The pH was adjusted to 7.5 using a solution of 1 M NaOH$_{(aq)}$. The reaction was left at rT and purified.

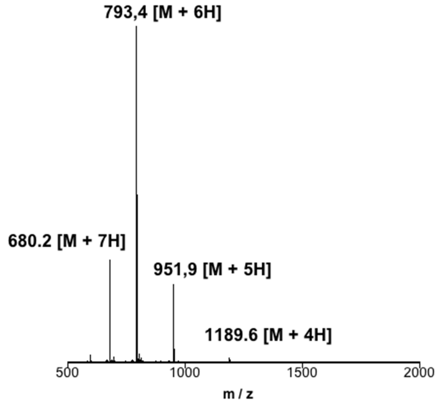
<table>
<tr><td colspan="2">Cys(Acm)KDKLNALAISVMNQWPGVKLRVTEGWDEDGHHSEESLHYEGR-S$(CH_2)_2$-SO_3Na</td></tr>
<tr><td>Exact mass: 5131.7</td><td>Yield: not determined</td></tr>
<tr><td colspan="2">


</td></tr>
</table>

9.1.7. Synthesis of F4: Cys^{122}-His^{159}-pCN-Phoc-MeDbz-Gly

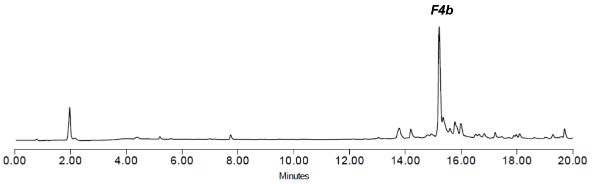
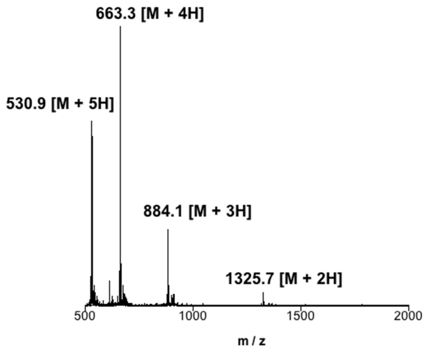
Starting from 600 mg (0.2 mmol) of Chemmatrix® resin and following the general procedure for the synthesis of *p*CN-Phoc-MeDbz peptides, 618 mg of crude (46 % recovery) were recovered after cleavage.

<table>
<tr><td colspan="2">CVDITTSDRDRSKYGMLARLAVEAGFDWVYYESKAHIH-<i>p</i>CN-Phoc-MeDbz-Gly</td></tr>
<tr><td>Exact mass: 4751.24</td><td>Yield: 46 % crude</td></tr>
<tr><td colspan="2">


</td></tr>
</table>

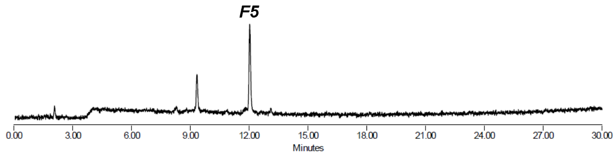
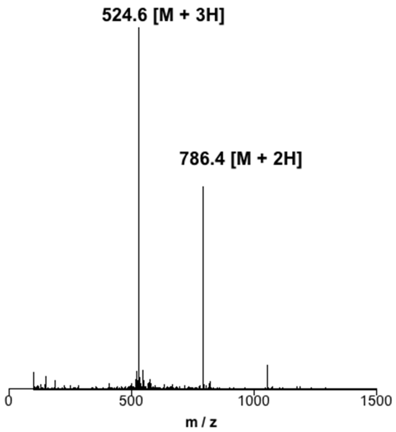
9.1.8. Synthesis of F4a: Cys^{122}-Leu^{141}-pCN-Phoc-MeDbz-Gly

Starting from 4 g resin (2.1 mmol) of Chemmatrix® resin and following the general procedure for the synthesis of *p*CN-Phoc-MeDbz peptides, 1.36 g (52 %) of crude were recovered after cleavage. After purification, 225 mg of pure peptide (17 %) yield were obtained.

CVDITTSDRDRSKYGMLARL-*p*CN-Phoc-MeDbz-Gly	
Exact mass: 2649.3	**Yield:** 17 %
 	

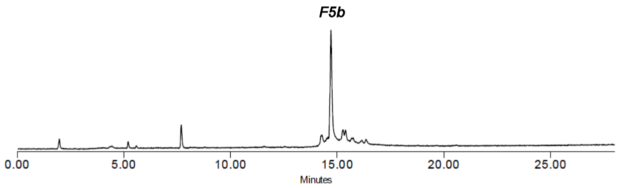
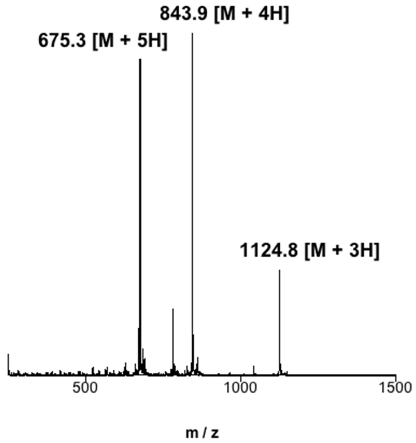
9.1.9. *Synthesis of F5: Cys^{160}-Gly^{173}-MeNbz*

Starting from 1 g (1.6 mmol) of Chemmatrix® resin and following the general procedure for the synthesis of *p*CN-Phoc-MeDbz peptides, 320 mg of protected crude were recovered after cleavage.

CSVKAENSVAAKSGG-MeNbz	
Exact mass: 1755.8	**Yield:** not determined
 	

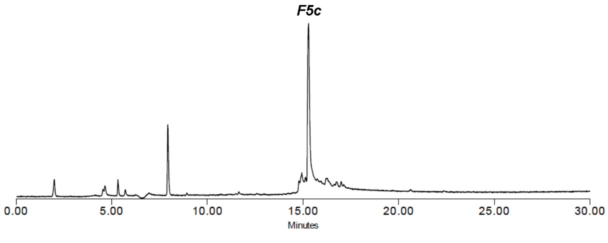
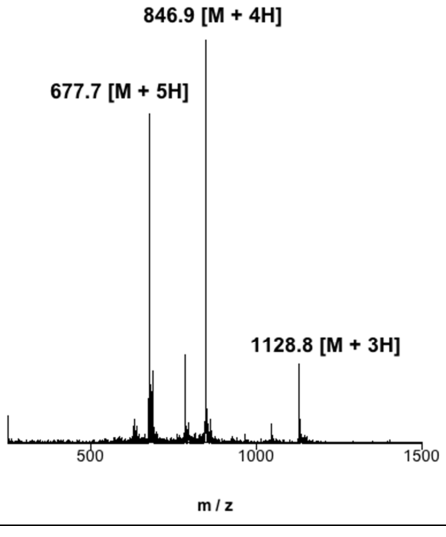
9.1.10. *Synthesis of F5a: Cys^{142}-Val^{168}-pCN-Phoc-MeNbz*

Starting from 2 g (1.0 mmol) of Chemmatrix® resin and following the general procedure for the synthesis of *p*CN-Phoc-MeDbz peptides, 317 mg (67 % recovered) of crude were recovered after cleavage.

CVEAGFDWVYYESKAHIHC(Acm)SVKAENSV-MeNbz	
Exact mass: 3371.51	**Yield:** 67 % crude
	

9.1.11. Synthesis of F5b: Thz^{142}-Val^{168}-pCN-Phoc-MeNbz

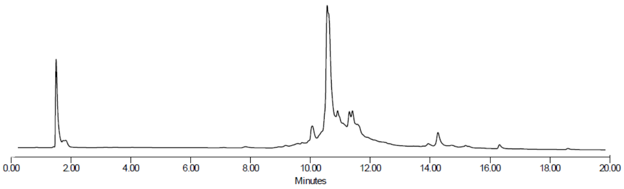
Starting from 2 g (1.0 mmol) of Chemmatrix® resin and following the general procedure for the synthesis of *p*CN-Phoc-MeDbz peptides, 317 mg (67 % recovered) of crude were recovered after cleavage.

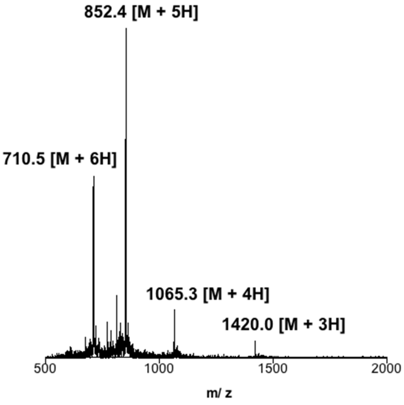
ThzVEAGFDWVYYESKAHIHC(Acm)SVKAENSV-MeNbz	
Exact mass: 3383.51	**Yield:** 67 % crude
	

9.1.12. Synthesis of F5c: Cys^{142}-Lys^{175}-PEG$_2$-Biotin

For the synthesis of F5d fragment, it was carried out in a 0.5 mmol scale (0.26 mmol) in a Chemmatrix® resin. The first aminoacid coupled to the resin was Fmoc-Lys(ivDde)-OH. Next, the peptide was elongated using the same protocol described before. After, the iv(Dde) protecting group of the lysine was removed by washings with hydrazine (4% in DMF) (3 x 1 h).

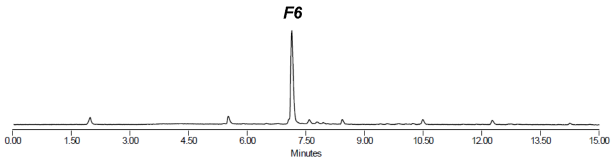
Next, two consecutive PEG moieties were coupled using 5 equiv of Fmoc-PEG-OH, 5 equiv HATU, 10 equiv DIEA, for 1h at 60 °C. After Fmoc deprotection, OSu-Biotin (5 equiv) was coupled using 5 equiv of HATU and 10 equiv of DIEA at 60 °C for 1 h. After peptide cleavage following the standard protocol, and purification, the peptide was obtained in 11 % yield.

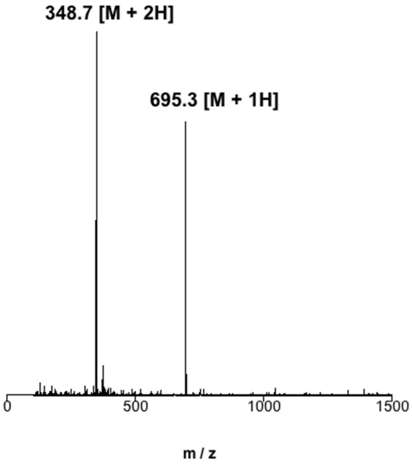
<table>
<tr><td colspan="2">CVEAGFDWVYYESKAHIHC(Acm)SVKAENSVAAKSGG-PEG$_2$-Biotin</td></tr>
<tr><td>Exact mass: 4257.0</td><td>Yield: 11 %</td></tr>
<tr><td colspan="2">


</td></tr>
</table>

9.1.13. *Synthesis of F6: Cys^{169}-Gly^{173}-MeNbz*

Starting from a 800 mg (0.4 mmol) of 2-CTC resin and and following the general procedure for the synthesis of MeNbz peptides, 368 mg of side-chain protected peptide crude were recovered.

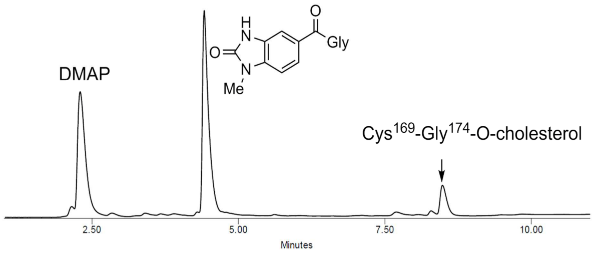
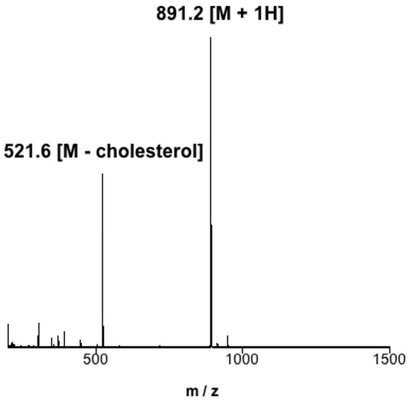
<table>
<tr><td colspan="2">CAKSG-MeNbz</td></tr>
<tr><td>Obs. mass: 694.3</td><td>Yield: not determined</td></tr>
<tr><td colspan="2">


</td></tr>
</table>

9.1.13.2.1. *Assembling of the Gly--cholesterol building block with the protected Cys^{169}-Gly^{173}-MeNbz*

100 mg (0.083 mmol) of protected F6 and 1 equiv of DMAP were dissolved in 6 mL of anhydrous DMF. Then a solution of gly-O-cholesterol (232 mg, 0.42 mmol) and DIEA (0.42 mL, 2.5 mmol) in anh. THF, was added. The mixture under N_2 atm was stirred at rT for 18 h. Next

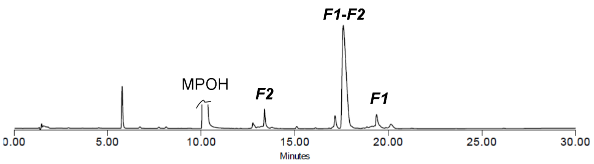
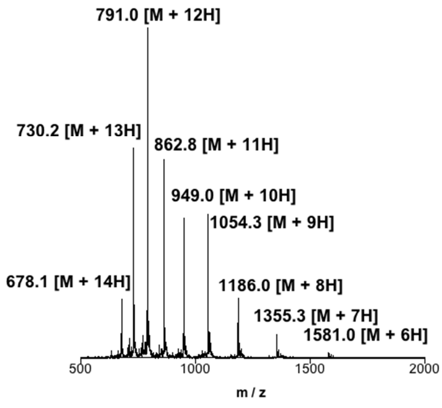
the solvents are removed under vaccum and the product is deprotected using a mixture of 95 % TFA, 2.5 % TIS, 2.5 % H_2O for 45 min. After, the mixture is precipitated on cold ether, and centrifuged. The pellet containing the peptide is dissolved in a mixture of 50% ACN:H_2O and lyophilized. The product was obtained in 33 % yield.

CAKSG-cholesterol	
Obs. mass: 890.1	**Yield:** 33 % crude
	

9.2. Fragment ligations

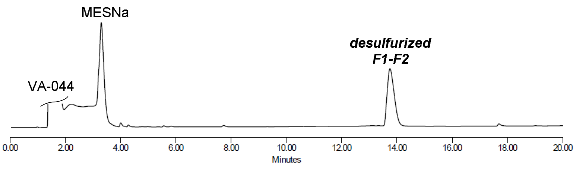
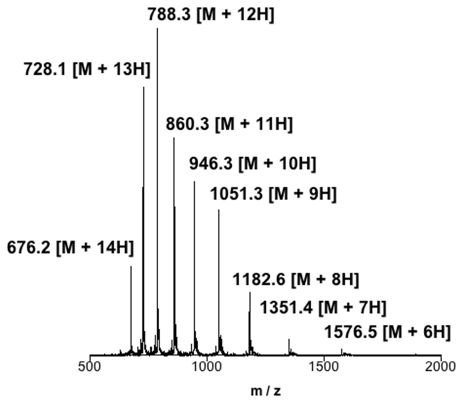
9.2.1. *NCL between F1 and F2: Palm-Ala^{1}-Gly^{34}-pCN-Phoc-MeDbz and Cys^{35}-Arg^{78}-pCN-PhCB-MeDbz-Gly*

45 mg (8.5 µmol) of Palm-Ala^{1}-Gly^{34}-MeDbz were dissolved in 6 mL of NCL buffer containing 100 mM of MPOH, and pH was raised to 7.5. When MeNbz was formed, 59.8 mg (8.5 µmol) of Cys^{35}-Arg^{78}-*p*CN-Phoc-MeDbz were added, and pH was adjusted to 7.0. The ligation was left at rT for 4 h and the workup was carried out following the protocol described above. After purification 22 mg (2.3 µmol, 22 % yield) of ligated Palm-Ala^{1}-Cys^{35}-Arg^{78}-*p*CN-PhCB-MeDbz-Gly were obtained.

Palm-AGPGRGFGKRRHPKKLTPLAYKQFIPNVAEKTLGCSGRYEGKISRNSERFKELTPNYNPDIIFKDEENT GADRLMTQR-*p*CN-Phoc-MeDbz-Gly	
Exact mass: 9480.0	**Yield:** 72 %
	

9.2.2. *Desulfurization of F1-F2: Palm-Ala^{1}-Cys^{35}-Arg^{78}-pCN-PhCB-MeDbz-Gly:*

Desulfurization was carried out following the general desulfurization procedure. 24 mg (2.33 µmol) of Palm-Ala^{1}-Cys^{35}-Arg^{78}-*p*CN-PhCB-MeDbz-Gly were dissolved in 1.5 mL of buffer B. The reaction was heated up to 37 °C for 2 h and monitored by UHPLC/ESI-MS. Upon completion, the solution was acidified with $HCl_{(aq)}$ 10% until pH 2, and purified by semi-preparative HPLC. After purification 11 mg (1.07 µmol, 60% yield)) of Palm-Ala^{1}-Ala^{35}-Arg^{78}-*p*CN-PhCB-MeDbz-Gly were obtained.

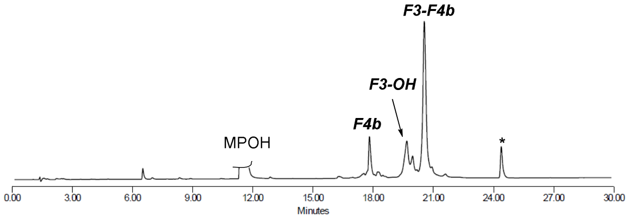
Palm-AGPGRGFGKRRHPKKLTPLAYKQFIPNVAEKTLGASGRYEGKISRNSERFKELTPNYNPDIIFKDEENTGADRLMTQR-*p*CN-Phoc-MeDbz-Gly	
Mass: 9448.0	**Yield:** 60 %
	

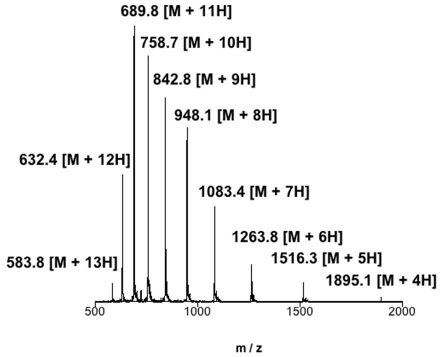
9.2.3. *KCL between F3 and F4a: Thz^{79}-Arg^{121}-MESNa and Cys^{122}-Leu^{141}-pCN-PhCB-MeDbz-Gly*

21 mg (3.72 µmol) of Cys^{122}-Leu^{141}-*p*CN-PhCB-MeDbz-Gly were dissolved in 2 mL of NCL buffer, once it was completely dissolved the solution was added onto 30 mg (5.31 µmol) of Thz^{79}-Arg^{121}-MESNa. Then, pH was adjusted to 6.5 and the reaction was left at rt for 2h. After purification 31 mg (3.55 µmol, 66 % yield) of ligated Thz^{79}-Cys^{122}-Leu^{141}-*p*CN-PhCB-MeDbz-Gly were obtained. During ligation around 10% of Thz^{79}-Arg^{121}-MESNa peptide suffers hydrolysis affording the Thz^{79}-Arg^{121}-OH peptide, which is no longer able to undergo NCL.

ThzKDKLNALAISVMNQWPGVKLRVTEGWDEDGHHSEESLHYEGRCVDITTSDRDRSKYGMLARL-*p*CN-Phoc-MeDbz-Gly

Obs. mass: 7576.6 **Yield:** 66 %



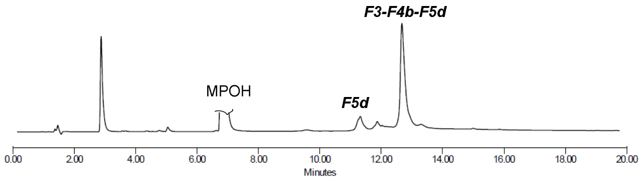


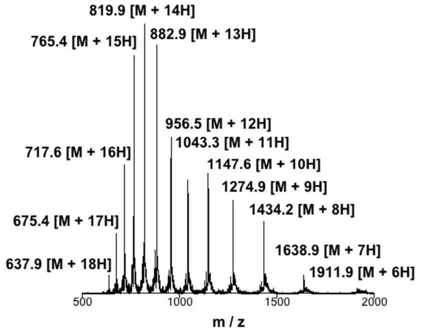
9.2.4. KCL between F3-F4a and F5c: Thz^{79}-Cys^{122}-Leu^{141}-pCN-PhCB-MeDbz-Gly and Cys^{142}-Cys^{160}(Acm)-Lys^{175}-PEG$_2$-Biotin

23 mg (5.42 µmol) of Cys^{142}-Lys^{175}-PEG-Biotin were dissolved in 3.6 mL of NCL buffer, and this solution was added onto 46 mg (5.41 µmol) of Thz^{79}-Cys^{122}-Leu^{141}-*p*CN-PhCB-MeDbz-Gly. Then, the pH was adjusted to 7.5 and the reaction was left at rt overnight. After purification 35 mg (2.82 µmol,) of ligated Thz^{79}-Cys^{122}-Cys^{142}-Cys^{160}(Acm)-Lys^{175}-PEG$_2$-Biotin were obtained.

ThzKDKLNALAISVMNQWPGVKLRVTEGWDEDGHHSEESLHYEGRCVDITTSDRDRSKYGMLARLCVEAGFDWVYYESKAHIHC(Acm)SVKAENSVAAKSGG-PEG$_2$-Biotin

Mass: 11465.5 **Yield:** 56 %

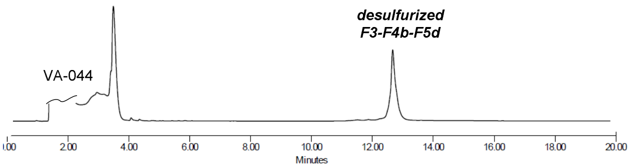


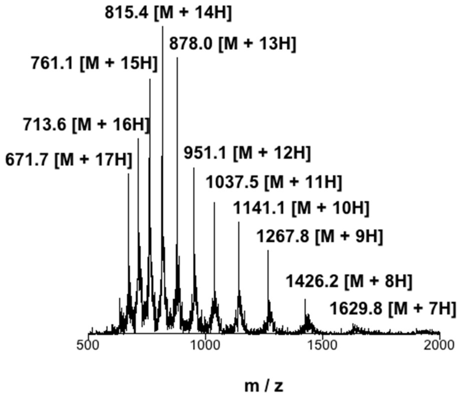


9.2.5. Desulfurization of F3-F4a-F5c: Thz^{79}-Cys^{122}-Cys^{142}-Cys^{160}(Acm)-Lys^{175}-PEG$_2$-Biotin:

15 mg (1.19 µmol) of Thz^{79}-Cys^{122}-Cys^{142}-Cys^{160}(Acm)-Lys^{175}-PEG-Biotin were dissolved in 2.4 mL of buffer B and the pH was adjusted to 6.5. The reaction was heated up to 60ºC for 45 minutes and monitored by UHPLC/ESI-MS. Upon completion, the solution was acidified with $HCl_{(aq)}$ 10%

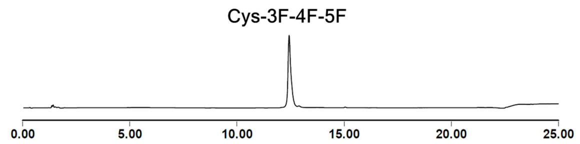
until pH 2, and purified by semi-preparative HPLC. Following purification 18 mg (1.43 µmol, 66% yield) of Thz^{79}-Ala^{122}-Ala^{142}-Lys^{175}-PEG$_2$-Biotin were obtained.

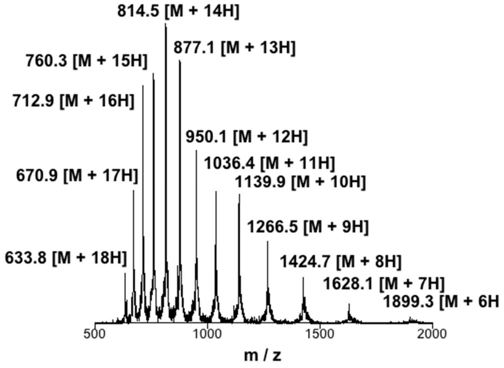
<table>
<tr><td colspan="2">ThzKDKLNALAISVMNQWPGVKLRVTEGWDEDGHHSEESLHYEGRAVDITTSDRDRSKYGMLARLAVEAG FDWVYYESKAHIHC(Acm)SVKAENSVAAKSGG-PEG$_2$-Biotin</td></tr>
<tr><td>Mass: 11401.6</td><td>Yield: 66 %</td></tr>
<tr><td colspan="2">


</td></tr>
</table>

9.2.6. Thz removal of F3-F4a-F5c: Thz^{79}-Ala^{122}-Ala^{142}-Cys^{160}(Acm)-Lys^{175}-PEG$_2$-Biotin:

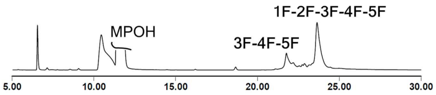
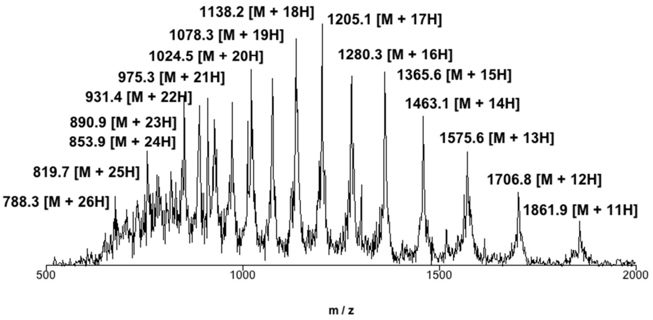
14 mg (1.11 µmol) of Thz^{79}-Ala^{122}-Ala^{142}- Cys^{160}(Acm)-Lys^{175}-PEG-Biotin were dissolved in 750 µL of degassed buffer containing 6 M GnHCl, 0.2 M NaPi, 20 mM TCEP, 0.2 M methoxylamine, at pH 4.5. The mixture was stirred at rt for Xh and monitored by HPLC and ESI-MS. When all Thz was removed the mixture was directly purified by semi-preparative HPLC. 9 mg (0.71 µmol, 64% yield) of Cys^{79}-Ala^{122}-Ala^{142}-Cys^{160}(Acm)-Lys^{175}-PEG$_2$-Biotin were obtained.

<table>
<tr><td colspan="2">CKDKLNALAISVMNQWPGVKLRVTEGWDEDGHHSEESLHYEGRAVDITTSDRDRSKYGMLARLAVEAGF DWVYYESKAHIHC(Acm)SVKAENSVAAKSGG-PEG$_2$-Biotin</td></tr>
<tr><td>Mass: 11389.6</td><td>Yield: 64 %</td></tr>
<tr><td colspan="2">


</td></tr>
</table>

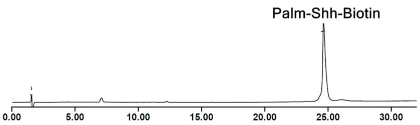
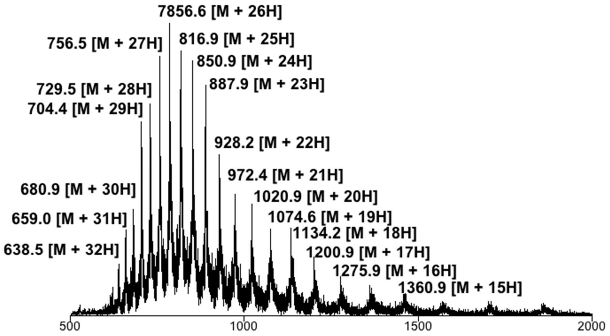
9.2.7. KCL between F1-F2 and F3-F4a-F5c: Palm-Ala^{1}-Arg^{78}-pCN-PhCB-MeDbz-Gly and Cys^{79}-Cys^{160}(Acm)-Lys^{175}-PEG$_2$-Biotin:

26 mg (2.05 µmol) of Cys^{79}-Ala^{122}-Ala^{142}-Cys^{160}(Acm)-Lys^{175}-PEG$_2$-Biotin were dissolved in 2 mL of NCL buffer. Once it was completely dissolved, the mixture was added onto 25 mg (2.43 µmol) of Palm-Ala^{1}-Ala^{35}-Arg^{78}-*p*CN-PhCB-MeDbz-Gly and the pH was adjusted to 7.5. The reaction was left at rt for 18h. After HPLC purification 26 mg (1.15 µmol, 49% yield) of Palm-Ala^{1}-Ala^{35}-Arg^{78}- Cys^{79}-Ala^{122}-Ala^{142}-Cys^{160}(Acm)-Lys^{175}-PEG$_2$-Biotin were obtained.

Palm-AGPGRGFGKRRHPKKLTPLAYKQFIPNVAEKTLGASGRYEGKISRNSERFKELTPNYNPDIIFKDEENTGADRLMTQRCKDKLNALAISVMNQWPGVKLRVTEGWDEDGHHSEESLHYEGRAVDITTSDRDRSKYGMLARLAVEAGFDWVYYESKAHIHC(Acm)SVKAENSVAAKSGG-PEG$_2$-Biotin	
Mass: 20469.4	**Yield:** 49 %
  	

9.2.8. Acm removal of F1-F2-F3-F4a-F5c: Palm-Ala^{1}-Arg^{78}-Cys^{79}-Cys^{160}(Acm)-Lys^{175}-PEG$_2$-Biotin:

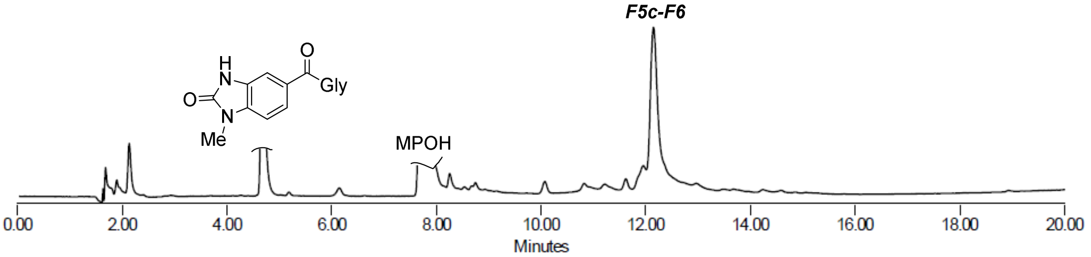
Acm was removed following the general procedure. 17 mg (0.75 µmol) of Palm-Ala^{1}-Arg^{78}-Cys^{79}-Cys^{160}(Acm)-Lys^{175}-PEG$_2$-Biotin were dissolved in 1 mL of a buffer containing 1:1 AcOH:H$_2$O, 50 mM silver triflate. The reaction was left overnight. After performing the work-up and further purification 11 mg (0.48 µmol, 65% yield) of the final product were obtained.

<table>
<tr><td colspan="2">Palm-AGPGRGFGKRRHPKKLTPLAYKQFIPNVAEKTLGASGRYEGKISRNSERFKELTPNYNPDIIFKDEENTGADRLMTQRCKDKLNALAISVMNQWPGVKLRVTEGWDEDGHHSEESLHYEGRAVDITTSDRDRSKYGMLARLAVEAGFDWVYYESKAHIHCSVKAENSVAAKSGG-PEG2-Biotin</td></tr>
<tr><td>Obs. mass: 20398.4</td><td>Yield: 65 %</td></tr>
<tr><td colspan="2">  </td></tr>
</table>

9.3. KCL of other strategies

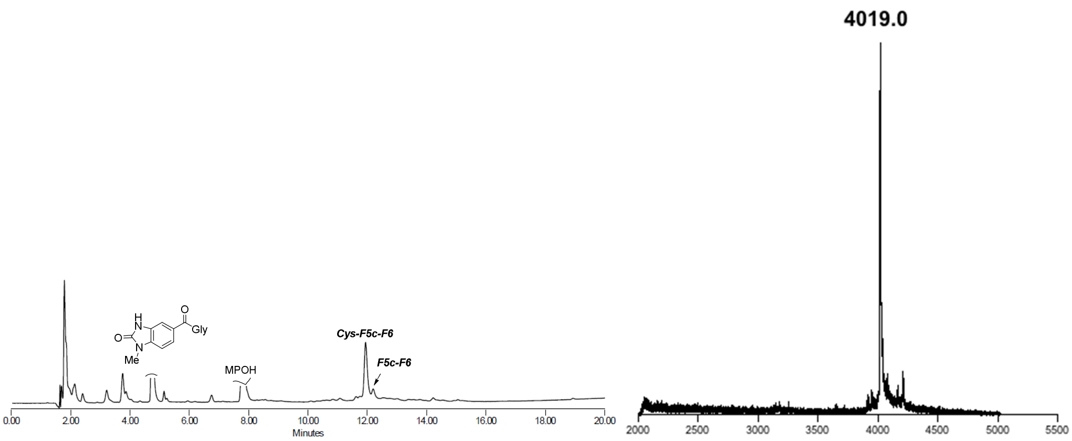
9.3.1. KCL between F5a and F6 and methoxylamine removal

100 mg (27.0 µmol) of F5b and 26 mg (27.0 µmol) of F6-O-cholesterol fragments were dissolved in 15 mL NCL buffer containing 100 mM MPOH at 30 °C and pH was adjusted to 7.0. The ligated product was obtained after 18 hours of reaction. Next, the work up of the ligation was carried out following the general procedure for NCL. Next, the Thz removal was carried out upon the addition on the aqueous layer containing the peptide product, of TCEP until 20 mM and methoxylamine until 0.3 M. The reaction was stirred at rT and the free cysteine peptide was observed after 18 hours. Once the reaction was completed, the reaction mixture was centrifuged at 5000 rpm, the, the supernatant was discarded and the pellet containing the peptide was dissolved in H_2O (containing 0.045 % TFA) and centrifuged again to remove the salts. The pellet was finally resuspended in H_2O:ACN (1:1) and lyophilized. The crude was obtained in 62 % yield (67 mg, 16.6 µmol) and used without further purification.

<table>
<tr><td colspan="2">ThzVEAGFDWVYYESKAHIHC(Acm)SVKAENSVCAKSG-cholesterol</td></tr>
<tr><td>Obs. mass: 4027.7</td><td>Yield: not determined</td></tr>
<tr><td colspan="2"> </td></tr>
<tr><td colspan="2"></td></tr>
</table>

CVEAGFDWVYYESKAHIHC(Acm)SVKAENSVCAKSG-cholesterol

Obs. mass:4015.7 | **Yield:** 62 %

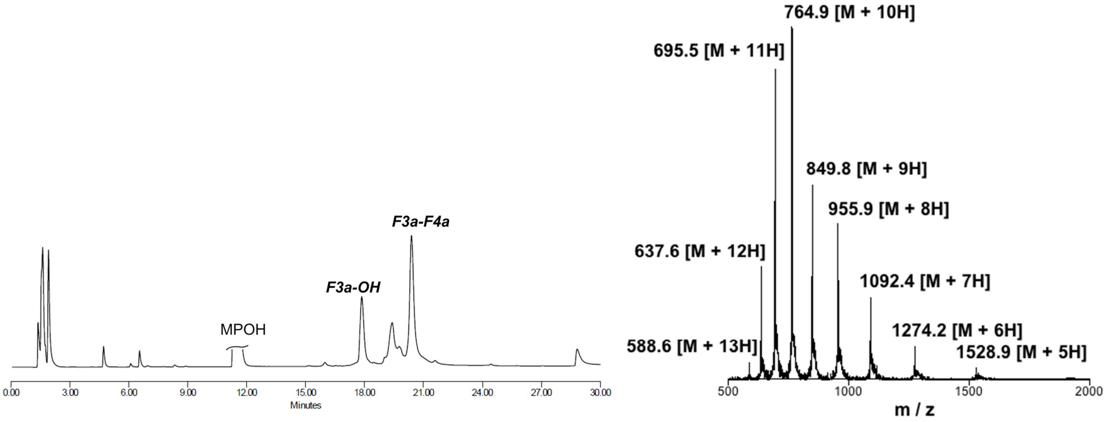


9.3.2. *KCL between F3a and F4a: Cys(Acm)^{79}-Arg^{121}-MESNa and Cys^{122}-Leu^{141}-pCN-PhCB-MeDbz-Gly*

36 mg (13.6 µmol) of Cys^{122}-Leu^{141}-*p*CN-PhCB-MeDbz-Gly were dissolved in 6 mL of NCL buffer, once it was completely dissolved the solution was added onto 50 mg (9.7 µmol) of Thz^{79}-Arg^{121}-MESNa. Then, pH was adjusted to 6.5 and the reaction was left at rt for 2h. After purification 38 mg (4.9 µmol, 51 % yield) of ligated Cys(Acm)^{79}-Cys^{122}-Leu^{141}-*p*CN-PhCB-MeDbz-Gly were obtained.

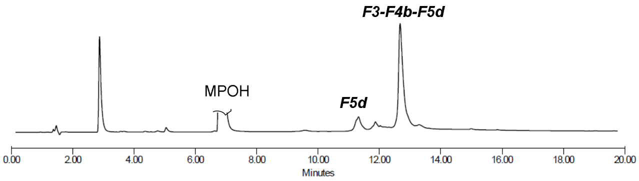
Cys(Acm)KDKLNALAISVMNQWPGVKLRVTEGWDEDGHHSEESLHYEGRCVDITTSDRDRSKYGMLARL-*p*CN-Phoc-MeDbz

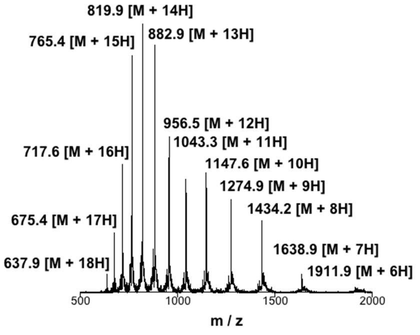
Exact mass: 7639.5 | **Yield:** 51 %



9.3.3. *KCL between F3a-F4a and F5c : Cys(Acm)^{79}-Arg^{121}-Cys^{122}-Leu^{141}-pCN-PhCB-MeDbz-Gly and Cys^{142}- Cys^{160}(Acm)-Lys^{175}-PEG$_2$-Biotin*

19 mg (4.5 µmol) of Cys^{142}-Lys^{175}-PEG-Biotin were dissolved in 3.0 mL of NCL buffer, and this solution was added onto 46 mg (4.9 µmol) of Cys(Acm)^{79}-Cys^{122}-Leu^{141}-*p*CN-PhCB-MeDbz-Gly. Then, the pH was adjusted to 7.5 and the reaction was left at rt overnight. After purification 25 mg (2.2 µmol,) of ligated product were obtained.

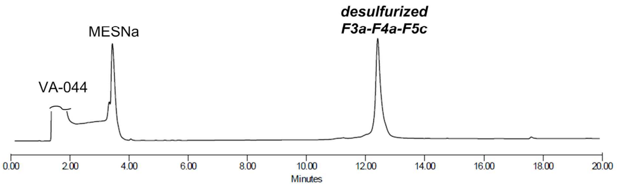
<table>
<tr><td colspan="2">Cys(Acm)KDKLNALAISVMNQWPGVKLRVTEGWDEDGHHSEESLHYEGRCVDITTSDRDRSKYGMLARLC VEAGFDWVYYESKAHIHC(Acm)SVKAENSVAAKSGG-PEG₂-Biotin</td></tr>
<tr><td>Mass: 11465.5</td><td>Yield: 48 %</td></tr>
<tr><td colspan="2">


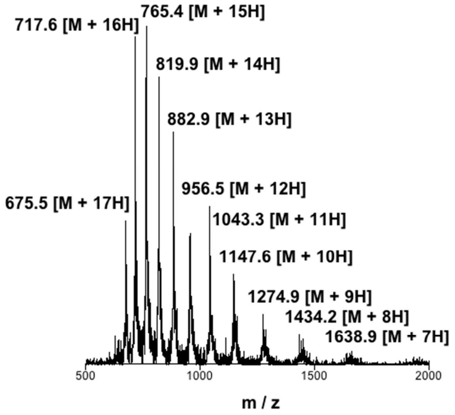


</td></tr>
</table>

9.3.4. *Desulfurization of F3a-F4a-F5c: Cys(Acm)^{79}-Cys^{122}-Cys^{142}-Cys^{160}(Acm)-Lys^{175}-PEG$_2$-Biotin:*

25 mg (2.2 µmol) of Cys(Acm)^{79}-Cys^{122}-Cys^{142}-Cys^{160}(Acm)-Lys^{175}-PEG-Biotin were dissolved in 1.4 mL of buffer B and the pH was adjusted to 6.5. The reaction was heated up to 60ºC for 45 minutes and monitored by UHPLC/ESI-MS. Upon completion, the solution was acidified with HCl$_{(aq)}$ 10% until pH 2, and purified by semi-preparative HPLC.

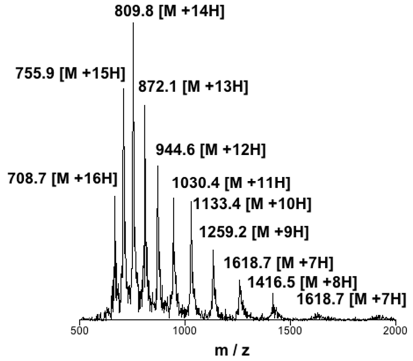
<table>
<tr><td colspan="2">Cys(Acm)KDKLNALAISVMNQWPGVKLRVTEGWDEDGHHSEESLHYEGRAVDITTSDRDRSKYGMLARLA VEAGFDWVYYESKAHIHC(Acm)SVKAENSVAAKSGG-PEG₂-Biotin</td></tr>
<tr><td>Mass: 11465.8</td><td>Yield: not determined</td></tr>
<tr><td colspan="2">




</td></tr>
</table>

9.3.5. Acm removal of F3a-F4a-F5c: Cys(Acm)^{79}-Ala^{122}-Ala^{142}-Cys^{160}(Acm)-Lys^{175}-PEG$_2$-Biotin

Acm was removed following the general procedure. 25 mg (0.9 µmol) were dissolved in 3 mL of a buffer containing 1:1 AcOH:H_2O, 50 mM silver triflate. The reaction was left overnight. After performing the work-up and further purification 9 mg (0.8 µmol, 70 % yield) of the final product were obtained.

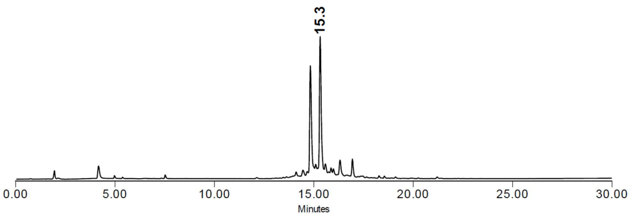
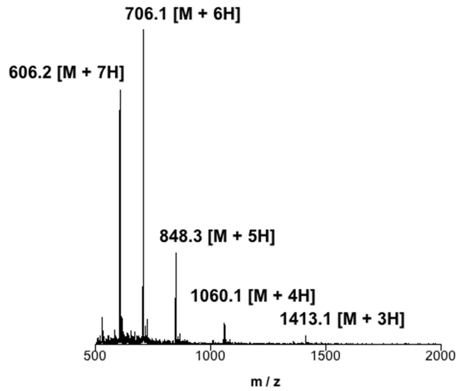
CKDKLNALAISVMNQWPGVKLRVTEGWDEDGHHSEESLHYEGRAVDITTSDRDRSKYGMLARLAVEAGF DWVYYESKAHIHCSVKAENSVAAKSGG-PEG$_2$-Biotin	
Mass: 11465.8	**Yield:** not determined
	

10. Synthesis of IleIle-ShhN-biotin

10.1. Peptide fragment synthesis

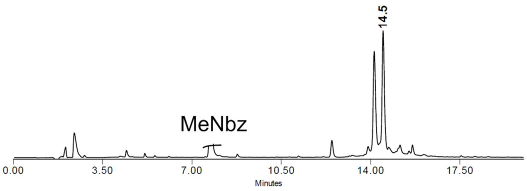
10.1.1. Synthesis of F1a: Ile^1-Gly^{35}-pCN-Phoc-MeDbz-Gly

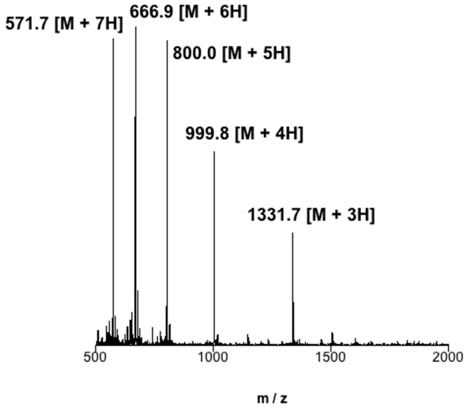
Starting from 1.75 g (0.82 mol) of Chemmatrix® resin and following the general procedure for the synthesis of *p*CN-Phoc-MeDbz peptides, 640 mg of crude were recovered after cleavage.

IICGPGRGFGKRRHPKKLTPLAYKQFIPNVAEKTLG-*p*CN-Phoc-MeDbz-Gly	
Mass: 4236.37	**Yield:** 18.3 %
	

10.1.2. Synthesis of F1a thioester: Ile^{1}-Gly^{35}-S(CH$_2$)$_2$SO$_3$Na

531 mg (0.13 mmol) of crude F1a were dissolved in 15 mL of a buffer containing 6 M Gn·HCl, 0.2 M phosphate, 20 mM TCEP and 100 mM MESNa. The pH was adjusted to 7.5 using a solution of 1 M NaOH$_{(aq)}$. The reaction was left at rT. After purification, 150 mg (28 % yield) of the product were obtained.

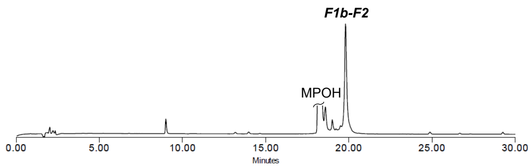
<table>
<tr><td colspan="2">IICGPGRGFGKRRHPKKLTPLAYKQFIPNVAEKTLG-S(CH$_2$)$_2$SO$_3$Na</td></tr>
<tr><td>Mass: 3995.22</td><td>Yield: 28 %</td></tr>
<tr><td colspan="2">

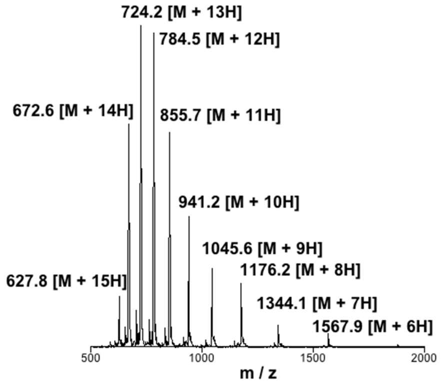

</td></tr>
</table>

10.2. Ligation of the fragments

10.2.1. KCL between F1a and F2: Ile^{1}-Gly^{35}-pCN-Phoc-MeDbz (or Ile^{1}-Gly^{35}-S(CH$_2$)$_2$SO$_3$Na) and Cys^{36}-Arg^{79}-pCN-PhCB-MeDbz-Gly

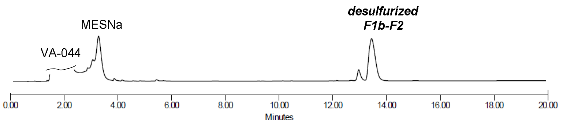
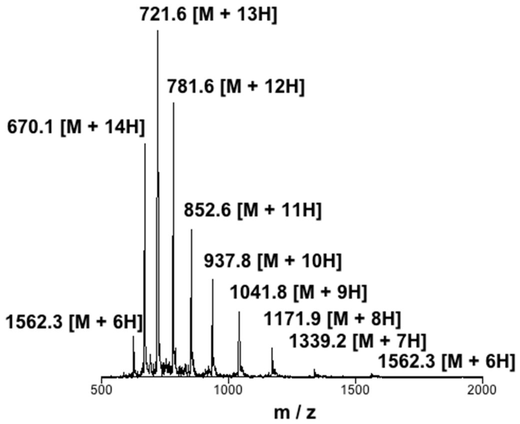
31 mg (7.3 µmol) of Ile^{1}-Gly^{35}-pCN-Phoc-MeDbz and 32 mg (5.7 µmol) of Cys^{35}-Arg^{78}-*p*CN-Phoc-MeDbz were dissolved in 3 mL of NCL buffer containing 100 mM of MPOH, and pH was adjusted to 6.85. The ligation was left at rT for 2h30 and the workup was carried out following the protocol described above. After purification 36 mg (3.8 µmol, 22% yield) of ligated Palm-Ala^{1}-Cys^{35}-Arg^{78}-*p*CN-PhCB-MeDbz-Gly were obtained.

<table>
<tr><td colspan="2">IIGPGRGFGKRRHPKKLTPLAYKQFIPNVAEKTLGCSGRYEGKISRNSERFKELTPNYNPDIIFKDEENT GADRLMTQR-<i>p</i>CN-Phoc-MeDbz-Gly</td></tr>
<tr><td>Mass: 9395.86</td><td>Yield: 52 %</td></tr>
<tr><td colspan="2">


</td></tr>
</table>

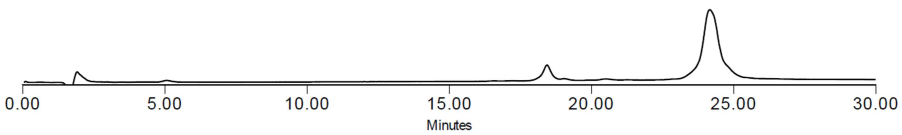
10.2.2. Desulfurization of F1a-F2: Ile^{1}-Gly^{35}- Cys^{36}-Arg^{79}-pCN-PhCB-MeDbz-Gly

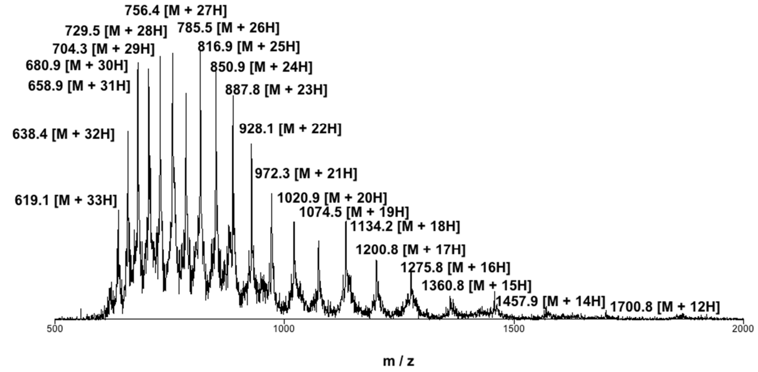
Desulfurization was carried out following the general desulfurization procedure 30 mg (3.2 µmol) of Ile^{1}-Gly^{35}- Cys^{36}-Arg^{77}-*p*CN-PhCB-MeDbz were dissolved in 2.0 mL of buffer B. The reaction was heated up to 37 °C for 2 h and monitored by UHPLC/ESI-MS. Upon completion, the solution was acidified with HCl$_{(aq)}$ 10% until pH 2, and purified by semi-preparative HPLC. After purification 13 mg (1.4 µmol, 43 % yield)) of Ile^{1}-Gly^{35}- Ala^{36}-Arg^{77}-*p*CN-PhCB-MeDbz were obtained.

<table>
<tr><td colspan="2">IIGPGRGFGKRRHPKKLTPLAYKQFIPNVAEKTLGASGRYEGKISRNSERFKELTPNYNPDIIFKDEENT GADRLMTQR-<i>p</i>CN-Phoc-MeDbz-Gly</td></tr>
<tr><td>Mass: 9363.89</td><td>Yield: 43 %</td></tr>
<tr><td>
</td><td>
</td></tr>
</table>

10.2.3. KCL between F1a-F2 and F3-F4a-F5c Ile^{1}-Arg^{79}-pCN-PhCB-MeDbz-Gly and Cys^{80}-Cys^{162}(Acm)-Lys^{176}-PEG$_{2}$-Biotin:

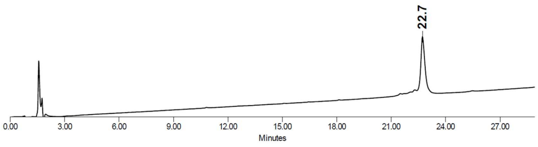
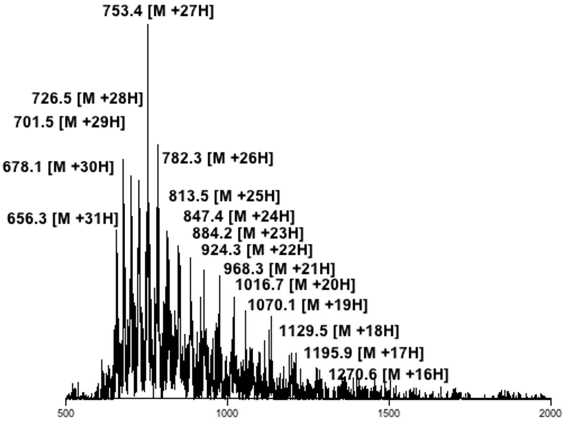
13 mg (3.0 µmol) of Ile^{1}-Arg^{79}-*p*CN-PhCB-MeDbz-Gly and 20 mg (3.6 µmol) of Cys^{80}-Cys^{161}(Acm)-Lys^{176}-PEG$_{2}$-Biotin were dissolved in 2 mL of NCL buffer, and pH adjusted to 6.9. The reaction was left at rt for 18h. After HPLC purification 10 mg (4.9 µmol, 13% yield) of Palm-Ala^{1}-Ala^{35}-Arg^{79}- Cys^{80}-Ala^{123}-Ala^{143}-Cys^{161}(Acm)-Lys^{176}-PEG$_{2}$-Biotin were obtained.

<table>
<tr><td colspan="2">IIGPGRGFGKRRHPKKLTPLAYKQFIPNVAEKTLGASGRYEGKISRNSERFKELTPNYNPDIIFKDEENT GADRLMTQRCKDKLNALAISVMNQWPGVKLRVTEGWDEDGHHSEESLHYEGRAVDITTSDRDRSKYGML ARLAVEAGFDWVYYESKAHIHC(Acm)SVKAENSVAAKSGG-PEG$_{2}$-Biotin</td></tr>
<tr><td>Mass: 20384.3</td><td>Yield: 13 % not determined</td></tr>
<tr><td colspan="2">
</td></tr>
</table>



10.2.4. Acm removal F1a-F2-F3-F4b-F5c-PEG$_2$-biotin: Ile^1 -Cys^{162}(Acm)-Lys^{176}-PEG$_2$-Biotin:ç

Acm was removed following the general procedure. 5.2 mg (0.23 µmol) of Ile^1-Cys^{161}(Acm)-Lys^{176}-PEG$_2$-Biotin were dissolved in 1 mL of a buffer containing 1:1 AcOH:H_2O, 50 mM silver triflate. The reaction was left overnight. After performing the work-up and further purification 3.0 mg (0.14 µmol, 62% yield) of the final product were obtained.

IIGPGRGFGKRRHPKKLTPLAYKQFIPNVAEKTLGASGRYEGKISRNSERFKELTPNYNPDIIFKDEENT GADRLMTQRCKDKLNALAISVMNQWPGVKLRVTEGWDEDGHHSEESLHYEGRAVDITTSDRDRSKYGML ARLAVEAGFDWVYYESKAHIHCSVKAENSVAAKSGG-PEG$_2$-Biotin	
Mass: 20313.3	**Yield:** 62 %
	

11. Folding of ShhN analogs

11.1. Electrophoresis

Electrophoretic separations were performed in one-dimensional microgels using sodium dodecyl sulfate-polyacrilamide gel electrophoresis (SDS-PAGE).

Preparation of the gel: The electrophoresis system used in the present thesis was Mini-Protean (BioRad). All gels were prepared with a depth of 1.0 mm. The glass-plate sandwich of the electrophoresis apparatus is prepared accordingly to manufacturer's instructions using two celan glass plates and a 1.0 mm spacer and is locked to the casting stand. The separating gel (7.5 mL)is prepared using 12.5 % of acrylamide, Tris-Base 0.375 M, pH 8.8, SDS 0.1 %, ammonium persulfate (APS) 0.05 % and TEMED 2.2 mM, and is added among the glass-plate sandwich. Next, isopropanol (100 µL) is added. The gel is allowed to polymerixe at rT. Meanwhile, the stacking gel is prepared using 3.3 % of acrylamide, Tris HCl 0.125 M, pH 6.8, SDS 0.1 %, APS 0.1 % and TEMED 6.6 mM. Then, isopropanol is eliminated and 2.5 mL of the stacking gel are added on the polymerized separating gel. A Teflon comb is inserted into the layer of the stacking gel to gform the wells. Plymerization is allowed at rT.

Preparation of the samples: 20 µL of the sample was mixed with 5 µL of loading sample buffer, and left at 95 °C for 5 min.

Gel electrophoresis: the polymerized gel sandwich is placed in the electrophoresis system and the buffer chamber is filled with electrophoresi buffer, shich contains Tris-Base 25 mM, glycine 0.19 M and SDS 0.1 %. The Teflon comb is removed, and the samples are loaded into the wells. Finally, electrophoresis buffer is added to completely cover the platinum electrode within the upper chamber, and the electrophoresis gel is run at a fix 120 V for 1 h.

Silver staining:

1. Incubate the gel with 100 ml (or the whole buffer prepared) Fixation solution for 15 min (leave the gel covered into a tapper at the shaker)
2. 3 washings of 3 min in 30 ml of 30 % EtOH in H2O (500 mL of washing buffer is prepared in advance)
3. Incubate the gel in 100 ml Pretreatment solution for 1 min (shaker)
4. 3 washings of 20 s in milliQ H2O
5. Incubate the gel in 100 ml Impregnation solution for 20 min (shaker)
6. 2 washings of 20 s in milliQ H2O
7. Incubate 100 ml Revealing solution – control how the bands appear! (at this point we shake the gel and control by the eye how the silver precipitates and reveals. We can do it as long as we want to see the bands appear)
8. 2 washings of 2 min in milliQ H2O

9. Incubate 100 ml Stop solution for 10 min

10. Preserve the gel in milliQ H2O and covered.

11.2. Dialysis

200 µL of the protein dissolved in a denaturant buffer (buffer 1) were placed in a cassette with a membrane cutt-off of 10 kDa. The display was placed in a recipient with a larger volume (25-fold the colume of buffer 1) of the dialysis buffer (buffer 2), and incubated for the desired time. When equilibrium was reached, the dialysis buffer was changed. After the dialysis experiment the solution containing the protein was further analyzed by circular dichroism.

11.3. FAST dilution

The protein was dissolve at 25 µM in a denaturant buffer (6 M Gn·HCl, 0.2 M phosphate buffer). Next, it was diluted 20-30-fold in the corresponding refolding buffer and incubated for a concrete period of time (from 30 minutes to 1 hour). Afterwards, the solution containing the folded protein was centrifuged in order to precipitate any aggregate. Finally, the supernatant was concentrated until the desired protein concentration.

Best conditions to properly fold ShhN were using the folding buffer consisting in in 25 mM phosphate, 150 mM NaCl, 2 mM DTT, 0.5% DDM, 0.5 µM $ZnCl_2$ buffer.

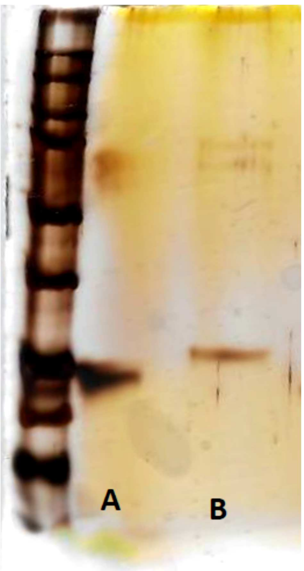


Figure 1. SDS-PAGE after silver staining showing the protein supernatant after folding (well A), and the protein precipitate (well B) after folding.

11.4. ELISA

Folding of synthetic Palm-ShhN-biotin and IleIle-ShhN-biotin was assessed by ELISA using the monoclonal antibody 5E1 [1].

1. Nunc ®MaxiSorp™ 384 well plates were coated with 30 µL of the target protein from a 50 nM solution in PBS. Control wells were coated with BSA at 2 µg/mL in PBS. The plates were left overnight at 4 °C.
2. Blocking was performed by adding 70 µL of blocking buffer into each well, and incubated for 1 hour 30 minutes at rT.
3. ELISA plates were washed 4 times with PT buffer.
4. Then the plates were incubated with 30 µL from a stock solution of 5E1 at 50 nM in PBT, for 1 h at rT.
5. Plates washed x6 with PT.
6. 30 µL of secondary antobody (host Goat, antiMouse HRP diluted 6000 fold in blocking buffer) were added and incubated with the plates for 45 min at rT.
7. Plates are washed x6 with PT buffer.
8. 30 µL of fresh TMB substrate were added into all wells, and incubated until the color developed.
9. 30 uL of 1M H_3PO_4 were added into each well to stop the reaction.
10. The plates were read in the plate reader at 450nm.

12. Phage display

12.1. Library preparation for phage selection rounds

1. 1 to 1.5 mL from the phage library solution stock were transferred to a 50mL falcon tube. Then 30 mL of PBS and 10 mL of PEG/NaCl buffer were added, and the tubes chilled in ice bath for 20 minutes.
2. The tubes were centrifuged at 8500 rpm for 20 minutes in order to precipitate the phages in a small pellet. The supernatant was discarded.
3. The remaining supernatant was carefully removed with 1mL pipette.
4. The phage pellet was resuspended in 1 mL PBT buffer, and the solution was transferred into a clean 1.5 mL microcentrifuge tube.
5. The phage-containing microcentrifuge tube was centrifuged at 1500 rpm, 4 °C for 10 minutes. The supernatant was transferred into clean 1.5 mL microcentrifuge tube. This supernatant was named **input**.

12.2. Protein reconstitution or refolding

- **Commercial ShhN** from Peptrotech was reconsituted in ddH_2O water at 1 mg/mL concentration. The protein solution was kept at 4 °C.
- **Recombinant Sonic Hedgehog (rShhN)** was reconstituted in ddH_2O to a concentration of 10 µM, kept at 4 °C.
- **N-biotinylated Shh** was reconstituted in ddH_2O to a concentration of 1 mg/mL. Protein solution was kept at 4 °C.
- Synthetic analogues **IleIle-Shh-Biotin** and **Palm-Shh-Biotin** were refolded in refolding buffer containing 150 mM NaCl, 0.5% DDM, 25 mM NaPi, 1 µM $ZnCl_2$, 25 mM DTT, pH 8.0, at 5 µM concentration, protein solutions were kept at 4 °C.

12.3. Biopanning selections

Coating of the target onto the microtiter plate could be direct or indirect, depending on the targets bearing or not a biotin residue. Direct coating was the election for those targets lacking biotin (commercial ShhN and recombinant ShhN), whereas pre-coating with neutravidin was chosen when a biotin residue was attached at the target (synthetic analog Palm-ShhN-Biotin, IleIle-Shh-Biotin and N-terminal Biotinylated Shh).

1. Coating of 96-well Maxisorp immunoplates:
    a. **Direct coating:** each well was coated with 100 µL of target protein from a 5 µg/mL stock solution in PBS. Control wells were coated with BSA at 5 µg/mL in PBS. The plates were left overnight at 4 °C.

       The plates were blocked adding 200 µL of blocking buffer into each well, and incubated for 1 hour 30 minutes at room temperature.
    b. **Pre-coating with neutravidin:** each well was coated with 100 µL of neutravidin from a 5 µg/mL stock solution in PBS. Plates were left overnight at 4 °C.

       Then, the plates were blocked adding 200 µL of blocking buffer into each well. Incubated for 1h 30 minutes at room temperature, and washed 4 times with PT buffer. The target protein was coated by adding 100 µL from a 300 nM stock solution in PBT and incubated for at least 30 minutes. Control wells were incubated with 100 µL PBT.
2. The plates were washed 4 times with PBT.
3. The wells were incubated with 100 µL of phage input (from step 5.1.5) for 1h at room temperature.

4. The solution was discarded and the plates washed 10 times with PT buffer to remove the unbound phages.
5. 100 µL of 0.1 M HCl were added into each well, and incubated for 5 minutes to elute the binding phages.
6. The eluted phages from the wells coated with the target were joined in a microcentrifuge tube (called **output**). The eluents from the control wells were transferred into separate microcentrifuge tubes (named **control**).
7. The pH of the eluents was adjusted by adding 1/8 volumes of 1.0 M Tris solution at pH 11.
8. **Phage amplification:**
    a. Half of the volume from the output and control solutions (500 µL) was added to 10 volumes (5 mL) of growing *E. Coli Omnimax* (OD_{550} 0.8) in 2YT/tet media in 15 mL falcon tubes. They were incubated for 20 minutes in the shaker at 37 °C and 200 rpm.
    b. 5 µL of M13K07 helper phage solution (10^{10} phage/ µL) were added into the tubes from pervious step and incubated at the shaker, 37 °C, 200 rpm for 1h.
    c. 5 mL from each tube were transferred into 250 mL baffled flasks containing 30 mL of 2YT/carb/kan and incubated in the shaker at 200 rpm, 37 °C overnight.
    d. The suspension containing the amplified phages and bacteria was transferred into clean 50 mL falcon tubes and centrifuged at 8500 rpm, 4 °C for 10 minutes to precipitate the cells.
    e. The supernatant was transferred into clean 50 mL falcon tubes, then 10 mL of PEG/NaCl buffer were added and chilled in ice bath for 20 minutes.
9. Repeat steps 5.1.2 – 5.3.5 for phage input preparation.
10. Repeat steps 5.3.1 – 5.3.8 for biopanning selection.
11. *Determination of the enrichment ratio:*
    a. To determine the enrichment of the libraries after each round of selections, both the **input** (from step 5.1.5), **output** (from step 5.3.6) and **control** (from step 5.3.7) had to be plated. To do so, six serial dilutions from the output and control (10 µL in 90 µL of PBS) and nine dilutions from the input of each library were carried out.
    b. Then, 10 µL from each serial dilution were transferred to a plate which already contained 90 µL of **growing media** (*Omnimax E.Coli* in 2YT/tet, OD_{550}=0.8) in each well.
    c. Next, 5 µL from each serial dilution from previous step were plated on LB/carb to determine the number of phage eluted. Additional 5 µL were plated on LB/tet

plates to detect any contamination. And additional 5 µL were plated on LB/kan plates (to determine cells infected with helper phage). The plates were left overnight at the incubator at 37 °C.

  d. An increase of colonies should be observed in LB/carb plates after each round of selection.

12. The rounds of selections were repeated until desired enrichment was observed. In our case, we carried out 4 rounds of biopanning.

12.4. Selection of colonies

1. The dilution from the titration plates from step 5.3.11.c, showing 5 to 15 grown colonies was selected. We selected one dilution per library.
2. To seed the selected dilutions, 10 µL from the corresponding dilutions in PBS (from step 5.3.11.a) were transferred into 90 µL of fresh growing media.
3. Let in the incubator at 37 °C overnight.
4. We used one blue box per library screened, and filled each well with 450 µL of 2YT/M13K07/Carb media.
5. Using sterile toothpicks 96 colonies from each dilution previously plated were picked and let stand in the blue box well for 5 minutes.
6. The toothpicks were removed and the blue boxes left in the shaker at 37 °C, 200 rpm overnight.
7. The blue boxes were kept at 4 °C until use.

12.5. ELISA

3. Coating of the antigen on Nunc® MaxiSorp™ 384 well plates:
    a. **Direct coating:** each well was coated with 30 µL of the target protein from a 2 µg/mL stock solution in PBS. Control wells were coated with BSA at 2 µg/mL in PBS. The plates were left overnight at 4 °C.

       Blocking was performed by adding 70 µL of blocking buffer into each well, and incubated for 1 hour 30 minutes.

    b. **Pre-coating with neutravidin:** each well was coated with 30 µL of neutravidin from a 5 µg/mL stock solution in PBS. The plates were left overnight at 4 °C.

       Blocking was performed by adding 70 µL of blocking buffer into each well, and incubated for 1 hour 30 minutes. The plates were washed 4 times with PT buffer and coated with the target protein by adding 30 µL from a 100 nM stock solution in PBT and incubated for at least 30 minutes. Control wells were incubated with 30 µL of PBT buffer.

4. ELISA plates were washed 4 times with PT buffer.
5. The phages in the blue boxes from step 5.4.7 were centrifuged at 3200 rpm, 4 °C for 10 minutes.
6. Phage supernatants from the previous step, were diluted 4-fold with PBT buffer (100 µL of phage supernatant were transferred into 300 µL of PBT buffer) and transferred into new blue boxes.
7. The coated wells were incubated with 30 µL of diluted phages, from step 5.5.3, for 1h at room temperature.
8. The solution was removed and the plates washed 8 times with PT buffer.
9. 70 µL of HRP/anti-M13 antibody conjugate (diluted 5000-fold in PBT) were added into each well, and incubated for 30 minutes at room temperature.
10. The solution was removed and the plates were washed 6 times with PT buffer and 2 times with PBS buffer.
11. 50 µL of fresh TMB substrate were added into each well, and the color was allowed to develop for 10 minutes.
12. The reaction was stopped upon the addition of 50 µL of H_3PO_4 and the plates were read at 450 nm in a plate reader.
13. The ELISA data obtained was processed as following: the absorbance signal obtained from the wells coated with the target were divided between the absorbance value from their control (BSA for the direct coating; neutravidin for the protein requiring pre-coating). Ratio values above 5 were taken as positive binding. Identified positive binders were PCR amplified and sequenced.

12.6. PCR

1. 0.5 µL from phage supernatant was added on 25 µL of the following PCR mix to carry out PCR from the selected binders:
2.

FOR: 5’ ATGTTTTTTCATTTGGGT 3’
REV: 5’ AAAGGAGCCTTTAATTGT 3’

Component	Amount	Final
Phage supernatant	0.5 µL	
Ampli*Taq* PCR buffer 10x	2.5 µL	1x
Phagemid-sequencing Forward primer	0.750 µL	300 nM
Phagemid-sequencing Reverse primer	0.750 µL	300 nM
Deoxynucleotide mix 10 mM of each dNTP	0.625 µL	250 µM of each dNTP
Ampli*Taq* DNA polymerase	0.1 µL	0.5 U per reaction
ddH_2O	Up to 25 µL	

U = units

3. The DNA was amplified using the following PCR program:

Component	Temperature (ºC)	Time	Cycles
Initial denaturation	94	5 min	1x
Denaturation	94	30 sec	25x
Annealing	60	30 sec	
Elongation	72	1 min	
Final elongation	72	7 min	1x
Cooling	4	Unlimited time	

4. The PCR reactions were analyzed by electrophoresis on 2% TAE/agarose gel run for 30 minutes at 100 V.
5. If the PCR ran well, we proceed to the clean up reaction. For that, 2 µL of PCR clean up mix were added to 5 µL of the PCR product.
6. The PCR clean up reaction was incubated at 37 °C for 15 min and then at 80 °C for 15 min.
7. The samples were sequenced.

12.7. Competitive ELISA

1. We first carried out competitive ELISA pre-coating the plates with netravidin. Unfortunately, this approach showed incongruent results, as explained in the discussion of this chapter. Because of that, when carrying out the competitive ELISA, the coating of the target was direct.

2. For direct coating of the antigen on Nunc® MaxiSorp™ 384 well plates, each well was coated with 30 µL of target protein from a 100 nM stock solution in PBS. Control well was incubated with BSA at 2 µg/mL in PBS. The plates were left overnight at 4 °C.
3. Blocking was performed by adding 70 µL of blocking buffer into each well, and incubated for 1 hour 30 minutes at room temperature.
4. Blue boxes containing selected phages supernatant (from step 5.5.3) were centrifuged at 3200 rpm at 4 °C for 10 min. Then the phages supernatant were diluted 3-fold with PBT buffer and 100 µL of these solutions were transferred into clean blue boxes.
5. At the same time, dilutions of the antigen at 2 µM, 1 µM, 0.5 µM, 0.25 µM, 0.125 µM and 0.0625 µM in PBT, were prepared.
6. To achieve the desired final concentration of phages and antigen in the solution mix, 100 µL from the diluted phages (from step 5.7.4) and 100 µL from the antigen dilutions (from step 5.7.5) were mixed to finally have dilutions with the phages diluted 6-fold and the antigen at different concentrations: 1 µM, 0.5 µM, 0.25 µM, 0.125 µM, 0.0625 µM and 0.03125 µM.
7. The solution mixture containing the diluted phages and the antigen at a specific concentration were incubated for 1 h at room temperature previously to addition on the antigen-coated plate.
8. The antigen coated wells were incubated with 100 µL of the mix from previous step for 1h at room temperature.
9. Plates were washed with PT 8 times.
10. 70 µL of HRP/anti-M13 antibody conjugate (diluted 5000-fold in PBT) were added into each well, and incubated at room temperature for 30 minutes.
11. The solution was removed and the plates washed 6 times with PT buffer and 2 times with PBS buffer.
12. 50 µL of fresh TMB substrate were added into each well, and the color was allowed to develop for 10 minutes.
13. The reaction was stopped upon the addition of 50 µL of H_3PO_4 and the plates were read at 450 nm in a plate reader.

    To analyze the results the absorbance observed in the well incubated with a solution containing the 6-fold diluted phage but no antigen, was considered the maxim signal obtained. Thus, the half of this value would be the IC_{50} of the phage for the antigen. Absorbance values (measured on the plate reader) coming from the wells incubated with serial dilutions of the antigen and fixed concentration of phage were divided into the control well absorbance signal (well coated with the target but incubated with PBT).

	PBT control	1µM Ag	0.5 µM Ag	0.25 µM Ag	0.125 µM Ag	0.0625 µM Ag	0.03125 µM Ag	no Ag
P32NA_H3	0,124	0,097	0,105	0,162	0,2	0,203	0,297	0,58
P32NA_F11	0,029	0,099	0,186	0,281	0,485	0,926	0,987	3,5
P32NA_F2	0,021	0,043	0,055	0,084	0,139	0,176	0,739	3,5
P32_H4	0,071	0,066	0,089	0,116	0,151	0,359	0,483	1,189
P32_G11	0,022	0,043	0,076	0,117	0,189	0,518	0,865	3,5
P32_A10	0,035	0,047	0,067	0,086	0,118	0,221	0,518	2,023
P32_D7	0,032	0,057	0,141	0,164	0,281	0,604	0,961	3,5
P32_C6	0,471	0,487	0,875	0,699	2,994	4,788	1,109*	0,312*

Table 1. Competitive ELISA data expressed in Abs at 450 nm**.**

12.8. 12.8. Affinity maturation. Switch from pVIII to pIII phagemid vector

12.8.1. pIII Phagemid vector miniprep

1. 2 µL of pIII phagemid vector DNA were mixed with 10 µL KCM
2. The mix was chilled in ice for 2-5 min, and then 10 µL of *E. Coli* CJ236 cells were added.
3. The mix was incubated in ice for 20 min.
4. Then it was placed at room temperature for 10 min.
5. 200 µL of 2YT were added and the mixture was incubated at 37 °C in the shaker for 1h.
6. 100 µL of the mixture were plated on carb plates, and left in the incubator at 37 °C overnight.
7. One single colony from the plate was introduced into 5 mL of 2YT/carb, and left in the shaker at 37 °C, 200 rpm overnight.
8. The media containing the cells and the phagemid was transferred into a 50 mL falcon tube and centrifuged at 8000 rpm for 3 min. The supernatant was discarded.
9. DNA extraction from the pellet was carried out using the Qiagen DNA purification Miniprep Kit (**Figure *1***).
10. 1% TAE/agarose gel was run at 100V for 45 min to check the miniprep reaction.
11. The final DNA concentration was determined using the nanodrop (plate reader).

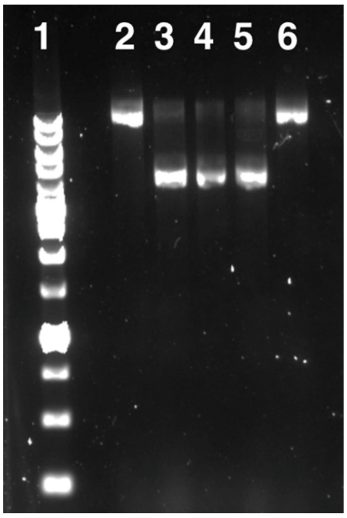


Figure 1. Analysis of 5 minipreps on a 1% TAE/agarose gel. Well 1) 1 kb DNA ladder; Well 2,3,4,5 and 6 corresponding to different minipreps of pIII phagemid vector. Minipreps with the desired size, 5000 pb, were in wells 2,3,4. Samples from wells 1 and 5 were discarded.

12.8.2. Plasmid digestion from selected clones from pVIII vectors.

1. PCR amplification of the desired inserts from selected clones in pVIII vectors was performed using the following primers designed and PCR mix:

FORWARD primer 5'-GGA TAT GCA TCC GAT TAT AAA GATG-3'
REVERSE primer 5'-TCC AGA GCT CCC GGC GTG AGC TTTC-5'

Component	Amount	Final
Phage supernatant	1 µL	
Ampli*Taq* PCR buffer 10x	10 µL	1x
Phagemid-sequencing Forward primer	3 µL	300 nM
Phagemid-sequencing Reverse primer	3 µL	300 nM
Deoxynucleotide mix 10 mM of each dNTP	2.5 µL	250 µL of each dNTP
Ampli*Taq* DNA polymerase	0.4 µL	0.5 U per reaction
ddH$_2$O	Up to 100 µL	

U = units.

2. PCR program used was the following:

	Temperature (ºC)	Time	Cycles
Initial denaturation	94	5 min	1x
Denaturation	94	30 sec	28x
Annealing	50	30 sec	
Elongation	72	1 min	
Final elongation	72	7 min	1x
Cooling	4	Unlimited time	

3. PCR products were cleaned up using PCR purification Kit from Qiagen. Final elution volume was 50 µL. A 2% TAE/agarose gel at 200 V for 8 minutes was run to assess the PCR amplification (***Figure 2***). Clone P32_A10 was not PCR amplified properly.

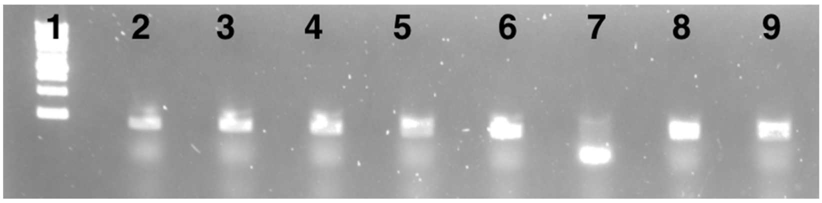


Figure 2. PCR products from the amplification of the unique clone genes. Well 1) 1 kb DNA ladder; clone P32_NA_H3; well 2) clone P32_NA_F11; well 3) clone P32_F2; well 4) clone P32_H4; well 5) clone P32_G11; well 6) clone P32_D7; well 7) clone P32_A10; well 8) clone P32_C6. All PCR inserts amplified, except well 7, showed the expected size below 250 pb.

4. The concentration of the final purified PCR products were measured on the nanodrop (plate reader).
5. Digestion of the amplified insert was carried out by adding 5 µg of the PCR amplified insert to the following mix:

Component	Amount	Final
Insert	variable	5 µg
NsiI HF	1 µL	100 U
SacI HF	1 µL	100 U
Cutsmart buffer	10 µL	1 x
dH_2O	Up to 50 µL	

6. The digestion mixture was incubated for 2h at 37 °C.
7. The pIII phagemid vector was also digested following the same procedure.
8. Gel purification was carried out to obtain pure pIII digested vector. For that 1% TAE/agarose gel was run at 100V for 45 min. The corresponding band was cut with a sterile razor and gel extraction DNA was performed using Qiagen Gel Extraction Kit. We did not take a picture of the gel since the UV light could damage the DNA.

12.8.3. Ligation

1. The following ligation reaction was set up to fuse the selected amplified inserts from pVIII vector into pIII phagemid vector.

Component	Amount	Final
pIII phagemid vector	Variable	10 ng
DNA insert	Variable	10 ng
T4 DNA ligase buffer x10	1 µL	1x
10 mM ATP	1 µL	
T4 DNA ligase	0.5 µL	
ddH$_2$O	Up to 10 µL	

2. The ligation reaction was incubated for 2 h at room temperature.
3. Next, 10 µL of the ligated products were transformed into *E. Coli* CJ236. The mix containing the ligated products was transferred into a microcentrifuge tube containing 30 µL KCM x 1 and chilled in ice for 2 min.
4. Same volume of cells was added and chilled in ice for 20 min.
5. The mix was placed at room temperature for 10 min.
6. 200 µL of 2YT/carb were added and left in the shaker at 37 °C, 200 rpm for 1h.
7. 100 µL of the culture were plated in LB/carb plates and left in the incubator at 37 °C overnight.

12.9. Cross-reactivity ELISA of Palm-Shh-biotin binders against recombinant ShhN

1. With the clones identified to bind to Palm-Shh-biotin antigen we carried out corss-reactivity ELISA to ensure that they recognize the recombinant ShhN. We coat two wells per phage to be screened.
2. ELISA assays were carried out on a Nunc®Maxisorp^{TM} 384 well plate. Each well was directly coated with 30 µL of recombinant ShhN from a 2 µM stock solution in PBS.
3. The plate was left at 4 °C overnight.
4. Blocking was carried out by adding 70 µL of blocking buffer into each well, and the plates were incubated for 1 h 30 min at room temperature.
5. The plate was washed 4 times with PT buffer.
6. One well was incubated with the diluted phage supernatant from the selected clones (supernatant from blue box from step 5.5.3), and the control well was incubated with PBT.
7. The plates were incubated for 1h at room temperature.
8. The solution was removed and the plates washed 8 times with PT buffer.
9. 70 µL of HRP/anti-M13 antibody conjugate (diluted 5000-fold in PBT) were added into each well, and incubated at room temperature for 30 minutes.

10. The solution was removed and the plates washed 6 times with PT buffer and 2 times with PBS buffer.
11. 50 µL of fresh TMB substrate were added into each well, and the color was allowed to develop for 10 minutes.
12. The reaction was stopped upon the addition of 50 µL of H_3PO_4 and the plates were read at 450 nm in a plate reader.
13. The results were processed as following: absorbance value from the well incubated with the phage was divided into the absorbance value from the control well. Ratio values above 5 were considered positive (**Table *2***).

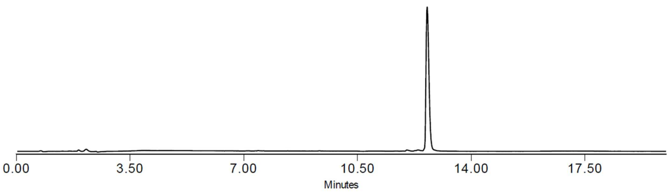
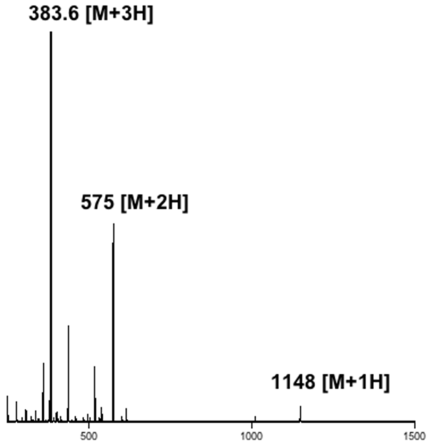
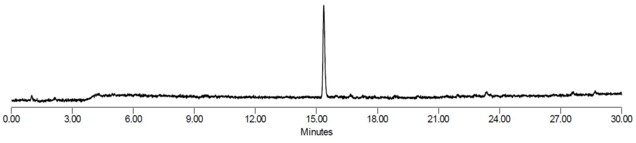
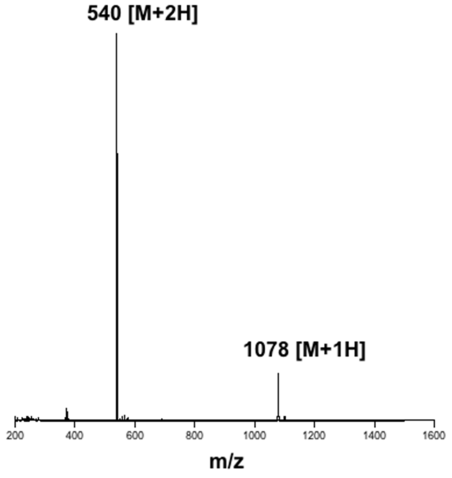
Clone	**a**	**b**	**c**	**d**	**e**	**f**	**g**	**h**
Ratio	6,603	14,596	8,256	4,76 *	7,087	9,716	5,1782	8,216

Table 2. Cross reactivity ELISA results. Ratio = Absorbance from phage incubated well / Absrobance from control well. (*) value close to 5, we considered it positive in cross-reactivity ELISA. Clone a = P32NA_H3; clone b = P32NA_F11; clone c = P32NA_F2; clone d = P32_H4; clone e = P32_G11; clone f = P32_A10; clone g = P32_D7; clone h = P32_C6.

Product characterization

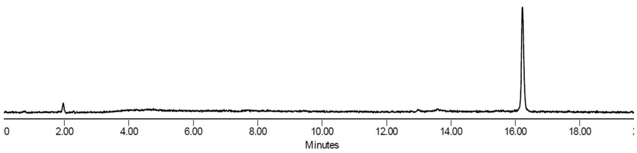
***p*-cyanophenyl chloroformate:** ^{1}H-RMN (400 MHz, CDCl$_3$): 7.8 (d, *J= 9.0,* 2H), 7.38 (d, *J= 9.0*, 2H). ^{13}C-NMR (100.6 MHz, CDCl$_3$): 111.56 (q), 116.47 (CN), 121.85 (CH$_2$ x2), 134.26 (CH$_2$ x2), 154.16 (q), 149.00 (q).

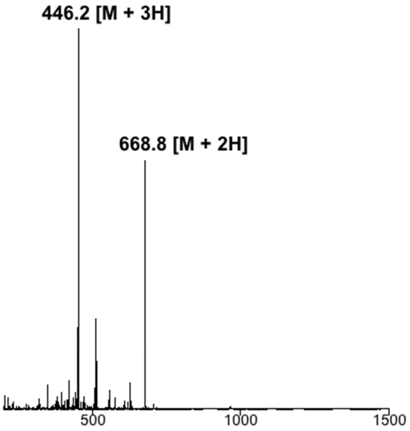
***p*-(methylsulfonyl)phenyl chloroformate:** ^{1}H-RMN (400 MHz, CDCl$_3$): 7.90 (d, *J= 8.0,* 2H), 7.34 (d, *J= 7.0*, 2H).

Peptide 1 LYSAA*p*NO$_2$-Phoc-MeDbz	**Exact mass:** 1149.54
	
Peptide 2 VSYRAL*p*NO$_2$-Phoc-MeDbz	**Exact mass:** 1077.50
	

Peptide 3 VSKSYRAR*m*NO_2-Phoc-MeDbz

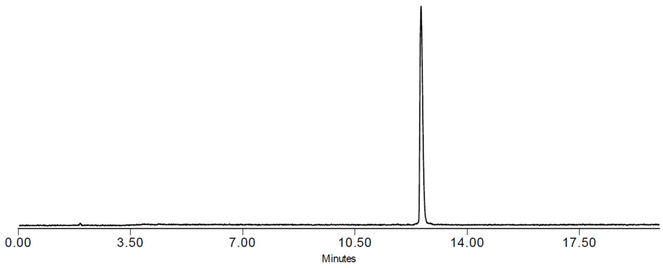
Exact mass: 1351.51

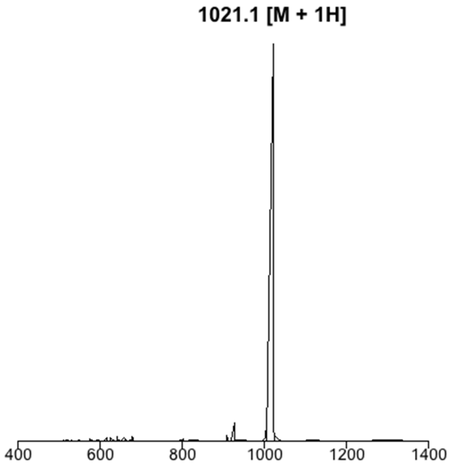




Peptide 4 LSYRAS-Phoc-MeDbz

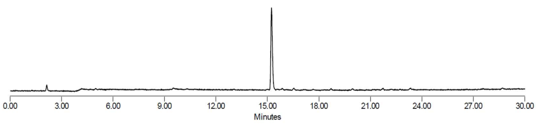
Exact mass: 1020.48

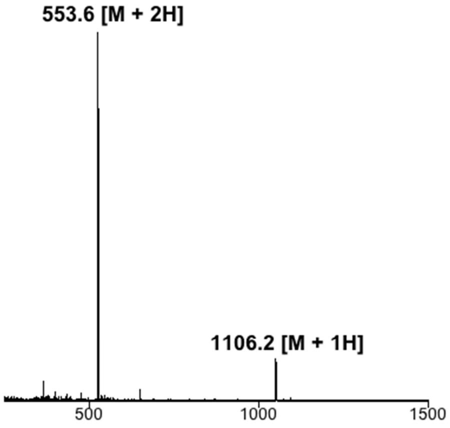


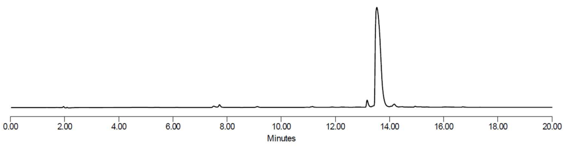


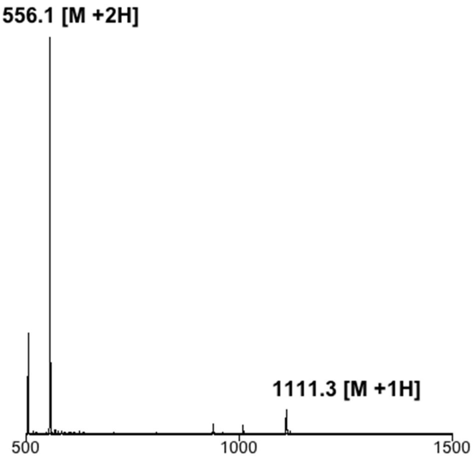
Peptide 5 VSYRAL*p*F-Phoc-MeDbz

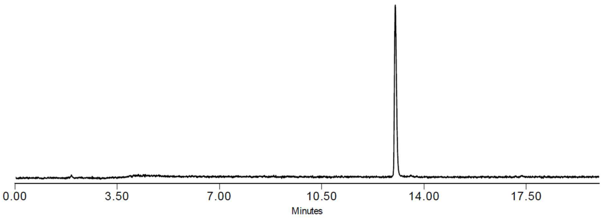
Exact mass: 1050.51

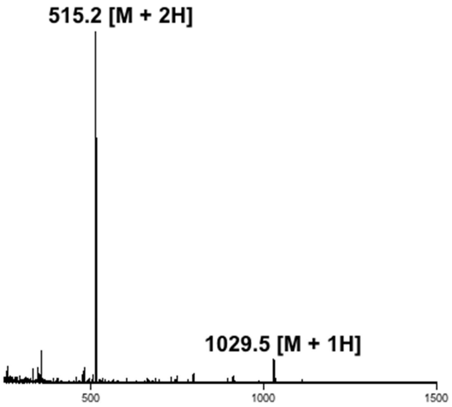


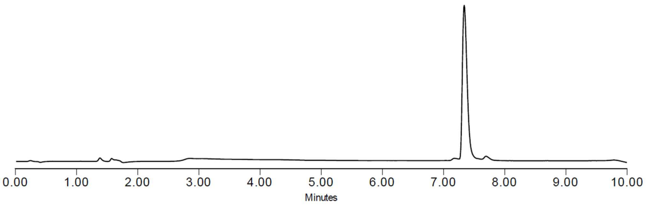


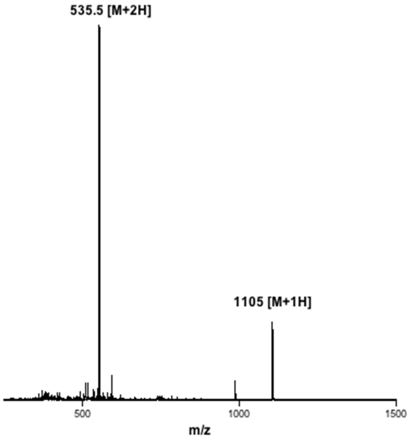
<table>
<tr><td><i>Peptide 6</i> VSYRAL<i>p</i>SO_2Me-Phoc-MeDbz</td><td>Exact mass: 1110.50</td></tr>
<tr><td colspan="2">




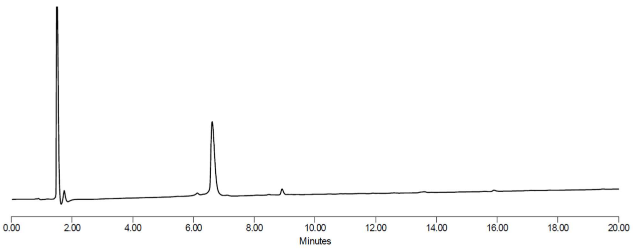
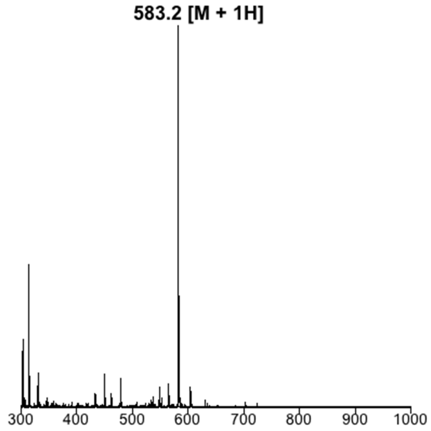
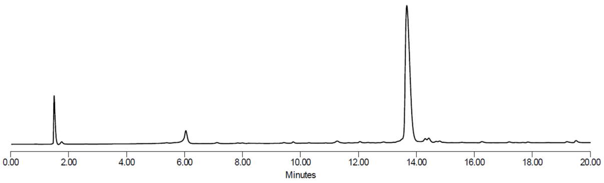
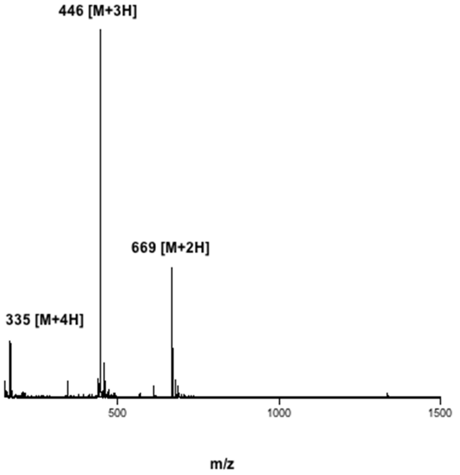
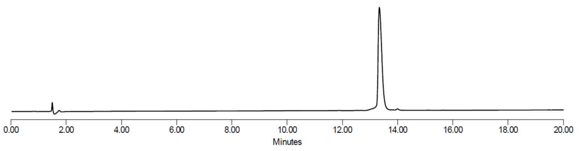
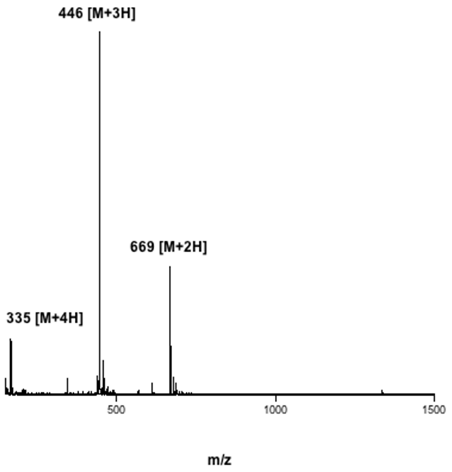
</td></tr>
<tr><td><i>Peptide 7</i> LSYRAA<i>p</i>CN-Phoc-MeDbz</td><td>Exact mass: 1030.13</td></tr>
<tr><td colspan="2">


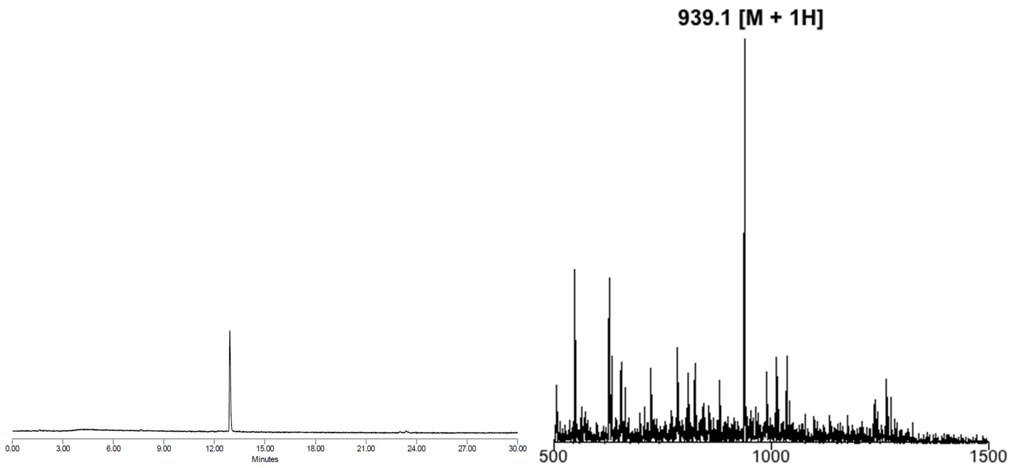


</td></tr>
<tr><td><i>Peptide 8</i> LSYAF<i>p</i>CN-Phoc-MeDbz</td><td>Exact mass: 964.42</td></tr>
<tr><td colspan="2">


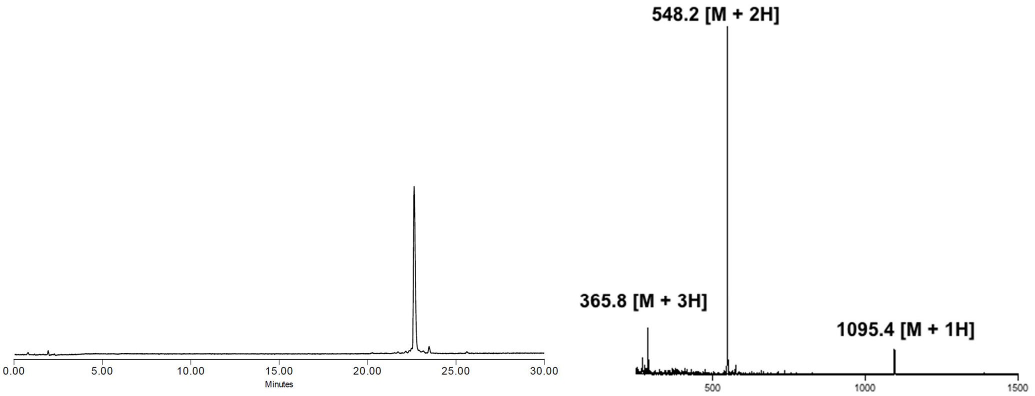


</td></tr>
</table>

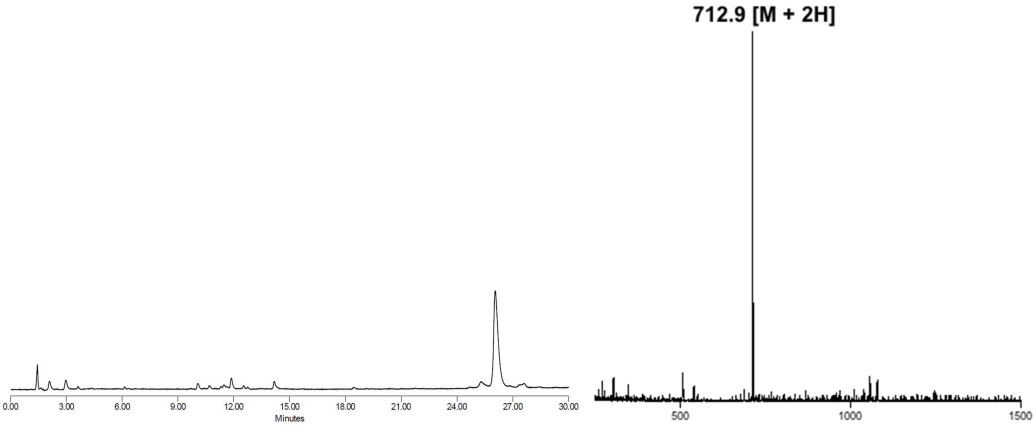
Peptide 9 CRAFS	**Exact mass:** 583.68
	
Peptide 11 LSYRAy-*p*CN-Phoc-MeDbz	**Exact mass:** 964.42
	
Peptide 12 LSYRAY-*p*CN-Phoc-MeDbz	**Exact mass:** 964.42
	
Peptide 15 VSYRAL-MeNbz	**Exact mass:** 937.5



Peptide 16 CSYRAF-*p*CN-Phoc-MeDbz | **Exact mass:** 1094.4

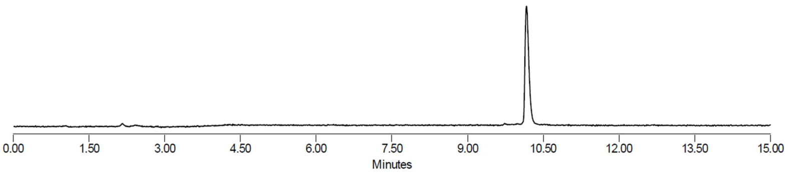


Peptide 19 Palm-C(Acm)PRAYR-*p*CN-Phoc-MeDbz | **Exact mass:** 1423.8



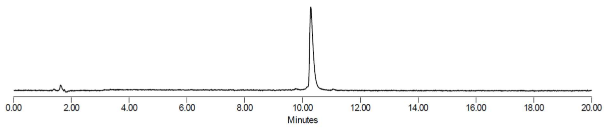
Peptide 21 VSYRAHCRAFS

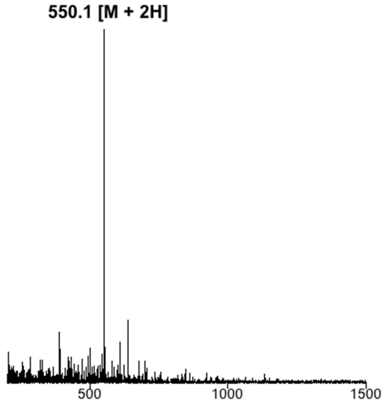
Exact mass: 1295.6



Peptide 23 VSYRAR-*p*CN-Phoc-MeDbz

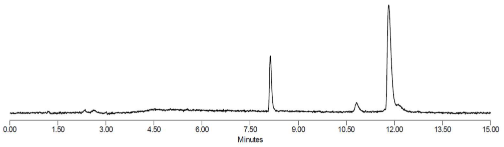
Exact mass: 1099.5

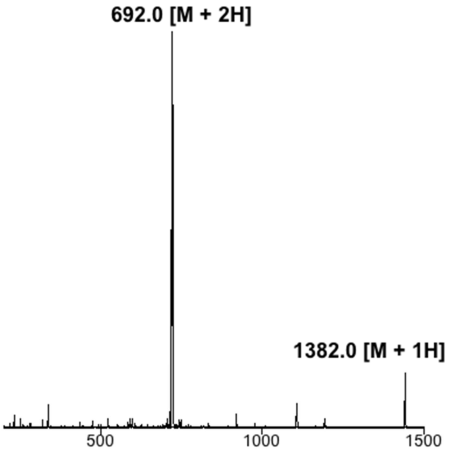




Peptide 24 ThzRNGIPCES-MeNbz

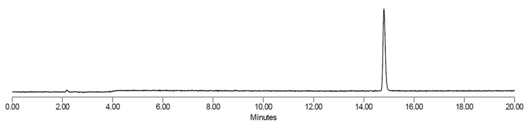
Exact mass: 1382.6

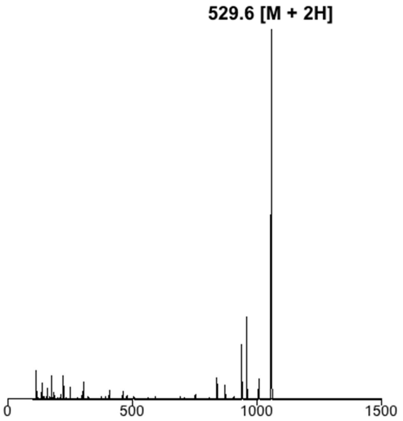




Peptide 26 VSYRAL-*p*CN-Phoc-MeDbz

Exact mass: 1056.5





ANNEXES

Recombinant expression of ShhN.

The nucleotide sequence from SinoBio ORF clone HG10372-M plasmid with a sequence identical to GenBank ref NM_000193.2 (full length human SHH clone) was used as template in in standard 50µl PCR reactions using KOD polymerase (as per manufacturer's instructions). 100ng of template was used per 50µl reaction. All primers were ordered from/synthesized by Sigma-Aldrich. PCR products were purified using Ampure XP beads (Beckman), eluted in EB and In-Fused (using standard In-Fusion protocol as per manufacturer's instructions) into KpnI-HindIII linearized pOPINS (Acta Crystallogr Sect F Struct Biol Cryst Commun. 2008 Apr 1; 64(Pt 4): 258–262.). After dilution with TE 5µl aliquots of the In-Fusion reactions were transformed into chemically competent Omnimax2 cells and plated on LB Agar supplemented with Kanamycin, IPTG and X-gal. Two colonies for each construct were picked for culturing and the production of plasmid minipreps. All plasmid inserts were fully sequence verified before use.

Expression plasmids for constructs 795-797 were transformed into Rosetta (DE3) pLysS E.coli, and the cells were plated on LB agar plates supplemented with Kanamycin (50µg/ml), Chloramphenicol (35µg/ml) and Glucose (1%w/v). Colonies were picked into LB supplemented with Kanamycin (50µg/ml), Chloramphenicol (35µg/ml) and Glucose (1%w/v) and grown with shaking at 37°C for 18 hrs. Overnight cultures were then diluted 1:100 in Overnight Express TB auto-induction media (Merck) supplemented with Kanamycin (50µg/ml) and Chloramphenicol (35µg/ml) to a final volume of 1l, grown with shaking at 37°C for 3hrs, grown with shaking at 25°C for a further 24hrs before harvesting by centrifugation at 5000g for 10 minutes at 4°C. Pellets not used immediately were stored at -80°C until required.

Cells were re-suspended by pipetting in approx. 5 volumes of Lysis buffer at 4°C before lysis using a Constant Systems cell disruptor at 25Kp.s.i. Lysates were cleared by centrifugation at 25000g for 30 minutes at 4°C and the supernatants loaded onto 1ml HisTrap HP columns (GE Healthcare) using an ÄktaXpress FPLC system (GE Healthcare). After sample loading the columns were washed with 10 column volumes of wash buffer before elution with 20 column volumes of elution buffer, collected in 1ml fractions. Peak fractions were analyzed by SDS-PAGE and those containing the N-His-SUMO-sSHH proteins were digested overnight at 4°C with N-His-SUMO protease (produced in IRB PECF).

Digested proteins were then buffer exchanged into TBS (50mM Tris pH8, 150mM NaCl) before re-loading onto fresh 1ml HisTrap HP columns. The flowthrough fractions containing the free hSHH fragments were collected and pooled.

Pooled fractions digested with His-SUMO protease (500µl of 1.25mg/ml) overnight at 4°C, desalted into TBS (50mM Tris pH8, 150mM NaCl) for reverse IMAC.