

ESTUDI DE LA REACCIÓ DE TRANS-SPLICING I CARACTERITZACIÓ FÍSICOQUÍMICA DE LA INTEÏNA PARTIDA Neq DNA POLIMERASA B DE L'HIPERTERMÒFIL N. EQUITANS

Verónica Gordo Hernández

ADVERTIMENT. L'accés als continguts d'aquesta tesi doctoral i la seva utilització ha de respectar els drets de la persona autora. Pot ser utilitzada per a consulta o estudi personal, així com en activitats o materials d'investigació i docència en els termes establerts a l'art. 32 del Text Refós de la Llei de Propietat Intel·lectual (RDL 1/1996). Per altres utilitzacions es requereix l'autorització prèvia i expressa de la persona autora. En qualsevol cas, en la utilització dels seus continguts caldrà indicar de forma clara el nom i cognoms de la persona autora i el títol de la tesi doctoral. No s'autoritza la seva reproducció o altres formes d'explotació efectuades amb finalitats de lucre ni la seva comunicació pública des d'un lloc aliè al servei TDX. Tampoc s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant als continguts de la tesi com als seus resums i índexs.

ADVERTENCIA. El acceso a los contenidos de esta tesis doctoral y su utilización debe respetar los derechos de la persona autora. Puede ser utilizada para consulta o estudio personal, así como en actividades o materiales de investigación y docencia en los términos establecidos en el art. 32 del Texto Refundido de la Ley de Propiedad Intelectual (RDL 1/1996). Para otros usos se requiere la autorización previa y expresa de la persona autora. En cualquier caso, en la utilización de sus contenidos se deberá indicar de forma clara el nombre y apellidos de la persona autora y el título de la tesis doctoral. No se autoriza su reproducción u otras formas de explotación efectuadas con fines lucrativos ni su comunicación pública desde un sitio ajeno al servicio TDR. Tampoco se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al contenido de la tesis como a sus resúmenes e índices.

WARNING. Access to the contents of this doctoral thesis and its use must respect the rights of the author. It can be used for reference or private study, as well as research and learning activities or materials in the terms established by the 32nd article of the Spanish Consolidated Copyright Act (RDL 1/1996). Express and previous authorization of the author is required for any other uses. In any case, when using its content, full name of the author and title of the thesis must be clearly indicated. Reproduction or other forms of for profit use or public communication from outside TDX service is not allowed. Presentation of its content in a window or frame external to TDX (framing) is not authorized either. These rights affect both the content of the thesis and its abstracts and indexes.



TESI DOCTORAL

***Estudi de la reacció de trans-splicing i
caracterització fisicoquímica de la inteïna
partida Neq DNA polimerasa B de
l'hipertermòfil N. equitans***

Verónica Gordo Hernández

2021



TESI DOCTORAL

***Estudi de la reacció de trans-splicing i
caracterització fisicoquímica de la inteïna partida
Neq DNA polimerasa B de l'hipertermòfil N.
equitans***

VERÓNICA GORDO HERNÁNDEZ

2021

PROGRAMA DE DOCTORAT EN QUÍMICA

Dirigida per:

Dr. Marc Ribó i Panosa

Tutor:

Dr. Marc Ribó i Panosa



El Dr. Marc Ribó i Panosa, de la Universitat de Girona,

DECLARO:

Que el treball titulat Estudi de la reacció de *trans-splicing* i caracterització fisicoquímica de la inteïna partida Neq DNA polimerasa B de l'hipertermòfil *N.equitans*, que presenta Verónica Gordo Hernández per a l'obtenció del títol de doctora, ha estat realitzat sota la meva direcció.

I, perquè així consti i tingui els efectes oportuns, signo aquest document.

Dr. Marc Ribó i Panosa

Girona, 15 de Març de 2021

Aquesta tesi ha estat realitzada gràcies a la beca de doctorat BR-UdG (2010) i amb el suport del Ministerio de Educación y Ciencia (Proyecto BFU2006-15543-CO2-02/BMC) i del Ministerio de Ciencia y Innovación (Proyecto BFU2009-06935/BMC).

Publicació derivada de la tesi

Verónica Gordo, David Aparicio, Rosa Pérez-Luque, Antoni Benito, Maria Vilanova, Isabel Usón, Ignacio Fita, Marc Ribó. *Structural Insights into Subunits Assembly and the Oxyester Splicing Mechanism of Neq pol Split Intein*. Cell. Chem. Bio, vol. 25, no. 7, pp. 871-879, 2018.
DOI: 10.1016/j.chembiol.2018.04.008

Abreviacions

A	adenina
ANS	8-anilino-1-naphthalenesulfonic acid
Amp	ampicil·lina
β ME	β -mercaptoetanol
C	citosina
CD	dicroisme circular
C-terminal	carboxi terminal
Da, kDa	Dalton, kilodalton
ddNTPs	dideoxiribonucleòtid
Deg, mdeg	(mil·li)graus d'el·lipticitat
DNA	àcid desoxiribonucleic
OD	densitat òptica
e	exponencial
EDTA	àcid etilendiaminotetracètic
G	guanina
FPLC	fast protein liquid chromatography
IntN	inteïna N
IntC	inteïna C
IPTG	isopropil- β -D-tiogalactopirànòsid
Kan	kanamicina
Kd	constant de dissociació
LB	medi Lüría-Bertrani
M, mM, μ M	molar, mil·limolar i micromolar
MALDI-TOF	espectrometria de massa de desopció / ionització làser assistida per matriu
Min	minut
Mw	massa molecular
N-ter	amino terminal
PAGE	electroforesi en gel de poliacrilamida
Pb	parells de bases
PCR	reacció en cadena de la polimerasa
PDB	banc de dades de proteïnes
pI	punt isoelèctric

psi	lliura per polzada quadrada
RMN	ressonància magnètica nuclear
RMSD	error quadràtic mig
RNA	àcid ribonucleic
Rpm	revolucions per minut
SDS	dodecil fosfat sòdic
SICLOPPS	l·ligació circular de pèptids i proteïnes mitjançant inteïnes partides
T	timina
TCA	tricloroacètic
TEMED	<i>N,N,N',N'</i> -tetrametiletildiamina
TEV	tobaco etch virus
Tris	trishidroximetilaminometà
UV	ultraviolat
Wt	wild type

Codi d'aminoàcids

<i>Codi de tres lletres</i>	<i>Codi d'una lletra</i>	<i>Aminoàcid</i>
<i>Ala</i>	A	Alanina
<i>Arg</i>	R	Arginina
<i>Asn</i>	N	Asparagina
<i>Asp</i>	D	Aspàrtic
<i>Cys</i>	C	Cisteïna
<i>Glu</i>	E	Glutàmic
<i>Gln</i>	Q	Glutamina
<i>Gly</i>	G	Glicina
<i>His</i>	H	Histidina
<i>Ile</i>	I	Isoleucina
<i>Leu</i>	L	Leucina
<i>Lys</i>	K	Lisina
<i>Met</i>	M	Metionina
<i>Phe</i>	F	Fenilalanina
<i>Pro</i>	P	Prolina
<i>Ser</i>	S	Serina
<i>Thr</i>	T	Treonina
<i>Trp</i>	W	Triptòfan
<i>Tyr</i>	Y	Tirosina
<i>Val</i>	V	Valina

Index General

Índex de figures	V
Índex de taules	VIII
Resum	1
Resumen	3
Summary	5
1. Introducció	7
1.1. Les inteïnes	9
1.1.1. Organització i estructura de les inteïnes.....	10
1.1.2. Inteïnes com a interruptors ambientals: Conditional Protein <i>Splicing</i> (CPS).....	14
1.1.3. Mecanisme de <i>splicing</i>	15
1.1.4. Classes d'inteïnes.....	19
1.1.5. Distribució dels aminoàcids nucleòfils en les posicions 1 i +1.....	20
1.1.6. <i>Trans-splicing</i>	22
1.1.7. Aplicacions.....	26
1.2. Adaptació de les proteïnes a elevades temperatures	30
1.2.1. Composició d'aminoàcids.....	31
1.2.2. Optimització de l'estructura terciària i interaccions no covalents.....	31
1.3. La inteïna partida Neq pol	34
2. Objectius	37
3. Material i Mètodes	41
3.1. Material biològic	43
3.1.1. Soques bacterianes.....	43
3.1.2. Vectors procarïotes utilitzats per a l'expressió de proteïnes recombinants.....	43
3.2. Tècniques de DNA recombinant	44
3.2.1. Extracció i purificació de DNA plasmídic.....	44
3.2.2. Digestió amb enzims de restricció.....	45
3.2.3. Mutagènesi dirigida.....	45
3.2.4. Disseny dels oligonucleòtids.....	45
3.2.5. Hibridació de nucleòtids.....	45
3.2.6. Electroforesi en gel d'agarosa.....	47
3.2.7. Purificació de fragments de DNA a partir de gels d'agarosa (GenClean).....	47
3.2.8. Lligació de l'insert de DNA amb el vector d'expressió utilitzant la T4 lligasa.....	47
3.2.9. Transformació de cèl·lules competents d' <i>E. coli</i>	47
3.2.10. Seqüenciació de DNA.....	48

3.3. Clonatge, expressió i purificació.....	48
3.3.1. Construccions per la caracterització de la inteïna N de <i>N. equitans</i>	48
3.3.2. Construccions per a la caracterització de la reacció de <i>trans-splicing</i> , del procés d'associació i del complex NeqN/NeqC.....	53
3.3.3. Producció i purificació de la Proteasa H6-TEV.....	60
3.4. Anàlisi i caracterització proteica	62
3.4.1. Anàlisi seqüencial de la NeqN i NpuN	62
3.4.2. Bradford	63
3.4.3. Electroforesi discontinua en gel d'acrilamida amb SDS (SDS-PAGE)	63
3.4.4. Anàlisi de solubilitat a 37°C.....	63
3.4.5. Estudi del procés de <i>trans-splicing</i>	64
3.4.6. Western-Blot	64
3.4.7. Proteòlisi limitada	65
3.4.8. Determinació de l'estabilitat tèrmica per UV.....	66
3.4.9. Determinació de l'estabilitat tèrmica per Calorimetria diferencial d'escombrat (DSC).....	67
3.4.10. Caracterització per dicromisme Circular (CD)	67
3.4.11. Assaig d'unió a ANS	68
3.4.12. Cromatografia d'exclusió molecular (SEC)	68
3.4.13. Ultracentrifugació analítica	69
3.4.14. Caracterització fisicoquímica del procés d'associació entre els dos fragments de la Neq pol	69
3.4.15. Formació del complex NeqN/NeqC <i>in vitro</i>	71
3.4.16. Cristal·lització i Difracció de raig X	72
4. Resultats i Discussió	75
4.1. Caracterització de la reacció de <i>trans-splicing</i> catalitzada per la inteïna partida Neq pol.....	77
4.1.1. Reacció de <i>trans-splicing</i> utilitzant com a exteïna N la MBP i com a exteïna C el domini GB1	79
4.1.2. Reacció de <i>cis-splicing</i> utilitzant com a exteïna N el domini GB1 i com a exteïna C el domini ZHER2	85
4.1.3. Reacció de <i>trans-splicing</i> utilitzant com a exteïna N el domini GB1 i com a exteïna C el domini ZHER2	91
4.1.4. Efecte de la Ser1 en la reacció de <i>trans-splicing</i> catalitzada pels precursors H6GB1(wt)-NeqN i NeqC-(TDS)ZHER2.....	98

4.2.	Caracterització de la inteïna N de <i>N.equitans</i>, NeqN	101
4.2.1.	Caracterització fisicoquímica de la NeqN	104
4.2.2.	Resolució de l'estructura tridimensional per cristal·lografia i difracció de raig X	117
4.3.	Caracterització del procés d'associació entre els dos fragments de la Inteïna partida Neq pol i del complex NeqN/NeqC.....	126
4.3.1.	Caracterització fisicoquímica del procés d'associació entre els dos fragments de la Inteïna partida Neq	127
4.3.2.	Caracterització fisicoquímica del complex NeqN/NeqC.....	130
4.3.3.	Resolució de l'estructura tridimensional per cristal·lografia.....	133
4.3.4.	Centre catalític de la Neq pol: Implicacions en el mecanisme de <i>splicing</i>	148
5.	<i>Conclusions</i>	151
6.	<i>Bibliografia</i>	157
7.	<i>Annex</i>.....	177

Índex de figures

Figura 1.1.	Precursor en cis.	11
Figura 1.2.	Esquema dels diferents tipus d'inteïnes trobades a la natura.	12
Figura 1.3.	Estructura esquemàtica dels precursors que catalitzen la reacció de <i>splicing</i>	12
Figura 1.4.	Esquema comparatiu de la reacció de <i>splicing</i> i la reacció de <i>Conditional Protein Splicing</i> (CPS).	14
Figura 1.5.	Mecanisme de <i>splicing</i>	16
Figura 1.6.	Reaccions de hidròlisi dels extrems terminals.....	19
Figura 1.7.	Mecanisme de <i>splicing</i> de les tres classes d'inteïnes.	20
Figura 1.8.	Distribució dels diferents nucleòfils en les posició 1 i +1.	21
Figura 1.9.	Model proposat per la formació d'inteïnes partides.....	23
Figura 1.10.	Esquema del mecanisme de <i>trans-splicing</i>	24
Figura 1.11.	Esquema del mecanisme de captura i col·lapse per l'associació dels dos fragments de la inteïna partida Npu DnaE.	25
Figura 1.12.	<i>Expressed Protein Ligation</i> (EPL).....	27
Figura 1.13.	Esquema de la generació de proteïnes cíclics a través de la reacció de <i>trans-splicing</i>	28
Figura 1.14.	<i>Conditional Protein Splicing</i> (CPS).	30
Figura 1.15.	Mapa del genoma de <i>N. Equitans</i> amb l'estructura de la inteïna partida Neq pol.....	35
Figura 3.1.	Esquema de l'obtenció de les diferents construccions utilitzades en el procés de caracterització de la NeqN.....	52
Figura 4.1.	Esquema del precursor N, H6MBP(GD)-NeqN, i del precursor C, NeqC-(TD)GB1.	79
Figura 4.2.	Coexpressió recombinant a 37°C i copurificació del complex format per H6MBP(GD)-NeqN i NeqC-(TD)GB1.	80
Figura 4.3.	Esquema simplificat de la reacció de <i>trans-splicing</i> i de les reaccions secundàries d'hidròlisi catalitzada pels precursors H6MBP(GD)-NeqN i NeqC-(TD)GB1.	81
Figura 4.4.	Anàlisi per SDS-PAGE de la reacció de <i>trans-splicing</i> a 50°C catalitzada pel complex format <i>in vivo</i> pel precursor N, H6MBP(GD)-NeqN, i el precursor C, NeqC-(TD)GB1.	82
Figura 4.5.	Esquema del precursor C, NeqC-(TD)GB1H6.	82
Figura 4.6.	Anàlisi per SDS-PAGE de la reacció de <i>trans-splicing</i> a 50°C catalitzada pel precursor N, H6MBP(GD)-NeqN, i el precursor C, NeqC-(TD)GB1 a diferents pHs.....	84
Figura 4.7.	Anàlisi per SDS-PAGE de la reacció de <i>trans-splicing</i> catalitzada pels precursors N i C, H6MBP(GD)-NeqN i NeqC-(TD)GB1H6 a 70°C.	84
Figura 4.8.	Esquema del precursor en cis H6MBP(D)-NeqSC-(TDS)ZHER2.	86
Figura 4.9.	Esquema de la reacció de <i>cis-splicing</i> catalitzada pel precursor H6GB1(D)-NeqSC-(TDS)ZHER2, i les possibles reaccions no desitjades d'hidròlisi.....	87
Figura 4.10.	Anàlisi de la reacció de <i>cis-splicing</i> per SDS-PAGE catalitzada pel precursor precursor H6GB1(D)-NeqSC-(TDS)ZHER2 a pH 8.0 en un rang de temperatures entre els 40°C i 70°C.	88

Figura 4.11.	Anàlisi per SDS-PAGE de la reacció de <i>cis-splicing</i> catalitzada pel precursor H6GB1(D)-NeqSC-(TDS)ZHER2 a pH 5.0 en un rang de temperatures entre 40°C i 70°C.	90
Figura 4.12.	Anàlisi per SDS-PAGE de la reacció de <i>cis-splicing</i> catalitzada pel precursor H6GB1(D)-NeqSC-(TDS)ZHER2 a pH 7.2 en un rang de temperatures entre 40°C i 70°C.	90
Figura 4.13.	Esquema de precursor N, H6GB(wt)-NeqN, i C, NeqC-(TDS)ZHER2.	92
Figura 4.14.	Anàlisi per SDS-PAGE de la complexitat del precursor N, H6GB1(wt)-NeqN i la NeqN.....	93
Figura 4.15.	Cinètica del procés de desnaturalització a 95°C amb 4 M d'urea analitzada per SDS-PAGE del precursor H6GB(wt)-NeqN i del complex format per H6GB(wt)-NeqN i NeqC-(TDS)ZHER2.	94
Figura 4.16.	Esquema de la reacció de <i>trans-splicing</i> i de les reaccions secundàries del precursor N, H6GB1(wt)NeqN, i el precursor C, NeqC-(TDS)ZHER2.	95
Figura 4.17.	Anàlisi per SDS-PAGE de la reacció de <i>trans-splicing in vitro</i> a 70°C catalitzada pels precursors: H6GB1(wt)-NeqN i NeqC-(TDS)ZHER2 en un rang de pH entre 3.5 i 9.0.	96
Figura 4.18.	Anàlisi per SDS-PAGE de la reacció de <i>trans-splicing in vitro</i> catalitzada pels precursors: H6GB1(wt)-NeqN i NeqC-(TDS)ZHER2 en un rang de temperatures entre els 40°C i 70°C.	97
Figura 4.19.	Anàlisi per SDS-PAGE de la reacció de <i>trans-splicing</i> catalitzada pel complex format <i>in vivo</i>	99
Figura 4.20.	Anàlisi per SDS-PAGE de la reacció de <i>trans-splicing</i> catalitzada pel complex format <i>in vivo</i> H6GB1(wt)-NeqN(S1C)/NeqC-(TDS)ZHER2 a pH 7.0 i temperatures de 70°C i 40°C.	100
Figura 4.21.	Anàlisi per SDS-PAGE de solubilitat de la T7-NeqN.	101
Figura 4.22.	Purificació per cromatografia de bescanvi aniònic de la construcció T7-NeqN.	102
Figura 4.23.	Anàlisi per SDS-PAGE del procés de purificació de la T7-NeqN per cromatografia d'afinitat Ni-NTA.	103
Figura 4.24.	Alineament de les seqüències de les inteïnes partides Neq pol, SspDnaE i Npu DnaE.	104
Figura 4.25.	Seqüència polipeptídica de la construcció NeqN.	105
Figura 4.26.	Anàlisi bioinformàtica de la NeqN i comparació amb la NpuN.....	106
Figura 4.27.	Anàlisi per SDS-PAGE de la solubilitat de les construccions H6TEV(wt)-NeqN i H6TEV-NpuN.	107
Figura 4.28.	Anàlisi per SDS-PAGE de la resistència de la NeqN a la digestió per part de la proteïna K.	109
Figura 4.29.	Anàlisi per MALDI-TOF de la digestió de la NeqN.....	109
Figura 4.30.	Anàlisi per SDS-PAGE <i>in vitro</i> de l'estabilitat conformacional de la inteïna partida NeqN mitjançant proteòlisi limitada.	110
Figura 4.31.	Corbes normalitzades de desplegament i replegament per temperatura de la NeqN a diferents pHs.	111
Figura 4.32.	Termograma obtingut per DSC del procés de desnaturalització de la NeqN. ..	112
Figura 4.33.	Espectres de CD de la proteïna NeqN.	113
Figura 4.34.	Espectre d'unió a ANS de la NeqN.....	114
Figura 4.35.	Caracterització de la NeqN per cromatografia d'exclusió molecular (SEC).	114
Figura 4.36.	Determinació experimental de la massa molecular de la NeqN.	116
Figura 4.37.	Estudi de l'efecte de la concentració en la transició de l'estat plegat-desplegat de la NeqN per DSC.	117
Figura 4.38.	Estructura tridimensional obtinguda per cristal·lografia de la NeqN.	119
Figura 4.39.	Diagrama de l'estructura secundària de la NeqN aïllada.	120

Figura 4.40.	Superposició de les estructures del dímer AD i del dímer CB.....	120
Figura 4.41.	Interaccions entre l'exteïna N i la NeqN.....	122
Figura 4.42.	Interaccions presents en la interfície dimèrica.....	123
Figura 4.43.	Comparació de l'estructura de la NeqN aïllada i la NpuN formant el complex NpuN DnaE. .	124
Figura 4.44.	Representació del caràcter electroestàtic i hidrofòbic de les dues meitats de NeqN assumint el model de dos lòbuls proposats per la inteïna Npu.....	124
Figura 4.45.	Anàlisi de dimerització de la NeqN per SEC.....	125
Figura 4.46.	Anàlisi per SDS-PAGE de la solubilitat del precursor C, NeqC.	126
Figura 4.47.	Determinació de la massa molecular del complex NeqN/NeqC.....	131
Figura 4.48.	Espectres de dicroisme circular de la NeqN, de la NeqC i del complex NeqN/NeqC.	132
Figura 4.49.	Termograma obtingut per DSC del procés de desnaturalització de la NeqN i del complex NeqN/NeqC.....	133
Figura 4.50.	Estructura cristal·logràfica obtinguda del complex NeqN/NeqC.....	134
Figura 4.51.	Sobreposició de les dues estructures del complex obtingudes dels cristalls hexagonals i dels cristalls tetragonals.	135
Figura 4.52.	Diagrama de l'estructura secundària de la NeqN i de la NeqC en el complex NeqN/NeqC.	137
Figura 4.53.	Complementació electroestàtica existent entre la NeqN i la NeqC.	139
Figura 4.54.	Estructura cristal·logràfica de la Neq N aïllada, on es ressalten els residus àcids implicats en el procés d'associació.....	140
Figura 4.55.	Distàncies entre els residus de l'exteïna N i la NeqC i l'exteïna C.....	141
Figura 4.56.	Sobreposició de la NeqN monomèrica.	142
Figura 4.57.	Esquema de l'estructura secundària de la NeqN aïllada i en el complex NeqN/NeqC.	143
Figura 4.58.	Efecte sobre el Trp52 del canvi conformacional de la NeqN provocat per l'associació amb la NeqC. .	144
Figura 4.59.	Alineament seqüencial basat en l'estructura de la NeqN formant complex amb altres inteïnes les quals presenten una elevada homologia estructural.	145
Figura 4.60.	Superposició de les estructures tridimensionals de la Neq pol amb altres inteïnes homòlogues.	147
Figura 4.61.	Centre actiu de la Neq pol.	149

Índex de taules

Taula 1.1.	Proteïnes més comuns envaïdes per inteïnes.	10
Taula 1.2.	Comparació del nombre de subunitats entre proteïnes hipertermofíliques i mesofíliques.	34
Taula 3.1.	Oligonucleòtids utilitzats en aquest treball.	46
Taula 3.2.	Resum de les construccions utilitzades en aquest treball.	61
Taula 4.1.	Resum dels precursors utilitzats per a catalitzar la reacció de <i>splicing</i>	78
Taula 4.2.	Masses moleculars dels productes obtinguts en la reacció de <i>trans-splicing</i> catalitzada pel precursor N, H6MBP(GD)-NeqN, i el precursor C, NeqC-(TD)GB1H6 <i>in vitro</i>	83
Taula 4.3.	Diferents tampons utilitzats per l'estudi de l'efecte del pH sobre la reacció de <i>trans-splicing</i>	83
Taula 4.4.	Masses moleculars dels productes obtinguts en la reacció de <i>trans-splicing</i> catalitzada pel precursor N, H6GB1(wt)-NeqN (S1C), i el precursor C, NeqC-(TDS)ZHER2.	99
Taula 4.5.	Col·lecció de dades i estadístics del refinament.	118
Taula 4.6.	Constants de velocitat de la reacció d'associació entre la NeqN i la NeqC a diferents temperatures.	127
Taula 4.7.	Energia d'activació de la reacció d'associació entre la NeqN i la NeqC.	128
Taula 4.8.	Constant d'afinitat i paràmetres termodinàmics del procés d'associació entre la NeqN i la NeqC.	128
Taula 4.9.	Col·lecció de dades i estadístics del refinament.	134
Taula 4.10.	Paràmetres fisicoquímics de la interacció predits pel servidor PISA.	138
Taula 4.11.	Ponts d'hidrogen entre exteïna N i exteïna C, i exteïna N i inteïna C.	141

Resum

L'*splicing* de proteïnes és un procés autocatalític compost per diferents etapes que implica l'eliminació d'una seqüència interna, anomenada inteïna, de la cadena polipeptídica precursora i la formació d'un enllaç peptídic entre les regions flanquejants, anomenades exteïnes N i C. Les inteïnes partides o *trans*-inteïnes són un tipus d'inteïnes que estan codificades en dos gens diferents juntament amb els gens de l'exteïna N i l'exteïna C.

El gen de la DNA polimerasa B de l'arqueobacteri hipertermòfil *Nanoarchaeum equitans* està dividit en dos marcs oberts de lectura. Els gens que codifiquen per les regions o dominis N-terminal i C-terminal de la polimerasa (exteïnes N i C) estan fusionats, respectivament, als gens que codifiquen per la inteïna N i la inteïna C de la inteïna partida Neq pol.

En la present tesi s'estudia la reacció de *trans-splicing* catalitzada per la inteïna partida Neq pol, així com es caracteritzen les propietats fisicoquímiques del fragment N-terminal, NeqN, i del complex NeqN/NeqC. D'altra banda, també s'estudia des d'un punt de vista cinètic i termodinàmic el procés d'associació entre els dos fragments de la inteïna.

La reacció de *trans-splicing* catalitzada *in vitro* per la inteïna Neq pol només es produeix a temperatures superiors als 50°C, amb una temperatura òptima de 80°C. Aquest fet indica que la temperatura regula la reacció de *trans-splicing* i que aquesta regulació pot produir-se per la pròpia inteïna o bé per les exteïnes. Amb l'objectiu d'esbrinar si la Neq pol és una inteïna dependent de temperatura, s'ha estudiat la reacció de *trans-splicing* en un context no natiu utilitzant diferents proteïnes com a exteïnes. Els resultats obtinguts suggereixen per una banda que la inteïna Neq pol no tolera els canvis d'exteïna, ja que només s'obté el producte de *trans-splicing* quan s'utilitza la GB1 i ZHER2 com a exteïnes N i C, respectivament, i sempre que es conservin els set aminoàcids salvatges de la exteïna N en la regió adjacent a la Ser1 de la inteïna N i els 3 residus nadius de l'extrem N-terminal de l'exteïna C. Aquest fet indica que els residus que es troben en les regions flanquejants del lloc d'unió inteïna-exteïna juguen un paper clau en la catàlisi de la reacció, segurament, orientant correctament els residus catalítics del centre actiu de la inteïna. I per altra banda els resultats obtinguts indiquen que la regulació de la reacció per la temperatura ve donada per la inteïna, ja que en aquest context no natiu el producte de la reacció de *trans-splicing* només s'observa a temperatures superiors als 50°C. La Neq pol és l'única inteïna partida descrita fins ara que conté com a residus catalítics una Ser en posició 1 i una Thr en posició +1, convertint-se en una excel·lent candidata per ser utilitzada en el marcatge isotòpic de proteïnes. Per a entendre millor el paper que juga en la reacció de *trans-splicing* la

Ser en posició 1, es va mutar per una Cys. Les dades experimentals obtingudes indiquen que la mutació Ser1Cys augmenta l'eficiència de la reacció.

Paral·lelament, s'ha caracteritzat des d'un punt de vista fisicoquímic i estructural la inteïna N lliure, la qual presenta propietats característiques de les proteïnes hipertermofíliques, com ara T_m elevada, resistència a la desnaturalització per agents caotròpics, augment del grau d'empaquetament del nucli hidrofòbic i interaccions aromàtiques, el que ha permès obtenir-la primera estructura tridimensional del fragment N-terminal d'una inteïna partida.

D'altra banda i gràcies al fet que la Neq pol és una inteïna depenent de temperatura s'ha pogut obtenir mitjançant cristal·lografia i difracció de raigs X l'estructura tridimensional del complex NeqN/NeqC i dur a terme la seva caracterització fisicoquímica. En l'estructura s'han pogut resoldre per primer cop els residus catalítics salvatges de la inteïna i també els residus nadius de les exteïnes N i C.

Finalment, utilitzant tècniques espectroscòpiques de fluorescència s'ha caracteritzat des d'un punt de vista cinètic i termodinàmic el procés d'associació dels dos fragments de la inteïna partida Neq pol. Com era d'esperar tenint en compte el caràcter hipertermofílic de la inteïna la velocitat d'associació augmenta gradualment amb la temperatura, obtenint el màxim de velocitat experimental a 64°C. Respecte a la termodinàmica del procés, els estudis experimentals mostren que es tracta d'un procés espontani ($\Delta G = -1.73$ kcal/mol) amb un canvi entròpic entàlpic positiu, que indica que el mecanisme d'associació dels dos fragments de la inteïna Neq pol és diferent del proposat per altres inteïnes partides.

Resumen

El *splicing* de proteínas es un proceso autocatalítico constituido por diferentes etapas que implica la eliminación de una secuencia interna, llamada inteína, de la cadena polipeptídica precursora y la formación de un enlace peptídico entre las regiones flanqueantes llamadas exteínas N i C. Las inteínas partidas o *trans*-inteínas son un tipo de inteínas que están codificadas en dos genes diferentes juntamente con los genes de la exteína N y de la exteína C.

El gen de la DNA polimerasa B del arqueobacterio hipertermófilo *Nanoarchaeum equitans* se encuentra dividido en dos marcos abiertos de lectura. Los genes que codifican para la región o dominios N-terminal i C-terminal de la polimerasa (exteínas N i C) están fusionados, respectivamente, con los genes que codifican para la inteína N i la inteína C de la inteína partida Neq pol.

En la presente tesis se estudia la reacción de *trans-splicing* catalizada por la inteína partida Neq pol, así como se caracterizan las propiedades fisicoquímicas del fragmento N-terminal, NeqN, y del complejo NeqN/NeqC. Por otro lado, también se estudia desde un punto cinético y termodinámico el proceso de asociación entre los dos fragmentos.

La reacción de *trans-splicing* catalizada *in vitro* por la inteína Neq pol solamente se produce a temperaturas superiores a los 50°C, con una temperatura óptima de 80°C. Este hecho sugiere que la temperatura regula la reacción de *trans-splicing* y que esta regulación puede producirse por la propia inteína o bien por las exteínas. Con el objetivo de conocer si la Neq pol es una inteína dependiente de temperatura, se ha estudiado la reacción de *trans-splicing* en un contexto no nativo utilizando diferentes proteínas como exteínas. Los resultados obtenidos sugieren por un lado que la inteína Neq pol no tolera los cambios de exteína, debido a que solo se obtiene el producto de la reacción de *trans-splicing* cuando se utiliza la GB1 i ZHER2 como exteína N i C, respectivamente, y siempre que se conserven los siete aminoácidos salvajes de la exteína N en la región adyacente a la Ser1 de la inteína N i los 3 residuos nativos del extremo N-terminal de la exteína C. Este hecho indica que los residuos que se encuentran en las regiones flanqueantes al sitio de unión inteína-exteína juegan un papel clave en la catálisis de la reacción; probablemente, orientando correctamente los residuos catalíticos del centro activo de la inteína. Por otro lado, los resultados obtenidos indican que la regulación de la reacción por la temperatura viene dada por la inteína, debido a que en este contexto no nativo solamente se observa el producto de *trans-splicing* a temperaturas superiores a los 50°C. La Neq pol es la única inteína partida descrita hasta ahora que tiene como residuos catalíticos una Ser en posición 1 i

una Thr en posición +1, convirtiéndose en una excelente candidata para ser utilizada en el marcaje isotópico de proteínas. Para entender mejor el papel que juega en la reacción de *trans-splicing* la Ser en posición 1, se ha mutado por una Cys. Los datos experimentales obtenidos indican que la mutación Ser1Cys aumenta la eficiencia de la reacción.

Paralelamente, se ha caracterizado desde un punto de vista fisicoquímico y estructural la inteína N libre, la cual presenta propiedades características de las proteínas hipertermófilas, como una T_m alta, resistencia a la desnaturalización por agentes caotrópicos, aumento del grado de empaquetamiento del núcleo hidrofóbico, interacciones aromáticas... lo que ha permitido obtener la primera estructura tridimensional del fragmento N-terminal de una inteína partida. También y gracias a que la Neq pol es una inteína dependiente de temperatura se ha podido obtener mediante cristalografía y difracción de rayos X la estructura tridimensional del complejo NeqN/NeqC, y realizar su caracterización fisicoquímica. En la estructura se han podido resolver por primera vez los residuos catalíticos salvajes de la inteína y también los residuos nativos de las exteínas N i C.

Finalmente, utilizando técnicas espectroscópicas de fluorescencia se ha caracterizado desde un punto de vista cinético y termodinámico el proceso de asociación entre los dos fragmentos de la inteína partida Neq pol. Como era de esperar teniendo en cuenta el carácter termofílico de la inteína, la velocidad de asociación aumenta gradualmente con la temperatura, obteniendo la velocidad máxima experimental a 64°C. Respecto a la termodinámica del proceso, los estudios experimentales demuestran que se trata de proceso espontáneo ($\Delta G = -1.73$ kcal/mol) con un cambio entálpico y entrópico positivo, lo que indica que el mecanismo por el cual los dos fragmentos asocian es diferente al propuesto por otras inteínas partidas.

Summary

Protein splicing is an autocatalytic process with different stages that involves the elimination of an internal sequence, called intein, from the precursor polypeptide chain and the formation of a peptide bond between the flanking regions called N and C exteins. The split inteins or trans-inteins are a type of inteins that are encoded in two different genes together with the N extein and C extein genes.

The DNA polymerase B gene of the hyperthermophilic archaeobacterium *Nanoarchaeum equitans* is divided into two open reading frames. The genes coding for the N-terminal and C-terminal region or domains of the polymerase (N and C exteins) are fused respectively with the genes coding for N intein and C intein of the starting split intein Neq pol.

The present thesis studies, the trans-splicing reaction catalyzed by the split intein Neq pol, as well as characterizes the physicochemical properties of the N-terminal fragment, NeqN, and the NeqN/ NeqC complex. On the other hand, the association process between the two fragments is also studied from a kinetic and thermodynamic point of view.

The *in vitro* catalyzed trans-splicing reaction by Neq pol intein only occurs at temperatures above 50°C, with an optimal temperature of 80°C. This fact suggests that temperature regulates the trans-splicing reaction, and that this regulation can be produced by the intein or by the exteins. With the aim of knowing if Neq pol is a temperature-dependent intein, the trans-splicing reaction was studied in a non-native context using different proteins as exteins. The results suggest that the intein Neq pol does not tolerate the changes of extein, because the product of the trans-splicing reaction is only obtained when GB1 and ZHER2 are used as N and C extein, respectively, and provided that the seven wild-type amino acids of N extein in the region adjacent to Ser1 of intein N and the 3 native residues of the N-terminal end of C extein are conserved. This fact indicates that the residues found in the regions flanking the intein-extein binding site plays a key role in catalyzing the reaction; probably, by orienting correctly the catalytic residues of the active center of the intein. In this context, the product of the trans-splicing reaction is only observed at temperatures above 50°C, which indicates that the regulation of the reaction by temperature is given by the intein. On the other hand, Neq pol is the only split intein described so far that has as catalytic residue a Ser in position 1 and a Thr in position +1, making it an excellent candidate to be used in the isotopic labeling of proteins. To better understand the role that the Ser in position 1 plays in the trans-splicing reaction, it was

mutated by a Cys. The experimental data obtained indicate that the Ser1Cys mutation increases the efficiency of the reaction.

In parallel, free N intein has been characterized from a physicochemical and structural point of view. Free N intein has the properties of hyperthermophilic proteins, such as a high T_m , resistance to denaturation by chaotropic agents, increased packing degree of the hydrophobic nucleus, aromatic interactions... which allowed to obtain the first three-dimensional structure of the N-terminal fragment of a split intein.

Also, and thanks to the fact that Neq pol is a temperature-dependent intein, it has been possible to obtain the three-dimensional structure of the NeqN/NeqC complex by crystallography and X-ray diffraction, and to carry out its physicochemical characterization. In the structure it has been resolved for the first time the wild catalytic residues of the intein and also the native residues of the N and C exteins.

Finally, the association process between the two fragments of the split Neq pol intein has been characterized from a kinetic and thermodynamic point of view using fluorescence spectroscopic techniques. As expected, and taking into account the thermophilic nature of intein, the rate of association gradually increases with temperature, obtaining the maximum experimental rate at 64°C. Regarding the thermodynamics of the process, our experimental studies show that it is a spontaneous process ($\Delta G = -1.73$ kcal/mol) with a positive enthalpic and entropic change, which indicates that the mechanism by which the two fragments associate is different to the one proposed by other split inteins.

1. Introducció

1.1. Les inteïnes

Les inteïnes es poden considerar anàlegs dels introns presents en els pre-mRNAs, ja que s'escindeixen autocatalíticament de la proteïna hoste (procés de *splicing*) per donar lloc a la proteïna activa. D'altra banda, les inteïnes també són elements genètics mòbils que envaeixen gens hostes a nivell de DNA i després s'expressen com a seqüències intermèdies dintre de la proteïna [1].

Les inteïnes van ser descobertes per primer cop l'any 1990 en la subunitat catalítica de l'ATPasa translocadora de H⁺ de la membrana vacuolar (TFP1/VMA1) de *Saccharomyces cerevisiae* [2], [3]. Els autors van observar que a partir d'un únic precursor proteic d'elevada massa molecular s'obtenien dues proteïnes, una que corresponia a la subunitat de l'ATPasa i una altra que corresponia a la inteïna.

Les inteïnes pertanyen a la superfamília Hint (Hedge-Hog/Intein), la qual està composta per altres dues famílies: dominis Hedgehog-Hint i Bacterial intein-like (BIL) domains, totes elles involucrades en processos postraduccional. Aquesta superfamília, com el seu propi nom indica, es caracteritza per tenir el domini Hint, el qual conté els trets fonamentals que són necessaris perquè es produeixi la reacció de *splicing*. Existeixen diferents formes de dominis Hint, però normalment, totes presenten una llargada entre 140-160 aminoàcids, contenen 4-6 motius conservats i adopten una estructura característica compacta formada per 3 fulls β i dues hèlixs α en forma de ferradura [4]. El correcte plegament d'aquest domini és tot el que necessiten les inteïnes per catalitzar la reacció de *splicing* [5].

Les inteïnes es troben àmpliament distribuïdes en els tres dominis de la vida, Eucarya, Eubacteria i Archaea, així com en virus i bacteriòfags. Trobant-se majoritàriament en els dominis Eubacteria i Archaea, exactament s'han detectat en un quart i la meitat dels genomes dels organismes d'aquests dominis seqüenciats fins ara, respectivament [6]–[12].

Tot i que les inteïnes es troben en proteïnes amb diverses funcions existeix una clara tendència d'aquestes a envair proteïnes relacionades amb el metabolisme del DNA, com ara polimerases, helicases, topoisomereses i ribonucleòtid reductases, particularment en el seu centre actiu (Taula 1.1). La integració de les inteïnes en gens implicats en la replicació, recombinació i replicació del DNA, i el metabolisme dels nucleòtids es pot explicar per la seva mobilitat. Per a moure's les inteïnes necessiten integrar-se en el DNA de l'hoste. S'ha vist que el DNA és més accessible per escindir-se i recombinar-se durant la replicació, per tant, és més eficient que les inteïnes envaeixen proteïnes que s'expressen durant aquest procés cel·lular [11], [12].

Es coneix poc sobre l'origen de les inteïnes, diversitat, evolució i funció en la proteïna hoste. Fins ara, les inteïnes es consideraven un paràsit, les quals no aportaven cap benefici a l'organisme hoste i es conservaven al llarg de l'evolució a causa de la seva localització en dominis altament conservats de proteïnes essencials [13]–[15] o bé gràcies al seu domini endonucleasa (HEN) [16]–[18].

Taula 1.1. Proteïnes més comuns envaïdes per inteïnes [11].

Proteïna	Procés en el que participa	Organisme hoste
Helicasa (DnaB)	Replicació del DNA	Bacteri
Mini-chromosome maintenance protein (MCM)/ cell division control protein (Cdc21)	Replicació del DNA	Arqueobacteri
Replication Factor C (RFC)	Replicació del DNA	Archaea
Recombinasa (RecA/RadA)	Reparació del DNA	Bacteri; Arqueobacteri
SF2 helicasa	Reparació del DNA	Bacteri
SF1 ATP-depenen helicasa	Reparació del DNA	Bacteri
Subunitat α de la subunitat catalítica de la DNA polimerasa III (DnaE)	Replicació del DNA	Bacteri
Subunitat τ de la subunitat catalítica de la DNA polimerasa III (DnaZX)	Replicació del DNA	Bacteri
DNA polimerasa I (PolB)	Replicació DNA	Arqueobacteri
DNA polimerasa II, subunitat DP2 (PolC)	Replicació DNA	Arqueobacteri
Topoisomerasa II- DNA girasa, subunitat B (GyrB)	Replicació DNA	Bacteri; Arqueobacteri; Bacteriòfag
DNA polimerasa I bacteriana (PolA)	Replicació DNA	Bacteri
DNA polimerasa II bacteriana (PolB)	Replicació DNA	Bacteri
SF6 Holliday junction ATP-dependent DNA helicasa (RuvB)	Replicació DNA	Bacteri
Phage terminase, large subunit	Empaquetament DNA	Bacteriòfag
Ribonucleòsid difosfat reductasa (RNR)	Síntesi de DNA	Eucariota; Arqueobacteri
Vacuolar ATP synthase catalytic subunit A (VMA)	Vacuolar H ⁺ -pump ATP	Eucariota; Arqueobacteri

Amb el recent descobriment d'inteïnes que només catalitzen la reacció de *splicing* davant d'un determinat estímul (temperatura, sals, potencial redox...) aquest pensament ha canviat. L'evidència d'aquestes inteïnes dependents d'estímuls suggereix que aquestes han evolucionat per adaptar-se al seu nínxol intracel·lular convertint-se en elements reguladors post-traduccionals que modulen la reacció de *splicing* d'acord amb les condicions ambientals en les quals es troba l'organisme hoste [19]–[26].

1.1.1. Organització i estructura de les inteïnes

Inicialment, l'element que catalitzava la reacció de *splicing* estudiada en *S. cerevisiae* s'anomenava simplement *spacer* [3]. Posteriorment, aquest element es va anomenar seqüència proteica interventora. I finalment, amb el descobriment de múltiples seqüències proteiques interventores es van acabar denominant inteïnes (**internal protein sequence**). Seguint el mateix

conveni, les seqüències aminoacídiques que flanquegen les inteïnes van rebre el nom d'exteïnes (**external protein sequence**). Les exteïnes unides als extrems N- i C- terminal de la inteïna són l'exteïna N i C, respectivament. La cadena polipeptídica que conté la inteïna i les dues exteïnes, s'anomena precursor (Figura 1.1) [1], [27], [28].

Per nomenar les inteïnes de cada espècie i de cadascun dels gens on es troben es va consensuar un sistema, en el qual s'abreua del nom de l'espècie o gènere que la conté seguit pel nom del gen. Per exemple, el gen PolB (DNA polimerasa tipus B) de *Nanoarchaeum equitans*, es troba envaït per la inteïna Neq DnaPolB. En el cas, de què la mateixa exteïna es trobi envaïda per diferents inteïnes, aquestes es diferencien a partir de valors numèrics. Un bon exemple és el gen Pol-II (DNA Polimerasa-II) d'*Haloquadratum walsbyi* el qual es troba envaït per dues inteïnes, Hwa Pol-II-1 i Hwa Pol-II-2.

D'altra banda, i amb l'objectiu de facilitar els estudis bioquímics comparatius entre diferents inteïnes, es va crear una nomenclatura. Aquesta nomenclatura es basa en la posició de l'aminoàcid en la seqüència del precursor. En el cas de la inteïna el primer residu de l'extrem N-terminal és el residu 1 i la numeració va creixent de manera seqüencial fins al final de la seqüència de la inteïna. En l'exteïna N el residu adjacent a la inteïna, és el residu -1. El residu +1 és el primer residu de l'exteïna C, el següent residu és el +2 i la numeració després continua com està indicat en la Figura 1.1.



Figura 1.1. Precursor en cis format per la inteïna unida per l'extrem N- i C- terminal a les dues exteïnes.

Totes les inteïnes que es van descobrir entre 1990 i 1997 contenen un domini endonucleasa (HEN) en la seva seqüència. Aquest domini es troba al mig del domini HINT (Figura 1.2-a) i és el responsable de la mobilitat genètica de les inteïnes [18].

No obstant això, s'han descobert altres inteïnes més petites, les quals reben el nom de mini-inteïnes. La primera a ser descoberta va ser la inteïna que envaïx el gen codificant per la subunitat A de la DNA girasa de *Mycobacterium xenopi* (Mxe GyrA) [29]. La principal característica d'aquesta classe d'inteïnes és que perden el domini HEN, i per tant, la capacitat d'envair nous gens. Òbviament, la pèrdua d'aquest domini comporta un escurçament de la seqüència (Figura 1.2-b).

Aquests dos tipus d'inteïnes es coneixen com a inteïnes en cis perquè tant la inteïna com les dues exteïnes formen part del mateix precursor polipeptídic i estan codificades en un únic gen.

Actualment, també es coneixen les inteïnes partides o inteïnes *trans*, les quals no tenen les regions N i C de la inteïna unides covalentment. En les inteïnes partides la inteïna N (IntN) juntament amb l'exteïna N (ExtN), i la inteïna C (IntC) amb l'exteïna C (ExtC), formen dues cadenes polipeptídiques separades que es troben codificades en dos gens diferents (Figura 1.2-c). En aquest tipus d'inteïnes la reacció de *splicing* (*trans-splicing*, en aquest cas) només es pot donar després de l'associació específica dels dos fragments de la inteïna, ja que d'aquesta manera es reconstrueix el centre actiu de la inteïna.

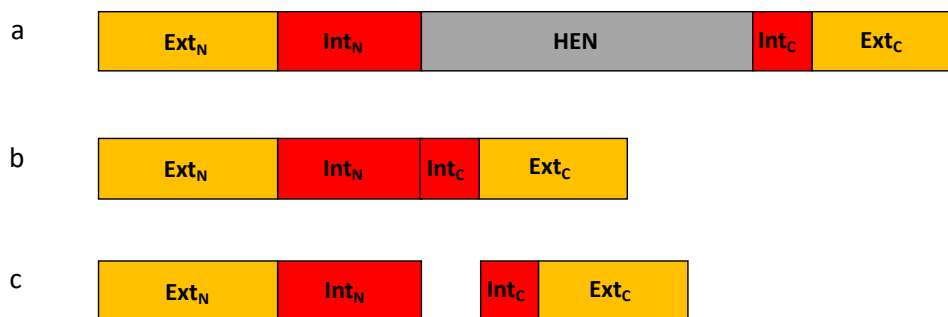


Figura 1.2. Esquema dels diferents tipus d'inteïnes trobades a la natura. (a i b) inteïnes en *cis*; (c) inteïna partida o *trans*.

Les inteïnes comparteixen un baix grau d'homologia entre elles, però nombrosos estudis realitzats han identificat blocs de residus conservats, els quals tenen un paper important en la reacció de *splicing*. Aquests blocs són els blocs A, B, F i G [27], esquematitzats en la Figura 1.3.

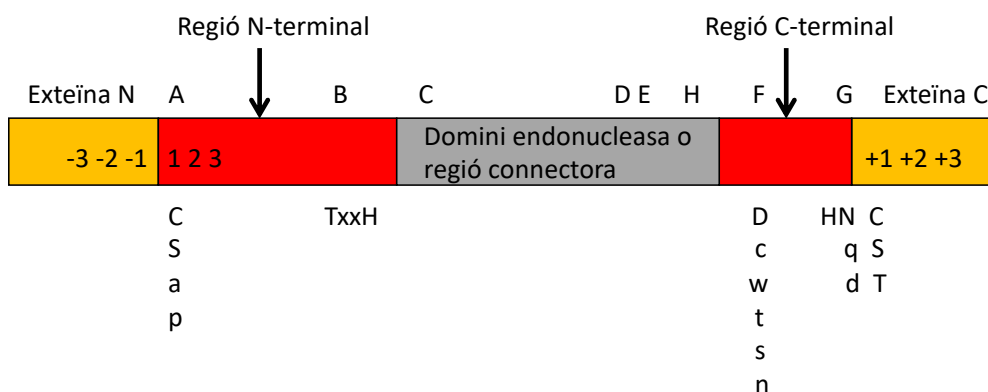


Figura 1.3. Estructura esquemàtica dels precursors que catalitzen la reacció de *splicing*. La inteïna (vermell) interrompuda pel domini endonucleasa o regió connectora (gris) formen el domini de *splicing*, el qual es troba unit a través d'un enllaç peptídic a les exteïnes N i C (groc). Els residus que conformen els llocs d'unió entre l'exteïna N i l'exteïna C s'indiquen seguint la numeració descrita en el text anteriorment. La localització dels blocs conservats es troba indicada en la part superior de l'esquema. Els residus més conservats es troben indicats en majúscula (Cys i Ser, en el bloc A; Thr i His en el bloc B; Asp en el bloc F; His i Asn en el bloc G; Cys, Ser i Thr en la posició +1) mentre que els menys conservats es troben indicats en minúscula (Ala i Pro en el bloc A; Cys, Trp, Thr, Ser i Asn en el bloc F; Gln i Asp en el bloc G).

Bloc A: Defineix el lloc d'unió N-terminal i està constituït per 13 residus. El primer residu en la majoria d'inteiïnes és una Cys o una Ser, en alguns casos poc freqüents s'hi troba un residu de Ala o Pro. Aquest residu intervé en el primer pas de la reacció de *splicing* formant l'intermediari tioèster o èster. La mutació Cys1Ala o Ser1Ala és una manera eficaç de bloquejar la reacció de *splicing*, tot i que, com es veu més endavant, hi ha inteïnes que han desenvolupat mecanismes que resolen aquest problema.

Bloc B: Format per 14 residus, on els residus 7 i 10 es troben altament conservats formant la seqüència ThrxxHis (x, representa qualsevol aminoàcid). Diferents estudis estructurals i mutacionals suggereixen que aquests residus juguen un paper clau en les primeres etapes de la reacció de *splicing* [30].

Bloc F: Constituït per 16 residus, la meitat dels quals es troben conservats. Normalment, el primer residu és un residu aromàtic (Phe, Tyr o Trp) seguit pels residus Val, Tyr, Asp, Leu, un residu polar (Cys, Ser o Thr) i una Val. Tot i que s'hi poden trobar altres residus, l'Asp en la posició 4 es troba fortament conservat. Es pensa que l'Asp en aquesta posició actua com a catalitzador i coordinador de la reacció de *splicing* [31].

Bloc G: Format per una curta seqüència de 8 residus que acaba en la majoria de casos amb una His en la penúltima posició, seguida d'una Asn altament conservada. En algunes inteïnes l'últim residu del bloc F és una Gln o Asp. Es pensa que la penúltima His d'aquest bloc juga un paper clau en la ciclació de la cadena lateral de l'Asn en l'últim pas de la reacció de *splicing*, tot i que hi ha inteïnes que han perdut aquest residu, com la Npu DnaE.

Tot i que la inteïna conté tota la informació necessària per dur a terme la reacció de *splicing*, s'ha vist que l'exteïna pot modular la seva activitat. Diversos estudis han demostrat que els residus +1, +2 i +3 de l'exteïna C modulen la cinètica de la reacció de *splicing* [32]–[36]. En alguns casos, la velocitat d'aquesta reacció també es veu afectada pel residu -1 de l'exteïna N [34], [36], [37].

La tolerància a modificacions en els residus que es troben en les regions d'unió entre la inteïna i l'exteïna és característica de cada inteïna. Per exemple, la inteïna partida Npu DnaE i la i Ssp DnaE presenten un 63% d'homologia de seqüència, permeten que es pugui donar la reacció de *splicing* de manera creuada entre les dues inteïnes. Les dues presenten els mateixos residus en les posicions +1, +2 i +3 de l'exteïna C (Cys-Phe-Asn) i en les posicions -2 i -1 de l'exteïna N (Gly-Tyr). Tot i això, presenten grans diferències sobre la tolerància dels residus de l'exteïna. La inteïna Ssp DnaE presenta una baixa tolerància a canvis en la posició +2 de l'exteïna C, acceptant tan sols el canvi per Tyr i Trp. Per altra banda, aquesta inteïna també és sensible als canvis en la posició -1 de l'exteïna N, la mutació Tyr-1Asp augmenta significativament la velocitat d'hidròlisi

de l'extrem N-terminal, en canvi, la mutació per Pro o Val la bloqueja [34]. En canvi, la Npu DnaE presenta una elevada tolerància a la substitució de residus en aquestes posicions [33].

1.1.2. Inteïnes com a interruptors ambientals: Conditional Protein *Splicing* (CPS)

Com s'ha comentat anteriorment, en els últims anys s'han descobert diferents inteïnes que actuarien com a reguladors post-traduccionals de l'activitat de la proteïna hoste en funció de condicions cel·lulars o ambientals, tals com la temperatura, la concentració de sals... [19]–[26]. De manera que en absència de l'estímul correcte la inteïna quedaria atrapada en el precursor proteic, i com a conseqüència la proteïna hoste es mantindria inactiva, i només davant de l'estímul adequat es promouria la reacció de *splicing* (Figura 1.4). Per aquesta raó, en aquests casos es parla de les inteïnes com a interruptors ambientals.

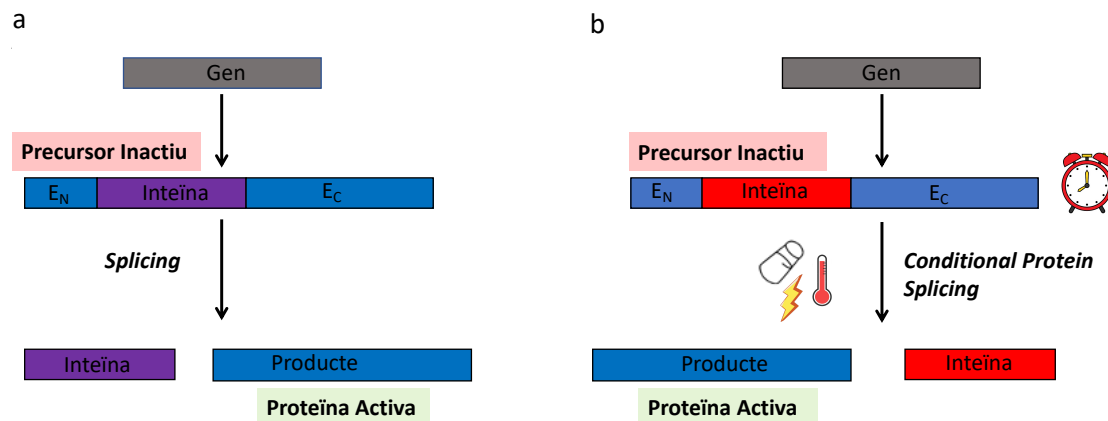


Figura 1.4. Esquema comparatiu de la reacció de *splicing* i la reacció de Conditional Protein *Splicing* (CPS). (a) La reacció de *splicing* es dona de manera simultània a mesura que s'expressa el precursor inactiu donant lloc a la proteïna activa; (b) en canvi la reacció CPS només es produirà davant de l'estímul específic, en absència d'aquest, la proteïna quedarà atrapada en el precursor inactiu.

Les inteïnes dependents d'estímul s'han trobat majoritàriament en organismes extremòfils envaint proteïnes involucrades amb processos de replicació, recombinació i reparació del DNA. Davant d'un estrès qualsevol, generalment, els organismes responen inhibint els processos relacionats amb el metabolisme del DNA, d'aquesta manera eviten la toxicitat per la replicació del DNA i estalvien ATP necessari per a mantenir els processos bioquímics per a poder sobreviure a l'estrès. Aquestes inteïnes bloquejarien la reacció de *splicing* i preservarien la integritat cel·lular davant d'un estrès determinat recuperant la seva activitat després de què aquest passes. En aquest cas, les inteïnes representarien un benefici evolutiu per a l'hoste i explicaria la marcada presència d'aquestes en el domini Archaea.

Existeixen inteïnes naturals que modulen la seva activitat en funció del potencial redox. Per exemple la inteïna de *P. abyssi*, la inteïna PolII, la qual forma un pont disulfur entre la Cys1 i la Cys+1 que bloqueja la reacció de *splicing* quan s'expressa en *E. coli* (BL21) [20]. Un altre exemple és la inteïna Mma PolII de l'arqueobacteri metanogènic *M. marisnigri*, la qual pot formar un pont disulfur intern que modula l'activitat de la inteïna en funció del potencial redox d'*E. coli* [21].

La reacció de *splicing* també pot ser regulada per la temperatura. Aquesta regulació pot tenir una gran rellevància a nivell fisiològic, ja que s'ha demostrat que l'activitat de diverses inteïnes d'organismes hipertermòfils està regulada per la temperatura. Per exemple, la inteïna hipertermofílica de *P. abyssi*, Pab PolII [19]. *P. abyssi*, és un arqueobacteri hipertermòfil anaeròbic que creix a temperatures properes a l'ebullició de l'aigua (temperatura òptima de creixement de 96°C). Du i col·laboradors han demostrat mitjançant ressonància magnètica nuclear (RMN) que aquesta inteïna, a més d'estar regulada pel potencial redox com s'ha comentat anteriorment, també es troba regulada per la temperatura a nivell estructural [19]. Pab PolII a temperatures mesofíliques, 25°C, presenta una elevada rigidesa estructural, la qual bloqueja la reacció de *splicing* a aquestes temperatures, a l'augmentar la temperatura aquesta rigidesa es perd, i com a conseqüència es recupera l'activitat catalítica de la inteïna. Un altre exemple de *splicing* regulat per la temperatura és el cas de les inteïnes de *P. Horikoshii* i *T. Sibericus*, Pho RadA i Tsi RadA, respectivament. Diferents treballs experimentals han demostrat que el precursor format per les exteïnes natives només pot catalitzar la reacció de *splicing* quan s'incuba a elevades temperatures. Aquesta regulació per temperatura es perd quan la inteïna s'expressa amb exteïnes no natives, la qual cosa indica que la regulació de la reacció de *splicing* en aquest cas ve donada per la interacció exteïna-inteïna.

També s'han identificat inteïnes que regulen la seva activitat en funció de la concentració de sals en el medi, com ara, la inteïna de l'arqueobacteri halòfil *H. Salinarum* sp NCRI, Hsa Pol-II, la qual només catalitza la reacció de *splicing* a concentracions de sal elevades [38].

1.1.3. Mecanisme de *splicing*

El mecanisme de *splicing* es troba exhaustivament estudiat i documentant [15], [28], [39], [40]. El mecanisme pel qual la inteïna s'escindeix del seu precursor consta de quatre passos (Figura 1.5):

- 1- Desplaçament N-S/O acil, dirigit pel residu catalític 1 de la inteïna (Cys/Ser). En aquest pas el sofre de la Cys o l'oxigen de la Ser ataquen nucleofílicament el grup carbonil de l'enllaç peptídic que uneix la inteïna amb l'exteïna N, com a conseqüència es forma un

- enllaç tioèster o èster. La molècula resultant d'aquest primer pas rep el nom d'intermediari lineal.
- 2- Transesterificació. El sofre de la Cys o l'oxigen de la Ser o Thr en posició +1 de l'exteïna C ataquen nucleofilicament l'enllaç tioèster/èster format en el pas 1 de la reacció. El producte d'aquest pas s'anomena intermediari ramificat.
 - 3- Ciclació de l'Asn del bloc G de la inteïna. Aquest pas és irreversible i implica l'escissió de la inteïna i l'obtenció de l'exteïna unida a través d'un enllaç (tio)èster.
 - 4- Reorganització de l'exteïna. En aquest pas es forma entre les dues exteïnes un enllaç amida donant lloc a la proteïna madura. Aquesta etapa final és favorable des d'un punt de vista termodinàmic i no necessita l'ajuda de la inteïna.

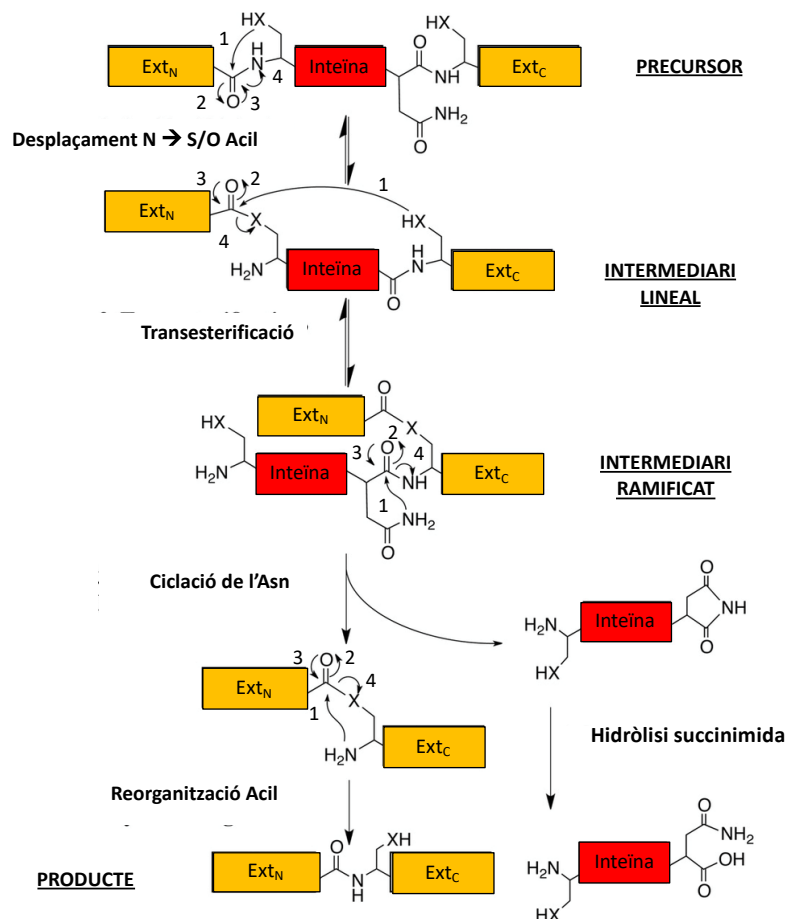


Figura 1.5. *Mecanisme de splicing. Mecanisme estàndard de la reacció de splicing catalitzada per inteïnes. X pot representar l'oxigen o el sofre de les cadenes laterals dels residus Ser, Thr o Cys. L'ordre en què els electrons salten en els diferents passos de la reacció de splicing estan numerats.*

Desplaçament N-S/O acil

Aquest pas des d'un punt de vista termodinàmic és desfavorable, ja que s'hidrolitza un enllaç peptídic per formar un enllaç d'elevada energia com és l'enllaç (tio)èster. La desestabilització de l'enllaç peptídic es considera la millor explicació per a la formació de l'intermediari lineal. En diverses estructures cristal·logràfiques s'ha trobat que l'enllaç peptídic -1 presenta una distorsió en la seva planaritat, la qual facilita l'atac nucleofílic per part de la Cys/Ser. En el cas de la inteïna Mxe GyrA s'ha vist que l'enllaç adopta una inusual configuració en *cis*, en el mini precursor Sce VMA aquest adopta una configuració distorsionada en *trans* i en el cas de la la inteïna Ssp DnaE aquest adopta una configuració normal en *trans* amb cert grau de tensió [41]–[44]. Altrament, estudis de RMN realitzats amb el precursor Mxe GyrA marcant exclusivament l'enllaç peptídic -1 amb els isòtops ^{13}C i ^{15}N , mostren que l'enllaç amida presenta una naturalesa altament polar, a conseqüència de la interacció directa amb el residu conservat His del bloc B mitjançant un pont d'hidrogen [45].

La distorsió en la planaritat de l'enllaç peptídic entre l'exteïna N i la inteïna no és suficient perquè es formi l'intermediari lineal. Per a la formació del nou enllaç (tio)èster es necessita: 1- la desprotonació del grup tiol o alcohol dels residus de Cys o Ser, respectivament; 2- l'estabilització de l'intermediari tetraèdric, i 3- la protonació del grup amida sortint. Utilitzant RMN, Wang i col·laboradors [30], han proposat un mecanisme en el qual la His10 del bloc B a través de canvis de pKa, té un paper essencial en la formació de l'intermediari lineal. En primer moment aquesta His actuaria com a base (pKa \sim 7,3), desprotonant el grup nucleofílic (hidroxil o tiol) facilitant l'atac nucleofílic i començant d'aquesta manera la reacció de *splicing*. A continuació, aquest mateix residu actuaria com a àcid (pKa < 3.5) protonant el grup amida sortint. En aquest mecanisme l'estabilització de l'intermediari tetraèdric es produiria gràcies a altres residus conservats del mateix bloc B. En el cas de la Mxe GyrA, l'estabilització es dona gràcies a l'Asn76 i His79, i en la inteïna Sce VMA per l'Asn76 i l'His79.

En algunes inteïnes l'Asp altament conservat del bloc F es troba interaccionant amb l'extrem N-terminal de la inteïna on es troba el residu catalític [31], [46], [47], el que suggereix que aquest residu també intervé en la formació de l'intermediari lineal. De fet es creu que també aquest Asp estabilitza l'intermediari tetraèdric.

Transesterificació

El segon pas en la reacció de *splicing* de les inteïnes consisteix en la trans-(tio)esterificació de l'exteïna N. En aquest pas l'exteïna N passa d'estar unida del primer residu catalític de la inteïna

al primer residu de l'exteïna C, mitjançant un enllaç (tio)-èster, formant-se d'aquesta manera l'intermediari ramificat. Aquest pas es troba poc caracteritzat, principalment per les dificultats a l'hora d'aïllar l'intermediari lineal. Generalment, s'assumeix que els requeriments perquè es realitzi aquest pas són els mateixos que per a la formació de l'intermediari lineal, la desprotonització de la cadena lateral del residu nucleòfil +1, el correcte posicionament per a l'atac de l'enllaç (tio)èster present en l'intermediari lineal, i l'estabilització de l'intermediari tetraèdric [48], [49]. Actualment, es desconeix com es posen en contacte el nucleòfil i l'enllaç (tio)èster de l'intermediari lineal, ja que en la majoria d'estructures d'intèines dipositades en el PDB existeix una gran distància, 8-9 Å, entre el primer residu de l'exteïna C i el primer residu de la inteïna [41], [43], [47]. Treballs recents [50]–[52] suggereixen que l'Asp del bloc F juga un paper clau en la transesterificació de les inteïnes. D'aquesta manera semblaria que els dos primers passos de la reacció de *splicing* estiguessin connectats a través d'una complexa xarxa de ponts d'hidrogen basada en l'acoblament de pKa que presenten el residu nucleòfil 1 del bloc A (Cys o Ser), la His del bloc B, i l'Asp del bloc F [52]. S'ha vist que en la inteïna Mtu RecA, quan l'Asp del bloc F es troba protonat, aquest presenta un pKa inusual de 6.1 i podria estabilitzar el nucleòfil en posició 1. Després de la formació de l'intermediari lineal el pKa d'aquest residu disminueix, per tant actuaria com a àcid desprotonant el residu catalític +1, i promovent d'aquesta manera la reacció de transesterificació [52].

Ciclació de l'Asparagina

La formació de l'intermediari ramificat implica l'escissió irreversible de la inteïna del precursor. En aquest pas, el nitrogen de la cadena lateral de l'Asn del bloc G, o amb menor freqüència la Gln, catalitza l'atac nucleòfil sobre el carboni carbonil de l'enllaç peptídic +1, formant-se succinimida. En la inteïna Mxe GyrA la formació de l'intermediari ramificat estimula la ciclació de l'Asn, on és la velocitat limitant de la reacció de *splicing* [50]. L'estructura de l'intermediari ramificat de la inteïna Mxe GyrA obtingut de manera semisintètica mostra que l'estructura i la dinàmica del centre actiu ha estat alterat per l'establiment d'interaccions que posicionen correctament la cadena lateral de l'Asn del bloc G per a l'atac nucleòfil [53]. Altrament, estudis computacionals i estructurals indiquen que les His dels bloc B i G són importants en aquest pas [50], tot i que algunes inteïnes han perdut la His del bloc G [54].

El mecanisme de *splicing* es pot veure afectat per diferents factors, com ara: mutacions dels residus claus, la presència d'agents nucleofílics o la utilització d'exteïnes no natives. Com a conseqüència, l'eficiència del procés disminueix produint-se reaccions secundàries no desitjades tals com la hidròlisi de l'extrem N- i/o C-terminal [34], [55]–[58]. La hidròlisi de l'extrem N-

terminal consisteix en la hidròlisi de l'enllaç (tio)èster format entre l'exteïna N i el nucleòfil 1 o +1 en els intermediaris lineal o ramificat. Com a producte s'obté l'exteïna N i la inteïna amb l'exteïna C (Figura 1.6). En el cas de la hidròlisi de l'extrem C-terminal aquest es produeix a conseqüència d'una ciclació prematura de l'Asn del bloc G, o de la Gln en algunes inteïnes. Com a resultat d'aquesta reacció, s'obté per una banda l'exteïna N unida a la inteïna i l'exteïna C lliure (Figura 1.6). Un tret important a tenir en compte és que les reaccions d'hidròlisi de l'extrem N- i C-terminal no estan coordinades. Les mutacions Cys1Ala o Ser1Ala són una manera eficient de bloquejar la primera etapa de la reacció, i per tant, la hidròlisi de l'extrem N-terminal, mentre que la hidròlisi de l'extrem C-terminal pot produir-se igualment [19], [31], [43]. Al contrari, passa amb les mutacions que prevenen la ciclació de l'Asn, com ara la substitució de l'Asn per Ala. És a dir, el bloqueig de la ciclació de l'Asn no impedeix que es formi l'intermediari lineal, el qual és susceptible a la hidròlisi [59], [60].

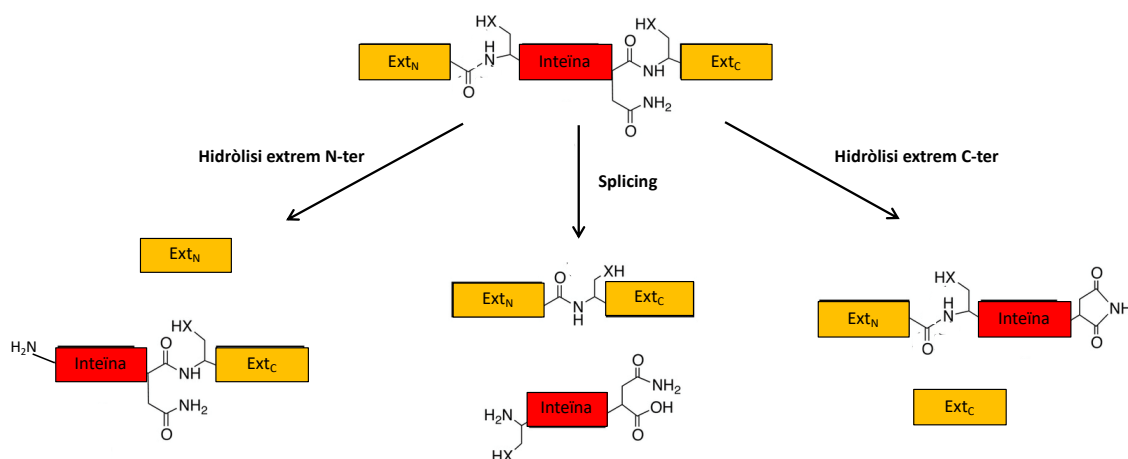


Figura 1.6. Reaccions d'hidròlisi dels extrems terminals. El mecanisme de *splicing* es pot veure afectat per diversos factors donant lloc a reaccions no desitjades com la hidròlisi de l'extrem N- i C-terminal.

1.1.4. Classes d'inteïnes

La majoria de les inteïnes descobertes fins ara presenten l'estructura en blocs descrita anteriorment (Bloc A, B, F i G). Aquestes inteïnes pertanyen a la classe I d'inteïnes. S'han trobat altres inteïnes amb variacions en l'estructura clàssica, les quals no utilitzen Cys o Ser en la posició 1 però sí en altres localitzacions. Aquestes inteïnes es divideixen en les classes 2 i 3 en funció del mecanisme de *splicing* que presenten [61].

Classe 2: Es caracteritza per tenir una Ala en posició 1. En la classe 1 d'inteïnes la presència d'Ala en la posició 1 bloqueja totalment la reacció de *splicing*. Per tant, les inteïnes de classe 2 han adaptat el centre catalític a aquesta pèrdua del residu nucleòfil i el que fan és directament formar l'intermediari ramificat, és a dir, el nucleòfil +1 ataca directament a l'enllaç peptídic entre la inteïna i l'exteïna N [46].

Classe 3: En aquest cas les inteïnes tampoc tenen el residu nucleofílic en la posició 1, si no que el tenen en una altra localització. La Cys encarregada de formar l'intermediari lineal es troba en el bloc F en el triplet conservat Trp-Cys-Thr [62]. Estudis mutacionals revelen que els residus Trp i Thr no participen en el mecanisme de *splicing*.

A la Figura 1.7 s'esquematitza el mecanisme de *splicing* dut a terme per les tres classes d'inteïnes.

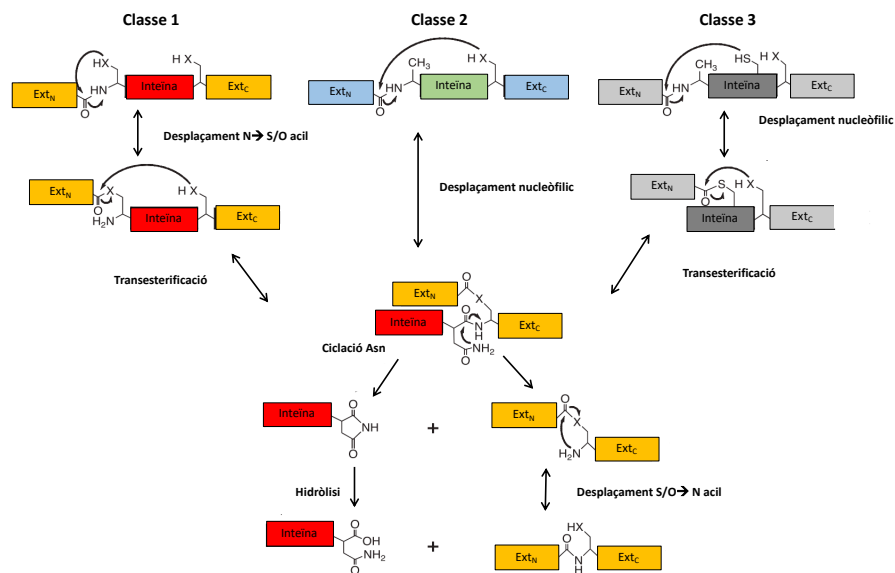


Figura 1.7. Mecanisme de *splicing* de les tres classes d'inteïnes. La majoria de les inteïnes descrites fins ara corresponen a la classe 1 d'inteïnes, la qual catalitza la reacció de *splicing* descrita prèviament. La classe d'inteïnes 2 i 3 catalitzen la reacció de *splicing* seguint altres vies, pel fet que generalment l'aminoàcid en posició 1 no és capaç de formar l'intermediari lineal. X representa l'àtom de sofre o oxigen de la cadena lateral de la Ser, Thr o Cys.

1.1.5. Distribució dels aminoàcids nucleòfils en les posicions 1 i +1

Liu i col·laboradors van analitzar la distribució dels aminoàcids nucleòfils en les posicions 1 i +1 i la seva combinació en més de 500 inteïnes reportades en el InBase (<http://www.neb.com/neb/inteins.html>) [63]. Com a resultat de l'estudi, van obtenir que el 77% de les inteïnes presenten en la posició 1 una Cys, el 10% una Ser i la resta, el 12%, tenen residus no catalítics, com ara l'Ala. Amb referència a la posició +1 van trobar que els residus Cys, Ser i

Thr s'hi troben en un 40%, 40% i 20%, respectivament. Quan van mirar la combinació dels dos residus 1/+1 van observar la següent distribució: la combinació Cys/Cys s'hi troba amb un 36%, sent la més abundant, seguida per la combinació Cys/Ser amb un 29%, la combinació Cys/Thr amb un 12%, i les combinacions Ser/Ser i Ser/Thr les dues amb un percentatge igual al 5% (Figura 1.8-a). Curiosament, la combinació Ser/Cys no es troba en cap inteïna natural. Aquesta distribució varia si es té en compte el grup taxonòmic (Figura 1.8-b).

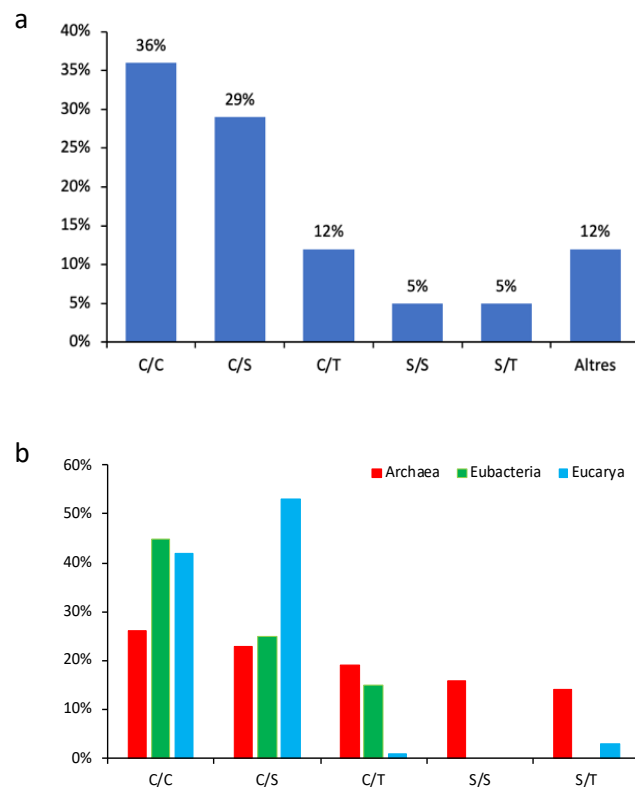


Figura 1.8. Distribució dels diferents nucleòfils en les posició 1 i +1. (a) Distribució per a totes les inteïnes, en l'eix de les X les diferents combinacions de residus en les dues posicions amb representació en les inteïnes, en l'eix de les Y el percentatge de cada combinació. (b) El mateix que abans, excepte que les inteïnes estan dividides en funció dels tres dominis de la vida, Archaea, Eubacteria i Eucarya.

Aquesta anàlisi mostra que el residu més freqüent en la posició 1 és la Cys (77%) seguit per la Ser (10%). En canvi en la posició +1 es troben en la mateixa freqüència tots dos residus, Cys i Ser (40%). Actualment, no s'ha descrit cap inteïna que presenti una Thr en posició 1, i només un 20% de les inteïnes presenten aquest residu en posició +1. Aquesta diferència respecte a la distribució dels residus en posició 1 i +1, indica que la posició 1 és més restrictiva a l'hora d'acomodar altres residus que no sigui la Cys en la posició +1. L'explicació per la qual la majoria d'inteïnes presenten una Cys en la posició 1 probablement rau en les propietats nucleòfiles d'aquesta que permeten pugui ser reactiva sense la necessitat d'altres residus assistents com

passa amb el cas dels residus de Ser i Thr [64],[65], i en la termodinàmica del primer pas de la reacció de *splicing* (desplaçament N-S acil) que és, com s'ha comentat anteriorment, altament desfavorable. La Thr igual que la Ser presenta un grup hidroxil en la seva cadena lateral, en canvi, la seva freqüència és molt més baixa. Aquest fet es pot explicar per què la Thr en la seva cadena lateral presenta un grup metil, el qual al tractar-se d'un grup apolar pot influir en les propietats del solvent del seu voltant atenuant d'aquesta manera la reacció de *splicing* [58].

1.1.6. *Trans-splicing*

Com s'ha comentat anteriorment, algunes inteïnes es troben de manera natural partides en dos fragments codificats en dos gens diferents, els quals es transcriuen i s'expressen de manera independent donant lloc al precursor N, format per l'exteïna N unida a la inteïna N, i al precursor C, constituït per la inteïna C fusionada amb l'exteïna C. Aquest tipus d'inteïnes només representen un 5% del total de les inteïnes descrites actualment [7]. Després de la traducció dels dos gens, les dues inteïnes interaccionen de manera no covalent i espontània per a reconstituir l'estructura característica del domini HINT.

Les seqüències d'inteïnes partides descrites fins ara són divergents i envaeixen almenys cinc proteïnes hoste diferents, el que suggereix que no presenten un origen evolutiu comú [66]–[68]. El lloc pel qual la majoria d'inteïnes es troben partides és el lloc d'inserció del domini HEN, entre el bloc B i F en moltes inteïnes contigües o *cis*. El mecanisme pel qual aquestes inteïnes presenten una estructura fragmentada en dues parts roman incert actualment, tot i que es pensa que el gen va ser fracturat per la inserció errònia i la posterior pèrdua d'un domini endonucleasa en una inteïna. La pèrdua d'aquest domini aniria acompanyada de la deslocalització de les dues parts del gen fragmentat en dos locis separats (Figura 1.9) [66].

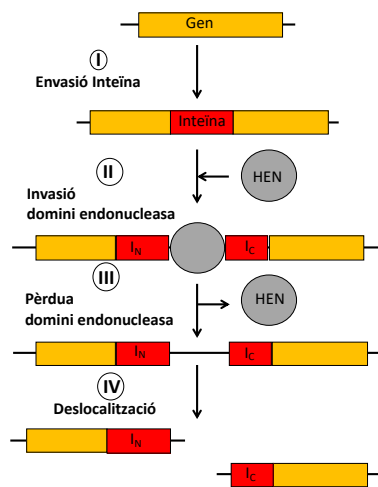


Figura 1.9. Model proposat per la formació d'inteïnes partides. On inicialment un gen és envaït per una inteïna (I). A continuació, el gen és fracturat per la invasió d'un domini endonucleasa (HEN) (II). Aquest domini endonucleasa en les inteïnes partides es perd (III), el que pot anar seguit d'una deslocalització de les dues parts del gen fracturat en dos locis diferents (IV).

L'organització estructural de les inteïnes partides és la mateixa que per a les inteïnes *cis* i conserven els quatre blocs seqüencials nomenats anteriorment, però en aquest cas es troben repartits entre la inteïna N i la inteïna C. Generalment, de manera natural la inteïna N sol tenir una longitud al voltant dels 120 residus i la inteïna C presenta una mitjana de 36 residus.

La família més gran d'inteïnes partides caracteritzada avui en dia es troba envaït el gen DnaE en almenys 20 espècies de cianobacteri [67]. Dintre d'aquesta família es troba la inteïna partida millor caracteritzada actualment, la inteïna partida de *Nostoc punctiforme* (Npu).

Avui dia, dipositades en el Protein Data Bank (<https://www.rcsb.org>) existeixen tan sols 9 estructures d'inteïnes partides, totes elles corresponen a construccions de la família DnaE de cianobacteris, concretament a les espècies *Synechocystis sp.* PCC6803 (Ssp) [47], [69], [70] i *Nostoc punctiforme* (Npu) [71], [72]. Cap d'elles correspon a la inteïna N o inteïna C aïllada i només dues de les estructures corresponen al complex IntN/IntC (codi PDB:4lx3 i 4qfq) amb les unitats unides de manera no artificial [73]. Dit d'una altra manera, en set de les nou estructures d'inteïnes dipositades en el PDB la inteïna N i la inteïna C es troben fusionades, expressant-se de manera recombinant a partir d'un únic gen, per a generar d'aquesta manera una inteïna contigua. En qualsevol cas, aquests estudis verifiquen que les inteïnes partides adopten el plegament en forma de ferradura característic del domini HINT.

El primer pas perquè es doni la reacció de *trans-splicing* és que els dos fragments de la inteïna partida interaccionin de manera específica per formar el complex IntN/IntC (Figura 1.10). Perquè la reacció de *trans-splicing* sigui eficient els dos fragments de la inteïna han d'interaccionar de

manera ràpida i amb una forta afinitat. L'alineament seqüencial de les inteïnes partides demostra que existeix una forta segregació de càrregues, amb els residus àcids concentrats en la inteïna N i els residus bàsics conservats en la inteïna C i que aquesta es troba absent en les inteïnes contigües o *cis*-inteïnes [74], [75]. Muir i col·laboradors han demostrat que aquesta segregació de càrregues és essencial per a la interacció dels dos fragments i han identificat quins són els residus implicats en les interaccions iòniques específiques que s'estableixen entre els dos fragments per a formar el complex [75].

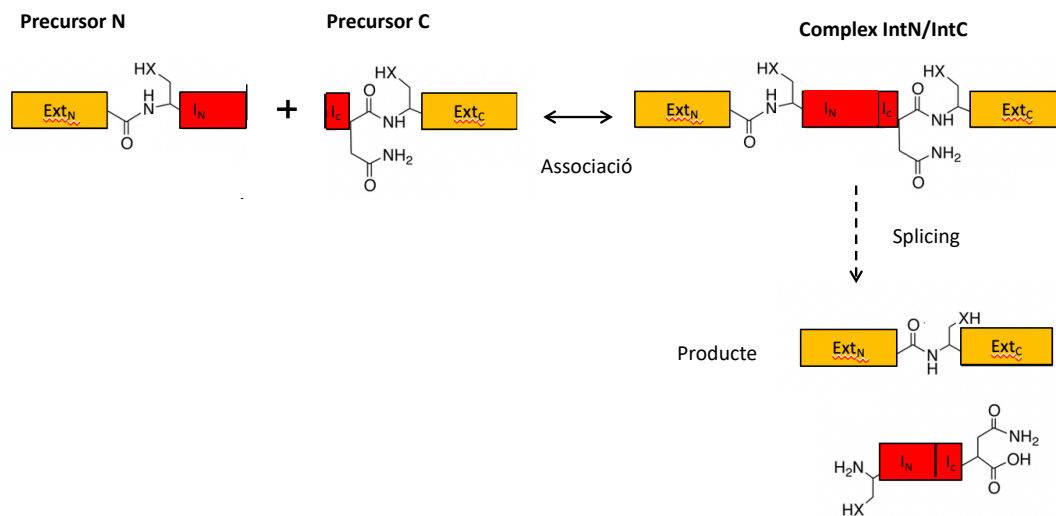


Figura 1.10. Esquema del mecanisme de *trans*-splicing.

Estudis cinètics i termodinàmics sobre el procés d'associació de les inteïnes partides de la família DnaE mostren que la interacció entre els dos fragments és ràpida i amb una forta afinitat. Per exemple estudis de fluorescència realitzats amb la inteïna partida Ssp DnaE presenten una constant d'afinitat de l'ordre nanomolar ($K_D \sim 30$ nM) i una constant de velocitat, k_{on} , de $\sim 10^7$ $M^{-1} s^{-1}$ [76]. De la mateixa manera, els fragments de la inteïna Npu DnaE també interaccionen fortament ($K_D \sim 2$ nM) [75] i amb una baixa cinètica d'associació més ràpida (k_{on} de $\sim 10^5$ - 10^6 $M^{-1} s^{-1}$) [35].

Per a entendre com dues proteïnes aïllades poden interaccionar i formar un complex amb l'estructura entrellaçada típica de les inteïnes, s'han fet diverses anàlisis estructurals i biofísiques utilitzant les inteïnes partides Ssp DnaE i Npu DnaE. Tots dos estudis convergeixen en què els dos fragments de la inteïna de manera aïllada no presenten una estructura tridimensional ben definida, la qual assolixen només després de la seva associació. Per tant, la inteïna N i la inteïna C es poden considerar proteïnes intrínsecament desestructurades (IDPs).

Liu col·laboradors utilitzant estudis biofísics i termodinàmics proposen que la inteïna N i la inteïna C de Ssp DnaE presenten un estat desplegat quan es troben aïllades, el qual canvia després de la interacció dels dos fragments assolint l'estructura típica en ferradura de les inteïnes [77].

D'altra banda Muir i col·laboradors utilitzant tècniques biofísiques, han demostrat que l'associació dels dos fragments de la Npu és un procés compost per diferents passos. Aquest model es basa en el fet que la inteïna N presenta una estructura dividida en dos lòbuls, els quals presenten diferents propietats biofísiques. El primer lòbul, NpuN1, format pels residus 1-50 presenta característiques de glòbul fos (elevada hidrofobicitat i baixa càrrega neta), i el segon lòbul, NpuN2, constituït pels residus 51-102 amb propietats característiques de proteïnes intrínsecament desestructurades (IDPs), com són una elevada càrrega neta i una baixa hidrofobicitat.

En primer lloc, en el mecanisme proposat per l'associació de la Npu DnaE, la NpuC seria capturada per la NpuN2 mitjançant interaccions electroestàtiques entre els residus bàsics de la inteïna C i els residus àcids del segon lòbul de la inteïna N, formant l'intermediari de plegament. Després d'aquesta captura inicial les regions flexibles col·lapsarien en un estat ordenat el qual estaria estabilitzat per interaccions hidrofòbiques entre la NpuC i el primer lòbul de la NpuN. Com a resultat s'obtindria l'estructura entrellaçada típica de les inteïnes [57] (Figura 1.11).

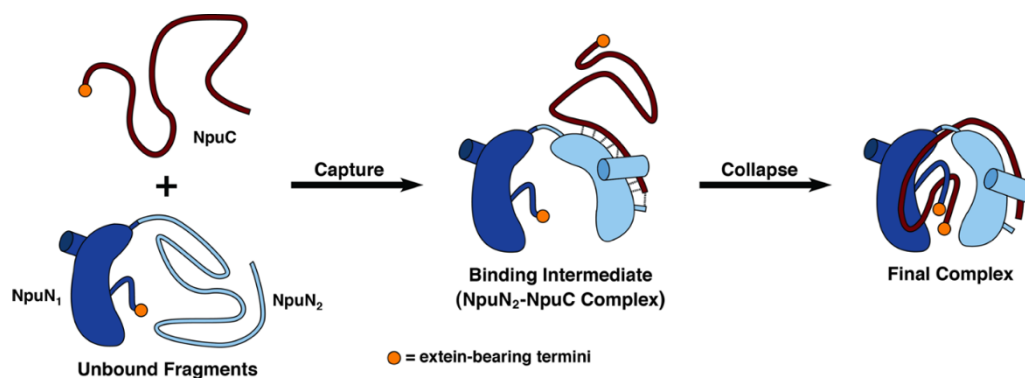


Figura 1.11. Esquema del mecanisme de captura i col·lapse per l'associació dels dos fragments de la inteïna partida Npu DnaE proposat per Muir i col·laboradors. La NpuC és capturada pel lòbul NpuN2 mitjançant interaccions electroestàtiques. Després aquest col·lapsaria donant lloc a una estructura ben definida, la qual estaria estabilitzada per interaccions hidrofòbiques entre la NpuC i el lòbul NpuN1. Esquema reproduït de Shah et al., 2013 [57].

1.1.7. Aplicacions

L'interès en l'estudi de les inteïnes i el seu mecanisme no es troba només en el coneixement en si, si no en les nombroses aplicacions que se'n deriven.

Marcatge isotòpic de proteïnes multidomini

Actualment, una de les principals tècniques per a resoldre l'estructura tridimensional de les proteïnes és la ressonància magnètica nuclear (RMN). Aquesta tècnica presenta una limitació i és que no es pot resoldre l'estructura tridimensional de proteïnes més grans de 100 kDa [78]. Cal dir que un dels problemes més grans d'aquesta tècnica és la dificultat d'interpretar espectres obtinguts de proteïnes grans per la gran quantitat de senyals que generen. El marcatge isotòpic diferencial ofereix un mitjà per estudiar de manera específica la regió d'interès en proteïnes grans en un context nadiu mitjançant RMN [79]–[81]. La reacció de *trans-splicing* catalitzada per inteïnes partides ofereix un mètode per al marcatge isotòpic diferencial [79]–[81]. Aquesta tècnica es pot dur a terme *in vitro* i *in vivo*. *In vitro* els dos precursors, precursor N i C, s'expressen de manera individual, i després en un tub d'assaig es realitza la reacció de *trans-splicing*. En aquest cas, per a obtenir un sol domini marcat de manera específica, el precursor que conté el domini d'estudi s'ha d'expressar amb medi mínim marcat isotòpicament. Per a la reacció *in vivo* els dos precursors es coexpressen en *E. coli* amb el sistema dual pskDuet01-pskBAD2. Aquests dos vectors a part dels elements que permeten el seu manteniment simultani en la cèl·lula, tenen promotors diferents, els quals permeten l'expressió independent dels dos precursors. Com a resultat de la producció s'obté el producte de *trans-splicing*. En aquest cas, en primer lloc s'indueix l'expressió del domini marcat, i un cop expressat aquest domini es recullen les cèl·lules i es resuspenen en medi no marcat induint-se l'expressió de l'altre domini [82].

Expressed protein ligation (EPL)

L'EPL permet la unió de proteïnes recombinants i pèptids sintètics en condicions aquoses [83], [84].

Aquesta tècnica permet incorporar de manera específica nombroses modificacions químiques a la proteïna diana com ara fluoròfors, molècules de DNA, modificacions post-traduccionals i aminoàcids no naturals en diferents posicions de la proteïna.

Perquè es doni l'EPL es necessita:

- Un enllaç α -tioèster en la proteïna d'estudi
- Un pèptid sintètic amb una Cys en l'extrem N-terminal

L'enllaç α -tioèster de la proteïna s'aconsegueix amb l'expressió recombinant de la proteïna d'estudi (exteïna N) fusionada amb una inteïna. La inteïna utilitzada ha de presentar una Cys en posició 1 i una mutació en l'última Asn, la qual bloqueja la ciclació d'aquesta. D'aquesta manera, el desplaçament N-C acil es dona formant-se un enllaç tioèster entre l'exteïna N i la inteïna, el qual pot ser hidrolitzat utilitzant l'agent reductor adequat. Com a resultat, s'obté la proteïna recombinant amb un tioèster en l'extrem C-terminal, el qual pot interaccionar i unir-se (trans-tioesterificació) amb el pèptid sintètic amb una Cys en l'extrem N-terminal (Figura 1.12).

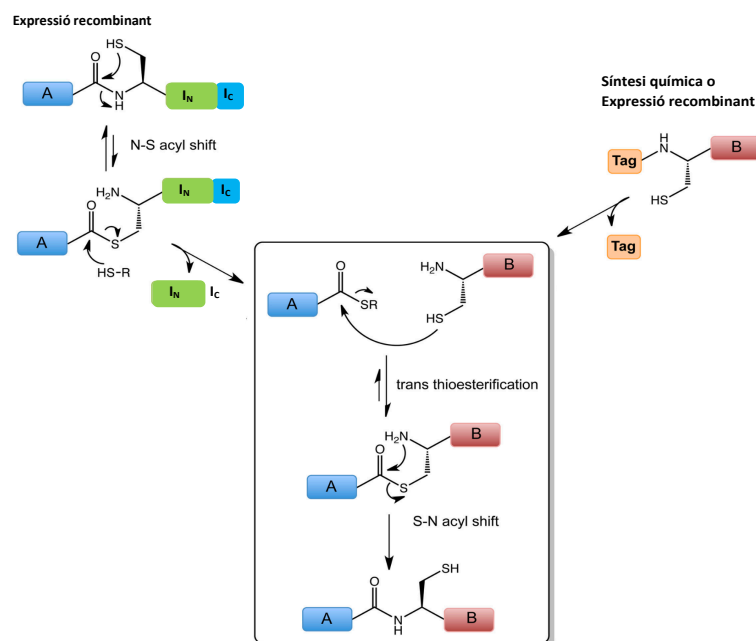


Figura 1.12. *Expressed Protein Ligation (EPL).* A l'esquerra es mostra la formació de la proteïna recombinant amb un α -tioèster mitjançant la tiòlisi de l'intermediari lineal catalitzada per la presència d'un excés d'agent reductor (HS-R). A la dreta es mostra la molècula sintètica amb una Cys en l'extrem N-terminal i la desprotecció d'aquesta. La regió emmarcada correspon a la lligació química pròpiament, en la que la Cys N-terminal catalitza la transesterificació de la proteïna α -tioèster, seguida d'un desplaçament S-N acil que genera el nou enllaç peptídic unint les dues molècules. Esquema reproduït de Vila-Perello and Muir, 2010 [84b].

Ciclació de proteïnes *in vivo* (SICLOPPS)

La ciclació de pèptids s'utilitza generalment en química mèdica per augmentar l'estabilitat i l'activitat del pèptid. La capacitat de generar pèptids cíclics dintre de la cèl·lula ofereix la possibilitat de crear una gran llibreria per a realitzar recerques d'alt rendiment. Això es pot fer inserint en cada extrem de la proteïna diana la inteïna N i la inteïna C d'una inteïna partida. Per a obtenir el pèptid circular la inteïna C s'ha d'inserir en l'extrem N-terminal d'aquest i la inteïna N en l'extrem C-terminal, la qual cosa permet després de la reacció de *trans-splicing* la formació

d'un enllaç peptídic entre els grups α -amí i α -carboxil terminal (Figura 1.13) [85]. Aquesta tècnica s'utilitza molt per a trobar inhibidors a diversos processos, com ara pèptids cíclics inhibidors de la α -sinucleïna [86].

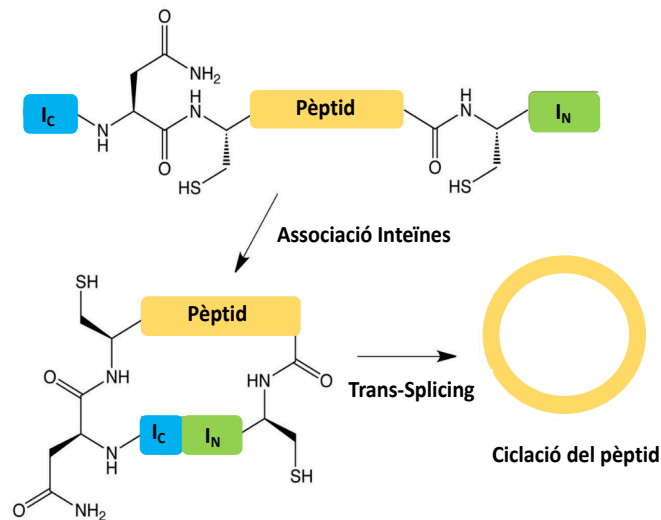


Figura 1.13. Esquema de la generació de proteïnes cíclics a través de la reacció de *trans-splicing*. El pèptid d'interès s'expressa flanquejat per la inteïna C, I_C , i la inteïna N, I_N , en l'extrem N- i C-terminal, respectivament. Esquema reproduït de Vila-Perello and Muir, 2010 [84b].

Conditional protein splicing (CPS)

Com ja s'ha explicat anteriorment en els últims anys s'han reportat diferents inteïnes l'activitat de les quals es troba regulada per factors ambientals. Disposar d'inteïnes les quals puguin modular l'activitat de les exteïnes *in vivo* sota un determinat estímul té un gran interès bioquímic i biomèdic. Imagineu-vos poder controlar en temps l'activitat d'una determinada proteïna? I d'un fàrmac?, podria solucionar molts problemes actuals reals en el camp de la biomedicina. Per aquesta raó, en els últims anys s'han realitzat nombrosos esforços per a obtenir de manera natural o sintètica inteïnes robustes que permetin el CPS.

Tant inteïnes en cis com en trans poden catalitzar CPS, però en el cas d'aquestes últimes la reacció rep el nom de PTS (*Protein Trans-Splicing*).

Actualment, existeixen nombroses estratègies per a modular de manera artificial l'activitat de les inteïnes, aquestes es poden englobar en tres tipus:

- 1- Inhibició de la reacció de *splicing* per introducció d'un grup protector amino (Figura 1.14-a).

La reacció de *splicing* es veu inhibida per la introducció d'un grup protector amino, el qual bloqueja la interacció entre els fragments d'una inteïna partida o bé en el cas de cis-inteïnes bloqueja la reacció de *splicing*. Aquest grup protector s'elimina amb llum ultraviolada. Amb aquest sistema, s'obtenen inteïnes les quals poden modular la seva activitat amb llum UV [87].

- 2- Regulació al·lostèrica de l'activitat de la inteïna per incorporació d'un domini d'unió a lligand (Figura 1.14-b).

La inserció del domini d'unió a lligand interromp la inteïna bloquejant la seva activitat. La unió del lligand al domini provoca un canvi conformacional en aquest, i com a conseqüència la inteïna recupera la seva activitat.

Un bon exemple d'aquest tipus de regulació, és el sistema desenvolupat per Liu i col·laboradors en el qual reemplacen el domini endonucleasa de la inteïna RecA per la regió del domini ER LBD responsable de la unió al seu lligand natural, genticina, i a la molècula petita sintètica 4-HT [88]. Dintre d'aquest context la inteïna només pot catalitzar la reacció de *splicing* en presència de genticina i/o 4-HT. Es creu que en absència del lligand el domini d'unió interacciona amb el complex Hsp90 prevenint el correcte plegament de la inteïna i per tant, bloquejant la seva activitat. La presència del lligand desplaçaria la interacció del domini amb el complex, i com a conseqüència la inteïna adquiriria el seu plegament natiu i restauraria la seva activitat

- 3- Interacció dels dos fragments d'una inteïna partida facilitada per dominis d'heterodimerització (Figura 1.14-c).

En aquest cas s'utilitzen inteïnes partides artificialment, els fragments de les quals presenten una baixa afinitat entre ells i es fusionen amb dominis d'heterodimerització, els quals associen en presència d'una molècula petita o de llum. Un exemple és el sistema sensible a la presència de Ca^{+2} , construït per Kwon i col·laboradors, on s'utilitza la inteïna partida artificialment Sce VMA de *S. Cerevisiae*. Per a promoure l'associació dels dos fragments, s'han utilitzat el domini d'unió a calci (CaM) i el domini d'unió a calmodulina (CaMBD). CaM s'uneix a l'extrem C-terminal de la inteïna N, i CaMB a l'extrem N-terminal de la inteïna C. En presència de calci el domini CaM és activat

estructuralment i pot unir-se al domini CaMBD, com a conseqüència els dos fragments de la inteïna associaran i es donarà la reacció de *splicing* [89].

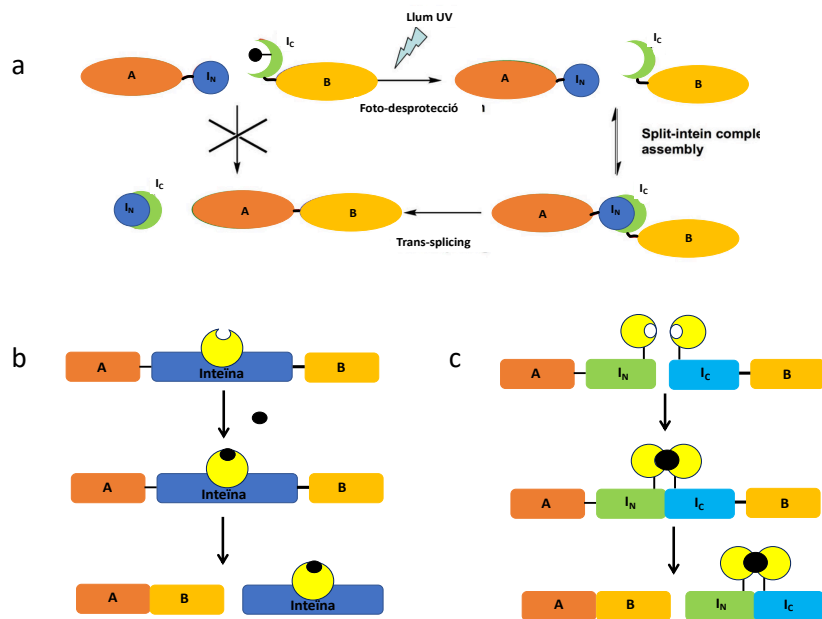


Figura 1.14. *Conditional Protein Splicing (CPS)*. (a) Inteïna partida amb un grup protector amino la qual només serà activa després d'irradiar-la amb la longitud d'ona correcta. (b) Inteïna fusionada a un domini d'unió a lligand, aquesta només catalitzarà la reacció de *splicing* quan s'uneixi el lligand. (c) Inteïna partida la qual només catalitzarà la reacció de *splicing* sota condicions químiques que induïxen la dimerització.

1.2. Adaptació de les proteïnes a elevades temperatures

Els organismes hipertermòfils són aquells que presenten una temperatura òptima de creixement per sobre dels 80°C [90]. Es troben representats en els dominis Bacteria i Archaea i colonitzen tant medis terrestres com aquàtics. Les proteïnes d'aquests organismes, proteïnes hipertermòfiles, s'han adaptat per a ser actives a elevades temperatures presentant una elevada termoestabilitat. Alguns d'aquests enzims són actius a temperatures per sobre del punt d'ebullició de l'aigua [91]. Aquestes proteïnes no treballen bé a temperatures mesofíliques, per sota dels 40°C.

Actualment, es pensa que els hipertermòfils són les primeres formes de vida que van sorgir a la terra [92]. Les proteïnes hipertermofíliques presenten un gran interès des d'un punt de vista evolutiu i molecular. Entendre les bases moleculars de l'elevada termoestabilitat d'aquestes proteïnes permetria desenvolupar noves proteïnes amb una elevada eficiència, així com un ampli ventall d'aplicacions biotecnològiques.

Avui en dia s'ha arribat al consens de què no existeix un únic mecanisme responsable de l'elevada termoestabilitat de les proteïnes hipertermòfiles, sinó que és el resultat de l'optimització de diferents mecanismes i interacció. A continuació, s'explicarà de manera resumida el que es coneix dels determinants moleculars que proporcionen aquesta elevada termoestabilitat.

1.2.1. Composició d'aminoàcids

Fukuchi i col·laboradors van realitzar un estudi estadístic on van comparar la composició d'aminoàcids de 47 proteïnes d'organismes termòfils, temperatura de creixement entre 65°C-70°C, amb 144 proteïnes d'organismes mesòfils. Amb aquest estudi van trobar que les proteïnes d'organismes termòfils presenten més aminoàcids carregats (Lys, Arg i Glu) que les proteïnes d'organismes mesòfils, a excepció de l'Asp el qual es troba en menor freqüència. Aquest increment en el nombre de residus carregats va acompanyat d'una disminució en el número de residus polars (Asn, Gln, Ser i Thr). El decrement en l'ús d'Asn, Gln i Asp és degut a la seva termolabilitat [93]. A elevades temperatures l'Asn i el Gln poden patir una reacció de desaminació, mentre que l'Asn i Asp també poden participar en la hidròlisi de l'enllaç peptídic. Aquesta degradació induïda per temperatura només és significant a temperatures per sobre dels 80°C a pH neutre. Les proteïnes d'organismes termòfils han adaptat els microentorns que envolten aquests residus per a reduir o fins i tot bloquejar aquest tipus de reaccions no desitjades, per exemple amb l'estabilització per ponts d'hidrogen, o bé la presència de cofactors. Per altra banda, també es va observar un augment en l'ús de residus hidrofòbics amb cadena lateral ramificada (Val, Leu i Iso) [94]. Aquests resultats es troben reforçats per altres estudis [95], [96]. Finalment, un altre tret característic és que els aminoàcids aromàtics es veuen afavorits en proteïnes (híper) termofíliques [97], [98]

1.2.2. Optimització de l'estructura terciària i interaccions no covalents

Diverses característiques estructurals estan relacionades amb l'elevada estabilitat de les proteïnes (híper)termofíliques [93].

Interacció amb el solvent i formació de ponts salins

Diversos estudis experimentals i bioinformàtics on es comparen proteïnes mesofíliques i hipertermofíliques demostren que aquestes últimes presenten un increment en el nombre de residus carregats en la superfície de la proteïna [99], [100], i per tant un augment en el número

d'interaccions amb el solvent i ponts salins. La presència d'abundants residus carregats en la superfície, els quals exposen les seves cadenes laterals al solvent, creen una densa xarxa de ponts d'hidrogen entre la proteïna i el solvent [99], [101]. Aquesta extensa capa de solvatació pot augmentar la termoestabilitat de la proteïna, ja que impedeix que les molècules d'aigua penetrin cap al nucli hidrofòbic, i per tant, les fa resistents a la desnaturalització [102][103], [104]. Per altra banda, els residus carregats també poden contribuir a l'estabilitat de la proteïna formant interaccions electroestàtiques. Es troba ben documentat que els ponts salins formats entre les cadenes laterals de dos aminoàcids carregats oposadament tenen un efecte estabilitzador en condicions mesofíliques si es troben enterrats dintre de la proteïna, ja que la formació d'aquestes interaccions presenta una gran penalització per desolvatació [105]. S'ha vist que en cas de proteïnes termofíliques aquest tipus d'interaccions contribueixen més a l'estabilitat de la proteïna, pel fet que la penalització per desolvatació amb la temperatura disminueix [106].

Nucli hidrofòbic, empaquetament molecular i flexibilitat

Un augment de les dimensions en el nucli hidrofòbic sembla ser un tret característic de l'adaptació a elevades temperatures de les proteïnes. Com bé s'ha comentat anteriorment, els ponts salins i les interaccions amb el solvent juguen un paper important en la termoestabilitat de les proteïnes. De tota manera, la presència d'un gran nombre de residus carregats dintre de l'estructura tridimensional nativa de la proteïna comporta un elevat cost entròpic, tot i tractar-se de proteïnes (híper)termofíliques [97]. Per a compensar aquest cos entròpic, les proteïnes (híper)termofíliques, han augmentat el seu contingut en residus hidrofòbics voluminosos, específicament en l'interior de la proteïna, la qual cosa comporta un augment del nucli hidrofòbic [99]. D'altra banda, l'índex hidrofòbic, el qual defineix el grau d'hidrofobicitat d'un aminoàcid (com més gran és l'índex més hidrofòbic és l'aminoàcid), dels residus polars augmenta amb la temperatura substancialment, en canvi, pels residus hidrofòbics és manté constant o augmenta sensiblement [97]. Aquest resultat, juntament amb els diferents estudis comparatius realitzats entre proteïnes (híper)termòfiles i mesofíliques que demostren que les proteïnes (híper)termòfiles tenen un nucli hidrofòbic més gran, suggereixen que les forces hidrofòbiques tenen un gran efecte estabilitzador a elevades temperatures [98], [107], [108]. L'augment en el nombre de residus hidrofòbics voluminosos, a més a més, comporta un augment en el grau d'empaquetament del nucli proteic [109]. Una elevada densitat d'empaquetament molecular es considera una altra característica de les proteïnes adaptades a elevades temperatures [110]–[114]. Un empaquetament dens del nucli hidrofòbic comporta

indirectament una reducció en les cavitats de l'estructura nativa, com a conseqüència, es redueix la probabilitat de que les molècules d'aigua del solvent penetrin en el nucli hidrofòbic, augmentant la resistència a la desnaturalització.

Un altre factor que es creu que augmenta l'estabilitat tèrmica de les proteïnes és la reducció de la flexibilitat de l'estat natiu [115]. Diferents estudis d'intercanvi H-D realitzat per RMN [116]–[118] i de fluorescència (Trp) [119] donen suport a la hipòtesi que les proteïnes hipertermofíliques són més rígides que les seves homòlogues mesofíliques a temperatures mesofíliques. Aquesta hipòtesi implica un canvi conformacional de la proteïna, la qual aniria d'un estat 'rígid' a baixes temperatures, on la proteïna no seria activa, a un estat 'flexible' a temperatures properes a la temperatura òptima de creixement de l'organisme, on la proteïna presentaria la seva activitat.

Així també, les regions desestructurades de les proteïnes així com els loops, actuen com a llocs de desnaturalització tèrmica, amb la qual cosa es postula una reducció d'aquestes regions com a possible mecanisme addicional per augmentar la termoestabilitat en proteïnes [102].

Ponts d'hidrogen

Les proteïnes d'organismes (híper)termòfils presenten un lleuger increment en el nombre de ponts d'hidrogen intramoleculars, el qual no pot explicar per si sol l'elevada estabilitat tèrmica d'aquestes proteïnes [97], [120]. Els ponts d'hidrogen contribuirien a la termoestabilitat de la proteïna des d'un punt de vista entàlpic, ja que la formació d'un pont d'hidrogen comporta una energia lliure mitjana de -1.3 kcal/mol [121].

Interaccions aromàtiques

L'abundància de residus carregats i aromàtics en proteïnes (híper)termofíliques obre la possibilitat d'augmentar les interaccions catió- π i les interaccions π - π en l'interior proteic. Mentre que existeix un consens en el paper estabilitzador de les interaccions catió- π [105],[122]–[124], el paper de les interaccions π - π es troba en debat. Probablement, el motiu d'aquest debat és perquè el principi de les interaccions π - π es poc conegut. En canvi, les bases de les interaccions catió- π són àmpliament conegudes. En proteïnes, aquestes interaccions normalment impliquen un residu d'Arg el qual interacciona a través del grup guanidini de la seva cadena lateral amb la cadena lateral d'un residu aromàtic [122], [124]. La interacció que s'estableix és de caràcter electroestàtic i ajuda a assolir l'estabilitat a elevades temperatures.

Ponts disulfur

Generalment, els ponts disulfur tenen un paper estabilitzador en les proteïnes, ja que disminueixen l'entropia de l'estat desplegat [125]. A causa de l'elevada reactivitat de les Cys i a la susceptibilitat que presenten els ponts disulfur a trencar-se a elevades temperatures el paper estabilitzador dels ponts disulfur no és clar en el cas de les proteïnes (híper)termòfiliques.

Oligomerització

S'ha vist que les proteïnes (híper)termofíliques presenten més tendència a presentar estats oligomèrics, a diferència de les seves homòlogues mesofíliques (Taula 1.2) [120]. L'oligomerització redueix la superfície exposada a solvent, resultant en un increment en l'entropia del sistema, aconseguint un estat termodinàmic més favorable. Per aquesta raó es pensa que l'oligomerització pot ser també un mecanisme especialitzat per a l'adaptació de les proteïnes a elevades temperatures.

Taula 1.2. Comparació del nombre de subunitats entre proteïnes hipertermofíliques i mesofíliques [120].

Proteïna	Nº de subunitats	
	Hipertermòfils	Mesòfils
PGK	2	1
β- Glucosidasa	4	1
Antranilat sintetasa	2	1
Dihidrofolat reductasa	1 o 2	1
Foforibosil-antranilat isomerasa	2	1
β- Glicosidasa	4	2
MkCH	3	2

1.3. La inteïna partida Neq pol

Nanoarchaeum equitans és un organisme paràsit, hipertermòfil i anaeròbic estricte, el qual creix exclusivament en la superfície del crenarchaea *Ignicoccus* sp. Kin4/I i en un rang de temperatures entre els 70°C i els 98°C (temperatura òptima 90°C).

Amb la seqüenciació del seu genoma es va identificar la inteïna Neq pol, la qual fragmenta en dos el gen de la DNA polimerasa B. El precursor N, constituït per la inteïna N (NeqN) i l'extrem N-terminal de la DNA polimerasa, es troba separat en el cromosoma per 83295 pb del precursor C, format per la inteïna C (NeqC) i l'extrem C-terminal de la DNA polimerasa (Figura 1.15) [68].

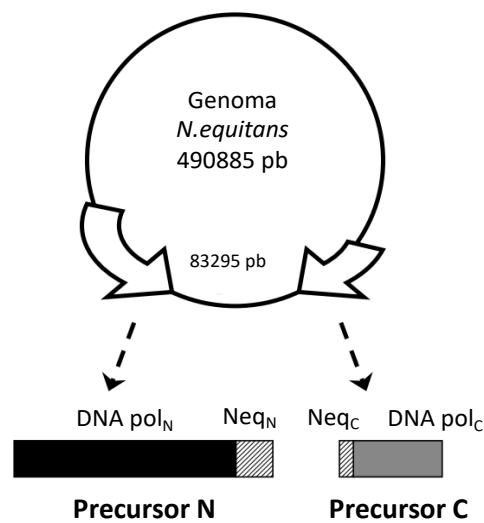


Figura 1.15. Mapa del genoma de *N. Equitans* amb l'estructura de la inteïna partida *Neq pol*. En el mapa del genoma s'indica la mida d'aquest en pb i la distància que separa els dos gens que codifiquen pels dos precursors proteics, el precursor N i el precursor C, estructura dels quals està esquematitzada. Esquema reproduït de Choi et al., 2005 [68].

La inteïna *Neq pol* és l'única inteïna partida descrita avui en dia del domini Archaea, la qual cosa explica la inexistència d'homologia a nivell de seqüència amb la part N-terminal de la resta d'inteïnes partides.

Així mateix aquesta inteïna és la primera inteïna partida descrita fins ara que presenta com a residus catalítics una Ser en posició 1 i una Thr en posició +1. La resta d'inteïnes descrites de la família DnaE presenten en totes dues posicions una Cys.

D'altra banda, la inteïna *Neq pol* és una de les inteïnes partides més petites. El fragment *NeqN* està format per només 98 residus mentre que el fragment *NeqC* en té 30 residus.

Per últim, però no menys important, cal esmentar el treball realitzat per Choi i col·laboradors on van demostrar que el producte de *trans-splicing* augmentava proporcionalment amb la temperatura a partir dels 50°C, arribant al seu màxim a una temperatura de 80°C. En canvi, a temperatures per sota dels 50°C el producte d'aquesta reacció era gairebé inapreciable [68], la qual cosa suggereix que la reacció de *trans-splicing* catalitzada per la inteïna partida *Neq pol* podria estar regulada per la temperatura.

2. Objectius

L'estudi de les inteïnes ha permès desenvolupar noves tècniques en el camp de la biotecnologia i biologia molecular. En els últims anys ha guanyant molt d'interès el Conditional Protein *Splicing* (CPS), on una inteïna modula la seva activitat en funció de factor externs, ambientals o moleculars, permeten regular d'aquesta manera l'activitat de la proteïna hoste. La inteïna partida Neq pol és una bona candidata per a ser utilitzada amb aquesta finalitat, ja que s'ha demostrat que només catalitza la reacció de *trans-splicing* a temperatures per sobre dels 50°C. D'altra banda des d'un punt de vista bioquímic, aquesta inteïna partida és interessant per dos motius:

- 1- és la primera inteïna partida hipertermofílica
- 2- és l'única inteïna partida amb la combinació catalítica Ser1/Thr+1.

En aquesta tesi s'han dut a terme tot un seguit d'estudis fisicoquímics, cinètics i estructurals per tal d'entendre el mecanisme pel qual els dos fragments de la inteïna associen a elevades temperatures i d'aquesta manera com la inteïna catalitza la reacció de *trans-splicing*. Els objectius concrets d'aquesta tesi són:

- 1- Caracteritzar la reacció de *trans-splicing* catalitzada per la inteïna Neq pol en un context exogen (unida a exteïnes que no són les pròpies) per tal d'entendre millor com la temperatura modula l'activitat de la inteïna i en especial el paper de la Ser en la posició 1.
- 2- Caracterització des d'un punt de vista cinètic i termodinàmic del procés d'associació dels dos fragments de la inteïna Neq pol.
- 3- Caracterització fisicoquímica i estructural del fragment N-terminal de la inteïna partida Neq pol de manera aïllada.
- 4- Caracterització fisicoquímica i estructural del fragment N-terminal de la inteïna partida Neq pol formant complex amb el fragment C-terminal.

3. Material i Mètodes

3.1. Material biològic

3.1.1. Soques bacterianes

E. coli DH5- α . Aquesta soca s'ha utilitzat per a obtenir múltiples còpies de DNA dels plasmidis utilitzats en aquesta tesi. Aquesta soca presenta una elevada taxa de transformabilitat, i el genotip *hsdR17(rk-,mk+)* el qual li permet metilar el DNA propi i li impedeix destruir DNA forani. Genotip: F-, ϕ 80 Δ lacZ Δ M15, Δ (lacZYA-argF), U169, deoR, recA1, endA1, *hsdR17(rk-, mk+)*, *phoA*, *supE44*, λ -, *thi-1*, *gyrA96*, *relA1*

E. coli BL21(DE3). Aquesta soca s'ha utilitzat per a l'expressió recombinant de les diferents construccions. Ha estat obtinguda de Novagen, USA. Aquesta soca és deficient per les proteases Lon i OmpT. La peculiaritat d'aquesta soca és que conté el pròfag DE3 que transporta el gen de la T7 RNA polimerasa sota el control del promotor lacUV5, el qual es induïble per IPTG. Genotip: F-, *ompT-*, *hsdS(rB- mB-)*, *dcm+*, *gal*, λ (DE3), *lon-*

3.1.2. Vectors procariotes utilitzats per a l'expressió de proteïnes recombinants

El gens que codifiquen pels fragments NeqN i NeqC de la Neq pol es van obtenir del vector comercial **pIDT-Smart-Kan-NEQ** (Integrated DNA technologies, USA), el qual té resistència a la kanamicina.

El gen que codifica per la NeqSC es va obtenir del vector comercial **pIDTSMART-AMP:NEQsc** (Integrated DNA technologies, USA), el qual té resistència a l'ampicil·lina.

El vector d'expressió **pSKDuet01** (Addgene, USA [33]) conté el promotor T7 el qual es reconegut per la T7 RNA polimerasa induïble per IPTG. Seguidament, hi ha la seqüència que codifica per una cua d'histidines, la proteïna GB1 i la NpuN. Aquest vector presenta resistència a kanamicina i un origen de replicació d'alt nombre de còpies RSF1030.

El plasmidi **pSKBAD2** (Addgene, USA [33]) conté un origen de replicació de baix nombre de còpies, ColEI. *A més a més, conté el promotor araBAD el qual s'indueix amb arabinosa i resistència a ampil·lina. Entre les dianes de restricció NdeI i Hind3 es troba la seqüència que codifica per la NpuC i la GB1.*

Aquests dos plasmidis presenten orígens de replicació diferents i compatibles, així com resistència a dos antibiòtics diferents, la qual cosa permet coexpressa-los en una mateixa cèl·lula bacteriana.

El plasmidi **pBADM-20** (EMBL, Germany), va ser dissenyat per augmentar la solubilitat de les proteïnes a expressar, facilitant d'aquesta manera la seva purificació. Aquest vector codifica per la proteïna tioredoxina A (TrxA) seguida d'una cua d'histidines i la seqüència diana de la proteasa TEV. L'expressió de la proteïna recombinant es troba sota el control del promotor araBAD, l'expressió del qual s'indueix amb arabinosa. D'altra banda, presenta un origen de replicació d'*E. coli*, pUC i com a marcador de selecció presenta resistència a l'ampicil·lina.

Tots els vectors descrits a continuació tenen el promotor T7, l'expressió del qual s'indueix amb IPTG.

El vector **pET-21a(+)** (Addgene, USA) conté en l'extrem N-terminal la seqüència de l'epítot T7, la qual facilita la detecció de proteïnes per tècniques immunohistoquímiques, i en l'extrem C-terminal una cua d'His. Aquest plasmidi té resistència a l'ampicil·lina i un origen de replicació.

El plasmidi **pET-22b(+)** (Addgene, USA) en l'extrem N-terminal con té la seqüència senyal pelB, la qual potencia la localització periplasmàtica de la proteïna, i en l'extrem C-terminal una seqüència de 6 His. Aquest plasmidi té resistència a l'ampicil·lina i un origen de replicació pBR322.

El vector d'expressió **pET-28a(+)** (Addgene, USA) en l'extrem N-i C-terminal codifica per una cua d'histidines. Aquest plasmidi té resistència a kanamicina i un origen de replicació

El plasmidi **pET-29a(+)** (Addgene, USA) en l'extrem N-terminal conté la seqüència que codifica pel pèptid S, el qual facilita la purificació de la proteïna recombinant en una columna d'afinitat amb S-ribonucleasa A, i la seqüència diana per la trombina. Aquest vector té resistència a la kanamicina i un origen de replicació pBR322.

3.2. Tècniques de DNA recombinant

3.2.1. Extracció i purificació de DNA plasmídic

L'obtenció d'una elevada quantitat de DNA plasmídic pur és un pas clau en biologia molecular, ja que constitueix el material de partida de moltes tècniques, com ara la transformació de cèl·lules competents i el clonatge de gens. Per a obtenir DNA plasmídic d'elevada qualitat s'ha utilitzat el kit comercial, QIAprep® Spin Miniprep Kit (Agilent Technologies, USA), seguint el protocol indicat per la casa comercial.

3.2.2. Digestió amb enzims de restricció

Els enzims de restricció utilitzats en aquest treball han estat subministrats per Roche S.L (Roche Molecular Biochemical, Germany)

Per a assegurar l'èxit de les digestions i reduir al màxim reaccions no desitjades el volum de l'enzim de restricció no supera en cap de les reaccions el 10% del volum total (concentració final de glicerol $\leq 5\%$), les reaccions s'incuben a 37°C durant un temps màxim de 3 h.

3.2.3. Mutagènesi dirigida

Per a introduir mutacions puntuals es va utilitzar el kit QuickChange™ (Agilent Technologies, USA) segons les instruccions de la casa comercial. Per tal d'eliminar el DNA parental, abans de transformar les cèl·lules DH5 α , el producte de la PCR es digereix amb *DpnI* a 37°C durant 1 hora.

3.2.4. Disseny dels oligonucleòtids

Per al disseny dels oligonucleòtids es van seguir les següents premisses. Els oligonucleòtids per a mutagènesi dirigida han de tenir una longitud aproximada de 21 nucleòtids, amb la mutació introduïda en el centre de la seqüència. Han de presentar un contingut GC > 40%, i si pot ser, l'extrem 3' hauria d'acabar amb una o més C o G.

La temperatura de fusió (T_m) es calcula seguint la següent fórmula:

$$T_m = 81.5 + 0.41(\%GC) - 675/N - \% \text{ de desaparellament}$$

on N és el nombre de nucleòtids totals de l'oligonucleòtid, %GC és el contingut en guanina + citosina en % i el % desaparellament és el percentatge de nucleòtids que no hibriden amb la seqüència patró.

Els oligonucleòtids utilitzats en aquest treball han estat subministrats per Roche Diagnostics S.L (Roche Molecular Biochemical, Germany).

3.2.5. Hibridació de nucleòtids

Els oligonucleòtids són subministrats en forma de liofilitzat. El primer pas per a realitzar la hibridació és resuspendre el liofilitzat amb 200 μ l d'H₂O MilliQ, seguidament determinar la seva concentració per absorbància a 260 nm, per acabar preparant concentracions equimolars dels dos oligonucleòtids. Finalment, es barregen volums iguals dels dos oligonucleòtids i s'incuben durant 1 hora a 90°C i es deixen refredar gradualment fins a temperatura ambient.

Taula 3.1. Oligonucleòtids utilitzats en aquest treball.

	Nom	Sequència (5'-3'-)	Objectiu	Plasmidi	Tm (°C)
1	T7 PROM	TAATACGACTCACTATAGGG	Seqüenciació		-
2	T7 TERM	CTAGTTATTGCTCAGCGGTG			-
3	pBADfw	ATGCCATAGCATTTTTATCC	Seqüenciació i ampliació del gen de la GB1.		-
4	pBADrev	GATTTAATCTGTATCAGG			-
5	native_NeqN1	CTAGCGGATCCTTTAAAGTGATTTATGGTG	Clonatge per a introduir 6 residus de l'extrem C-terminal de l'exteïna N de N.equitans	pET21-a(+)_T7-NeqN	
6	native_NeqN2	GCCTAGGAAATTTCACTAAATACCACCTAG			
7	wt_NeqN1	GTGATTTATGGTGATTCCATCATGGACACTG	Mutagènesi dirigida per a introduir l'Asp en posició -1 de l'exteïna N de N.equitans	pET21-a(+)_FKVIYG-NeqN	70.5
8	wt_NeqN2	CAGTGTCATGATGGAATCACCATAAATCAC			
9	D-1_NeqN1	GGGAAGGATTTCAGAAGGCGACTCCATCATGGACACTG	Mutagènesi dirigida per introduir l'Asp en posició -1 de l'extrem N-terminal de N.equitans	pSKDuet01_H6MBP-NeqN/pET-28a(+)_H6MBP-NeqN	85.3
10	D-1_NeqN2	CAGTGTCATGATGGAGTCGCCTTCTGAAATCCTTCCC			
11	Ser1Cys1	GATTTATGGTGATTGCATCATGGACACTG	Mutagènesi dirigida per a introduir la mutació Ser1Cys	pSKDuet01_H6GB1(wt)-NeqN	71.7
12	Ser1Cys2	CAGTGTCATGATGCAATCACCATAAATC			
13	Stop_NeqC1	CTACACCGTAACGGAAAAGCTTGC GGCCGCACTCG	Eliminació codó stop de l'extrem C-terminal de la GB1		82.7
14	Stop_NeqC2	CGTGTGCGGCCGCAAGCTTTCCGTTACGGTGTAG			
15	EcoR1	GGTTCATAGCTTCTAATGAATCAATGGTACCCAG	Amplificació del gen GB1	pSKBAD02	79.8

3.2.6. Electroforesi en gel d'agarosa

L'electroforesi en gel d'agarosa s'ha utilitzat per a purificar fragments de DNA, analitzar preparacions de DNA plasmídic i productes de PCR. En funció de la mida del DNA a analitzar el rang d'agarosa a oscil·lat entre 0.8-1.2%. L'agarosa (Ecogen, Spain) es dissolt escalfant-la al microones amb tampó TAE (40 mM Tris-Acetate, 2mM EDTA, pH 8.0). Un cop refredada, s'afegeix el bromur d'etidi (Sigma, USA) a una concentració final de 50 µg/ml. Un cop solidificat el gel, l'electroforesi es realitza amb tampó TAE. Les mostres es preparen barrejant-les amb tampó de càrrega x6 (Fermentas, USA) seguint una relació 5:1. Per estimar la mida molecular, s'utilitza un marcador de DNA (Fermentas, USA). L'electroforesi es realitza aplicant un corrent elèctric 110 V durant 30 minuts.

3.2.7. Purificació de fragments de DNA a partir de gels d'agarosa (GenClean)

La purificació dels fragments de DNA s'ha realitzat amb el kit comercial QIAquick® Gel Extraction Kit (Agilent Technologies, USA), el qual permet purificar fragments d'entre 70 pb i 10 Kpb. L'extracció i purificació es va realitzar seguint el protocol descrit per la casa comercial.

3.2.8. Lligació de l'insert de DNA amb el vector d'expressió utilitzant la T4 lligasa

Per a inserir de manera estable l'insert en el vector en aquesta tesi s'ha utilitzat la lligasa de T4 (Roche Molecular Biochemical, Germany). La reacció de lligació es munta respectant la proporció aconsellada per la casa comercial de 0.1-0.5 U/µg de DNA, i és deixa incubar tota la nit a 18°C.

El resultat de la reacció de lligació es pot utilitzar directament per a transformar cèl·lules d'*E. coli* DH5α. Si no es vol utilitzar immediatament aquesta reacció es pot congelar a -20°C.

3.2.9. Transformació de cèl·lules competents d'*E. coli*

La soca d'*E. coli* corresponent, DH5α o BL21(DE3), es va deixar créixer tota la nit a 37°C i 180 rpm amb medi Luria Bertani [126]. A partir d'aquest cultiu es va preparar un reinòcul (1/100) en 10 ml de medi LB fresc, el qual es va deixar créixer a 37°C i 180 rpm, fins arribar a una OD₅₅₀=0.6. Seguidament, el cultiu es va centrifugar a 4000 rpm a 4°C, durant 10 minuts. Les cèl·lules es van resuspendre amb 10 ml d'una solució freda de 100 mM CaCl₂ fred i estèril. Després de 30 minuts,

les cèl·lules es recullen per centrifugació, 4000 rpm 4°C durant 10 minuts, i es resuspenen amb 1 ml de la solució freda de 10 mM CaCl₂. Quan va ser necessari, la solució es va guardar a -80°C amb 20% de glicerol. Als 100 µl de cèl·lules competents s'afegeixen 5 µl de DNA provinent de mutagènesi dirigida o 1 µl de DNA plasmídic pur. Després de 1-2 hores en gel, s'incuba durant 1 min a 42°C, i ràpidament es torna a posar en gel una altra vegada 10 minuts. A continuació, s'afegeixen 600 µl de LB en el cas que el DNA vingui de mutagènesi dirigida o de clonatge, o 300 µl en el cas de DNA plasmídic, i es deixen incubar durant 1 hora a 37°C. Finalment, les cèl·lules es sembren en plaques d'agar que contenen l'antibiòtic adient per al plasmidi.

3.2.10. Seqüenciació de DNA

Totes les construccions emprades en aquesta tesi han estat verificades per seqüenciació. Per a la seqüenciació s'han utilitzat els oligonucleòtids T7PROM, T7TERM, pBADforward i pBAD reverse en funció del plasmidi (Taula 3.1).

El DNA ha estat seqüenciat utilitzant el mètode Sanger [127]. S'ha utilitzat el kit comercial ABI PRIMS® dRhodamine Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems, USA).

Les mostres provinents de la PCR de seqüenciació es van precipitar afegint 0.1 volums de 3M acetat sòdic a pH 5.5 i 3 volums d'etanol absolut fred. Després d'1 hora a -80°C, es van centrifugar, i el sediment es va rentar amb 500 µl d'etanol al 70% fred. Finalment, el sediment es resuspèn amb 20 µl d'H₂O formamida desionitzada (Bio-Rad, USA), i s'analitza en un analitzador ABI Prism 310 Genetic Analyzer (Applied Biosystems, USA) en el Laboratori d'Ictiologia Genètica de la Universitat de Girona.

3.3. Clonatge, expressió i purificació

3.3.1. Construccions per la caracterització de la inteïna N de *N. equitans*

T7-NeqN

La seqüència corresponent a la NeqN es troba indicada en negreta.

MASMTGGQQMGRG**SIMDTEIEVIENGIKKKEKLSDLFNKYYAGFQIGEKHYAFPPDLVYDGER**
VWKVYSIIKHETETDLYEINGITLSANHLVLSKGNWVKAKEYENKNN

Per a construir el plasmidi pET-21a(+)_T7-NeqN, el vector pIDT-Smart-Kan-_NEQ es va digerir amb els enzims de restricció *Bam*HI i *Hind*III. A continuació l'insert, NeqN, es va lligar en el vector comercial pET21-a(+), prèviament, digerit amb els mateixos enzims de restricció.

Amb el producte del clonatge, pET21-a(+)_T7-NeqN, es van transformar cèl·lules *E. coli* BL21(DE3), les quals van créixer en 1L de medi LB suplementat amb 100 µg/ml d'ampicil·lina a 37°C fins a una OD₆₀₀ = 0.6. L'expressió es va induir afegint 0.5 mM IPTG (Fermentas, USA) durant 5 hores. Després, les cèl·lules es van recollir per centrifugació i es van resuspendre amb 20 mM tampó fosfat, 500 mM NaCl, 1 mM EDTA a pH 8.0. Seguidament, les cèl·lules es van lisar passant-les dues vegades per la premsa de French (SLM-Aminco, Spectronic Instruments, UK) a 11.000 Psi, i la fracció soluble es va recuperar per centrifugació. La fracció soluble es va incubar durant 10 minuts a 70°C, tot seguit, per a separar les proteïnes desnaturalitzades es va centrifugar durant 20 minuts a 15.000 rpm a 4°C. La fracció soluble es va dialitzar enfront de 20 mM tampó fosfat, 1mM EDTA a pH 8.0, per ser purificada per cromatografia de bescanvi aniònic.

TrxH6TEVT7-NeqN

La seqüència és la corresponent al precursor abans d'eliminar la cua d' histidines amb la digestió de la proteasa TEV. El punt de tall de la proteasa s'indica amb una barra vertical i la seqüència corresponent a la NeqN es troba indicada en negreta.

MSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQN
 PGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSGHMHSHSSGENL
 YFQ | GAMASMTGGQQMGRG**SIMDTEIEVIENGIKKKEKLSDLFNKYYAGFQIGEKHYAFPPLY**
VYDGERWVKVYSIIKHETETDLYEINGITLSANHLVLSKGNWVKAKEYENKNN

Per l'obtenció del plasmidi pBADM_20-TrxH6TEVT7-NeqN va ser necessari realitzar un clonatge intermedi, pTYB4_T7-NeqN, on es va incorporar la diana de restricció *Nco*I, en l'extrem N-terminal de l'epítip T7, la qual era necessària per al seu clonatge en pBADM-20.

En primer lloc, el vector pET21-a(+)_T7-NeqN va ser digerit amb *Nhe*I i *Hind*III, obtenint com a resultat el fragment T7-NeqN, el qual va ser lligat al vector pTYB4 (New England Biolabs, MA, USA) prèviament digerit amb els mateixos enzims de restricció. Com a resultat es va obtenir el plasmidi pTYB4_T7-NeqN. Seguidament, es varen digerir simultàniament el vector d'origen, pTYB4_T7-NeqN, i el vector de destí, pBADM-20 (EMBL), amb *Nco*I i *Hind*III. L'insert que codifica per la T7-NeqN es va lligar al plasmidi pBADM-20 digerit, obtenint com a resultat el plasmidi pBADM_20-TrxH6TEVT7-NeqN.

Per a la seva expressió, es va fer créixer un cultiu d'*E. coli* BL21(DE3), prèviament transformat amb el vector pBADM-20_TrxH6TEVT7-NeqN, a 37°C en medi LB complementat amb 100 µg/ml d'ampicil·lina fins a arribar a una OD₆₀₀ de 0.6. Es va induir l'expressió amb 0.2% L-arabinosa (Fermentas, USA) i es va deixar créixer durant 4 hores a 37°C i 250 rpm. Després de la producció, les cèl·lules es van recollir per centrifugació i es van 20 mM fosfat, 500 mM NaCl, 5 mM Imidazol a pH 7.2. Seguidament, es van lisar passant-les dues vegades per la premsa de French (SLM-Aminco, Spectronic Instruments, UK), 11.000 psi. Per a la purificació del precursor, TrxH6TEVT7-NeqN, a partir de la fracció soluble, es van utilitzar columnes d'afinitat Ni-NTA (HisTrap™ FF FPLC column, GE Healthcare, Chicago, ILL, USA). L'elució es va aconseguir realitzant un gradient lineal de 5-250 mM imidazol en 30 minuts utilitzant com a tampó A: 20 mM tampó fosfat, 500 mM NaCl, 5 mM imidazol a pH 7.2, i com a tampó B: 20 mM tampó fosfat, 500 mM NaCl, 250 mM imidazol a pH 7.2. L'elució es va dialitzar enfront de 20 mM fosfat, 1mM EDTA a pH 8.0 i es va digerir amb la proteasa TEV. Per a realitzar la digestió es va utilitzar una relació 1:5 proteasa TEV: proteïna, i es va deixar incubar durant 3 hores a 30°C. El producte de digestió es va tornar a dialitzar enfront de 20 mM fosfat, 500 mM NaCl, 20 mM Imidazol a pH 8.0. Finalment, per a obtenir la T7-NeqN pura es va utilitzar una cromatografia d'afinitat Ni-NTA, on la columna es va rentar amb 5 volums de columna amb 20 mM fosfat, 500 mM NaCl, 20 mM Imidazol a pH 8.0 i l'elució es va realitzar amb 6 volums de columna amb 20 mM fosfat, 500 mM NaCl, 250 mM Imidazol a pH 8.0.

H6TEV(wt)-NeqN

La seqüència és la corresponent al precursor abans d'eliminar la cua d' histidines amb la digestió de la proteasa TEV. El punt de tall de la proteasa s'indica amb una barra vertical, la seqüència corresponent a la NeqN es troba indicada en negreta i els residus salvatges de l'exteïna N de *N.equitans* es troben subratllats.

MHHHHHSSGENLYFQ | GAMASGSFKVIYGD**SIMDTEIEVIENGIKKKEKLSDLFNKYYAGFQI**
GEKHYAFPPDLYVDGERWVKVYSI**IKHETETDLYEINGITLSANHLVLSKGNWVKAKEYENKN**
N

El plasmidi pET22-b(+)_H6TEV(wt)-NeqN es va construir en diferents passos (Figura 3.1). En primer lloc, es va realitzar un clonatge en el qual es va substituir la seqüència de l'epítot T7 (pET21-a(+)_T7-NeqN) per la seqüència que codifica pels residus de l'exteïna salvatge FKVIYG. A causa de la petita mida de l'insert a clonar, va ser necessari sintetitzar l'insert de manera

artificial. Per això es varen dissenyar dos oligonucleòtids complementaris (Oligonucleòtids 5 i 6; Taula 3.1), els quals presentaven en els seus extrems els extrems cohesius originats pels enzims de restricció *NheI* i *BamHI*. Prèviament al clonatge es va realitzar la hibridació dels oligonucleòtids (Apartat: 3.2.5). El casset format pels dos oligonucleòtids hibridats es va clonar en el vector pET21-a(+)-T7-NeqN digerit prèviament amb els enzims de restricció *NheI* i *BamHI*. Com a resultat es va obtenir el plasmidi pET21-a(+)-FKVIYG-NeqN.

Sobre el plasmidi pET21-a(+)-FKVIYG-NeqN i per a obtenir pET21-a(+)-(wt)-NeqN, es va realitzar mutagènesi dirigida (Apartat: 3.2.3) utilitzant els oligonucleòtids 7 i 8 descrits en la Taula 3.1.

Per a obtenir la construcció H6TEV(wt)-NeqN en el plasmidi d'expressió pET22-b(+), es varen realitzar dos clonatges. El primer, en el qual es va clonar a partir del vector pBADM-20_H6TEVT7-NeqN el fragment H6TEVT7-NeqN en el vector pET22-b(+). Per a realitzar aquest clonatge es varen utilitzar els enzims de restricció *NdeI* i *Hind3*. Com a resultat es va obtenir el plasmidi pET22-b(+)-H6TEVT7-NeqN. Finalment, i per a obtenir el plasmidi pET22-b(+)-H6TEV(wt)-NeqN, es va digerir el vector pET21-a(+)-(wt)-NeqN amb *BamHI* i *HindIII*. L'insert resultant de la digestió, (wt)-NeqN, es va lligar en el plasmidi pET22-b(+)-H6TEVT7-NeqN digerit, prèviament, amb els mateixos enzims de restricció.

Per a expressar H6TEV(wt)-NeqN, cèl·lules d'*E. coli* BL21(DE3) prèviament transformades amb el vector pET-22b(+)-H6TEV(wt)-NeqN es van fer créixer amb 2L de cultiu LB suplementat amb 100 µg/ml d'ampicil·lina a 37°C fins a una OD₆₀₀ = 0.6. L'expressió es va induir afegint 0.5 mM IPTG durant 5 hores a 37°C. Després de la producció, les cèl·lules es van recollir per centrifugació i es van resuspendre amb 20 mM fosfat, 500 mM NaCl, 20 mM Imidazol a pH 7.6. Posteriorment, es van lisar passant-les dues vegades per la premsa de French, 11.000 psi. La fracció soluble es va recuperar per centrifugació i es va injectar a la FPLC. Per a la purificació del precursor, H6TEV(wt)-NeqN, es van utilitzar columnes d'afinitat Ni-NTA (HisTrap™ FF FPLC column, GE Healthcare, Chicago, ILL, USA), i per a la seva elució es va realitzar un gradient lineal de 20 -175 mM imidazol en 21 minuts (Tampó A: 20 mM fosfat, 500 mM NaCl, 20 mM Imidazol a pH 7.6; Tampó B: 20 mM fosfat, 500 mM NaCl, 250 mM Imidazol a pH 7.6). La fracció pura es va dialitzar enfront de 20 mM fosfat, 50 mM NaCl a pH 8.0, després es va tractar amb la proteasa TEV, relació 1:5 proteasa TEV: proteïna, durant 6 hores a 25°C. El producte de la digestió es va dialitzar enfront de 20 mM fosfat, 300 mM NaCl a pH 8.0, i es va purificar per cromatografia d'afinitat Ni-NTA, on es recuperava la proteïna d'interès del través. Tot el procés de purificació es va seguir per 12%-SDS-PAGE. La fracció pura que conté la (wt)-NeqN es va dialitzar enfront d'aigua Milli-Q i es va liofilitzar (Annex E).

En aquest treball es refereix a aquesta construcció, (wt)-NeqN, com NeqN.

La massa molecular de la NeqN i la puresa es van confirmar per espectrometria de masses, MALDI-TOF, utilitzant un espectròmetre de masses Bruker-Biflex (Bruker, USA) en el Servei Tècnic de Recerca de la Universitat de Girona (Annex E).

Per als estudis estructurals de la variant wt-NeqN per difracció de raig X, es va seguir el mateix protocol amb unes petites modificacions. La primera és que en comptes d'utilitzar tampó fosfat es va utilitzar Tris-HCl, ja que el fosfat forma cristalls els quals podrien dificultar la resolució de l'estructura. L'altra modificació és que després de la purificació es va realitzar un pas addicional on la NeqN es va repurificar per cromatografia d'exclusió molecular (SEC). Després de la cromatografia les mostres es concentraven centrifugant a 4000 rpm a 4°C en tubs Amicon® (Merck) de 15 ml correctament equilibrats prèviament fins a obtenir la concentració desitjada.

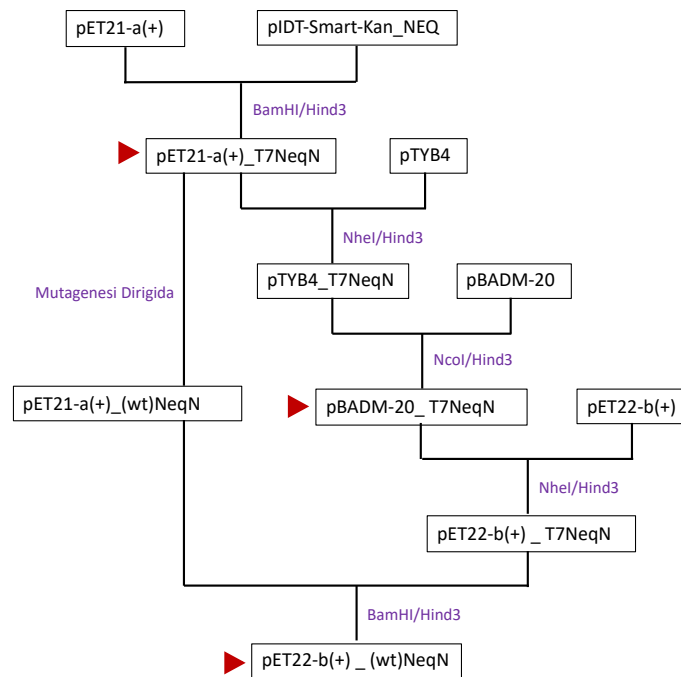


Figura 3.1. Esquema de l'obtenció de les diferents construccions utilitzades en el procés de caracterització de la NeqN. Les construccions utilitzades estan assenyalades amb un triangle.

H6TEVT7-NpuN

La seqüència és la corresponent al precursor abans d'eliminar la cua d'histidines amb la digestió de la proteasa TEV. El punt de tall de la proteasa s'indica amb una barra vertical i la seqüència corresponent a la NpuN es troba indicada en negreta.

MHHHHHSSGENLYFQ | GAMASMTGGQQMGRGS**CLSYETEILTVEYGLLP**IGKIVEKRIECTVY
SVDNNGNIYTQPVAQWHRGEOEVFEYCLEDGSLIRATKDHKFM**TV**DG**QMLP**IDEIFERELDLM
RVDNLPN

Per a obtenir el plasmidi pET22-b(+)_H6TEVT7-NpuN els vectors pSKDuet01 i pET22-b(+)_H6TEVT7-NeqN es van digerir amb els enzims de restricció *Bam*HI i *Hind*III i el producte de la digestió es va lligar. Aquesta construcció es va utilitzar com a control en l'assaig de solubilitat (Apartat: 3.4.4).

3.3.2. Construccions per a la caracterització de la reacció de *trans-splicing*, del procés d'associació i del complex NeqN/NeqC

La seqüència corresponent a la inteïna es troba senyalada en negreta, i els residus salvatges de les exteïnes subratllats.

H6MBP(GD)-NeqN

La seqüència és la corresponent al precursor N. La seqüència de la NeqN es troba indicada en negreta i els residus salvatges de l'exteïna N de *N.equitans* es troben subratllats.

MGSSHHHHHSSGLVPRGSHMKIEEGKLVIIWINGDKGYNGLAEVGGKFEKDTGIKVTVEHPDKL
 EEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYP
 IAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKY
 ENKYDIKDVGVNAGAKAGLTFVLVDIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAWSNID
 TSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLG
 AVALKSYEEELAKDPRIAATMENAQKGEIMPNI PQMSAFWYAVRTAVINAASGRQTVDEALKDA
 QTNSSNNNNNNNNNNNLGIEGRISEGD**SIMDTEIEVIENG**IKK**KEKLSDLFNKY**YAG**FQIGEKH**
YAFPPDLVYDGERWVKVYSIIKHETETDLYEINGITLSANHLVLSKGNWVKAKEYENKNN

Per a obtenir el plasmidi pSKDuet01_H6MBP(GD)-NeqN es van realitzar diferents passos. En primer lloc, es va digerir el plasmidi pIDT-Smart-Kan-NEQ amb els enzims de restricció BamHI i HindIII, les dianes dels quals flanquegen el gen de la NeqN. Paral·lelament, per a obtenir el gen de la MBP es va digerir el vector pHM45 (cedit pel Dr. Tom Muir, Rockefeller University, NY, USA [128]) amb les endonucleases NdeI i BamHI. Els fragments resultants de les dues digestions es van inserir en el vector d'expressió pET-28a(+), el qual prèviament havia estat digerit amb NdeI i HindIII. Com a resultat, es va obtenir el vector, pET-28a(+)H6MBP-NeqN. Seguidament, utilitzant els oligonucleòtids 9 i 10 (Taula 3.1) es va introduir per mutagènesi dirigida sobre el vector pET-28a(+)H6MBP-NeqN el residu salvatge Asp-1. Com a resultat es va obtenir el plasmidi pET-28a(+)H6MBP(GD)-NeqN. A posteriori, els vectors pET-28a(+)H6MBP(GD)-NeqN i pSKDuet01 es van digerir amb els enzims de restricció NcoI i HindIII. Finalment, el fragment H6MBP(GD)-NeqN es va lligar en el plasmidi pSKDuet01.

Per a l'expressió individual del precursor N, H6MBP(GD)-NeqN, cèl·lules d'*E. coli* BL21(DE3) van ser transformades amb el plasmidi pSKDuet01_H6MBP(GD)-NeqN i es van fer créixer en 200 ml de medi LB amb ampicil·lina (100 µg/ml) a 37°C fins a una OD600 = 0.6. L'expressió es va induir afegint 0.5 mM IPTG durant 5 hores a 37°C. Després de la producció les cèl·lules es van recollir per centrifugació i es van resuspendre amb tampó de lisi (20 mM fosfat a pH 7.4, 300 mM NaCl, 10 mM imidazol, 1 mM PMSF). Les cèl·lules es van lisar per sonicació, 3 cicles de 30 segons, i la fracció soluble es va recuperar per centrifugació. La fracció soluble es va barrejar amb 1ml d'agarosa Ni-NTA (Quiagen, GmbH, Germany), correctament equilibrada amb tampó de lisi, i es va incubar durant tota la nit a 4°C. Després de la incubació, es va carregar a una columna. Després de descartar el través, la columna es va rentar amb 5 volums de columna (VC) del tampó de rentat 1 (tampó de lisi), 5 VC del tampó de rentat 2 (tampó de lisi + 25 mM imidazol), 3VC de tampó de rentat 3 (tampó de lisi + 50 mM imidazol). La proteïna es va eluir amb tampó d'elució (tampó de lisi + 250 mM imidazol) en 2 fraccions de 3 VC. Les diferents fraccions obtingudes s'analitzen per 12% SDS-PAGE (Annex A). En aquest cas la fracció pura no es va poder analitzar per espectrometria de masses MALDI-TOF perquè no s'ionitzava amb prou eficiència.

H6GB1(wt)-NeqN

La seqüència és la corresponent al precursor N. La seqüència de la NeqN es troba indicada en negreta i els residus salvatges de l'exteïna N de *N.equitans* es troben subratllats.

MGSSHHHHHHSSGLVPRGSHMQYKLIILNGKTMKGETTTEAVDAATAEKVFKQYANDNGVDGEWT
YDDATKTYTVTEGSFKVIYGD**SIMDTEIEVIENGIKKKEKLSDLFNKYYAGFQIGEKHYAFPPD**
LYVYDGERWVKVYSIIKHETETDLYEINGITLSANHLVLSKGNWVKAKEYENKNN

Per obtenir el vector d'expressió pSKDuet01_H6GB1(wt)-NeqN es van digerir paral·lelament els vector pET-21a(+)-(wt)-NeqN, descrit anteriorment, i pSKDuet01 amb els enzims de restricció *Bam*HI i *Hind*III. El producte de la digestió es va lligar obtenint la construcció desitjada, H6GB1(wt)-NeqN, en pSKDuet01.

Aquest plasmidi s'utilitzà per a obtenir l'expressió del complex NeqN/NeqC *in vivo*. El procediment es detallarà més endavant.

H6GB1(wt)-NeqN(S1C)

La seqüència és la corresponent al precursor N. La seqüència de la NeqN es troba indicada en negreta i els residus salvatges de l'exteïna N de *N.equitans* es troben subratllats. En vermell està indicada la mutació Ser1Cys.

MGSSHHHHHHSSGLVPRGSHMQYKLIILNGKTMKGETTTEAVDAATAEKVFKQYANDNGVDGEWT
YDDATKTYTVTEGSFKVIYGD**CIMDTEIEVIENGIKKKEKLSDLFNKYYAGFQIGEKHYAFPPD**
LYVYDGERWVKVYSIIKHETETDLYEINGITLSANHLVLSKGNWVKAKEYENKNN

Per a obtenir el plasmidi pSKDuet01_H6GB1(wt)-(S1C)NeqN per mutagènesi dirigida es va introduir la mutació Ser1Cys utilitzant els oligonucleòtids 11 i 12 sobre el plasmidi pSKDuet01_H6GB1(wt)-NeqN (Taula 3.1)

Aquest plasmidi s'utilitzà per a obtenir l'expressió del complex NeqN/NeqC *in vivo*. El procediment d'expressió i purificació es detallarà més endavant.

NeqC-(TDS)H6

La seqüència és la corresponent al precursor C. La seqüència de la NeqC es troba indicada en negreta i els residus salvatges de l'exteïna C de *N. equitans* es troben subratllats.

MRYLGKKRVILYDLSTESGKFFVNGLVLHNTDSDPNSSSVDKLAAALEHHHHHH

El plasmidi pET-29a(+)_NeqC es va obtenir digerint el plasmidi pIDT-Smart-Kan_NEQ amb els enzims de restricció *NdeI* i *EcoRV*, les dianes de les quals flanquegen el gen de la NeqC. L'insert obtingut de la digestió es va lligar en el vector comercial pET29-a(+) prèviament digerit amb els mateixos enzims de restricció.

Per a l'expressió de la construcció NeqC-(TDS)H6 *E. coli* BL21(DE3) transformada prèviament amb el plasmidi pET-29(+)_NeqC-(TDS)H6 es va fer créixer en 2L de medi LB suplementat amb ampicil·lina (100 µg/ml) a 37°C fins a una OD600 =0.6. L'expressió es va induir amb 0.5 mM IPTG, i es va deixar créixer el cultiu durant 5 h a 37°C. Després de recollir les cèl·lules per centrifugació, es van resuspendre amb 20 mM fosfat a pH 7.2, 500 mM NaCl, 8M urea, 20 mM imidazol i es varen deixar 1 hora a temperatura ambient amb agitació suau. La fracció insoluble es va eliminar per centrifugació a 14000 rpm a temperatura ambient 30 minuts. La fracció soluble es va incubar amb 6 ml de resina Ni-NTA (Qiagen, GmbH Germany) prèviament equilibrada amb de 20 mM fosfat a pH 7.2, 500 mM NaCl, 8M urea, 20 mM imidazol durant 1.5 hores a temperatura ambient sota agitació suau, i es va carregar a la columna. Per a la purificació la columna es va rentar amb 5 VC de 20 mM fosfat a pH 7.2, 500 mM NaCl, 8M urea, 20 mM imidazol, seguidament amb 5 VC de 20 mM fosfat a pH 7.2, 500 mM NaCl, 8M urea, 50 mM imidazol, i finalment la proteïna es va eluir amb 5 VC de 20 mM fosfat a pH 7.2, 500 mM NaCl, 8M urea, 250 mM imidazol en fraccions de 1.5 ml. Tots els passos de la purificació es van seguir per 15%-SDS-PAGE. Les fraccions que contenen la proteïna d'interès es van dialitzar en diferents passos per eliminar gradualment la urea. En un primer pas les fraccions es van dialitzar enfront de 1L de 20 mM fosfat, 150 mM NaCl, 5 M Urea, 1 mM EDTA a pH 8.0, després amb 1.5 L de 20 mM fosfat, 150 mM NaCl, 2M Urea, 1 mM EDTA a pH 8.0 i finalment, es varen realitzar 2 diàlisis enfront de 20 mM fosfat, 150 mM NaCl, 1 mM EDTA a pH 8.0. La mostra pura es va dialitzar en front d'aigua milli-Q, després de descartar per centrifugació la proteïna precipitada es va liofilitzar.

En aquest treball es refereix a aquesta construcció, NeqC(TDS)-H6, com NeqC.

La massa molecular de la *NeqC-(TDS)H6* i la puresa es van confirmar per espectrometria de masses, MALDI-TOF, utilitzant un espectròmetre de masses Bruker-Biflex (Bruker, USA) en el Servei Tècnic de Recerca de la Universitat de Girona (Annex B).

NeqC-(TD)GB1

La seqüència és la corresponent al precursor C. La seqüència de la *NeqC* es troba indicada en negreta i els residus salvatges de l'exteïna C de *N.equitans* es troben subratllats.

MRYLGKKRVILYDLSTESGKIFYVNGLVLHNTDIGSEFNQYKLIILNGKTLKGETTTTEAVDAAT
AEKVFQYANDNGVDGEWYDDATKTYTVTE

El vector d'expressió pSKBAD2_*NeqC*-(TD)GB1 es va obtenir en diferents passos. El primer pas va ser obtenir el vector pRSET_*NeqC*-(TD)eGFP. Sobre el vector pET-29a(+)_*NeqC*-(TDS)H6 i utilitzant els oligonucleòtids 1 i 2 (Taula 3.1) es va amplificar el gen de la *NeqC*. El gen amplificat i el vector pRSET-eGFP van ser digerits, simultàniament, amb els enzims de restricció *NdeI*/*EcoRI*, seguidament, el producte de la digestió es va lligar, obtenint, pRSET_*NeqC*-(TD)eGFP. En aquest clonatge es perd el residu natiu Ser en posició +3. A continuació, el vector pRSET_*NeqC*-(TD)eGFP va ser digerit amb les endonucleases *NdeI* i *HindIII*. L'insert obtingut, *NeqC*-(TD)eGFP, es va clonar en el vector d'expressió pET-28a(+), el qual prèviament havia sigut digerit amb els mateixos enzims de restricció. Com a resultat d'aquest clonatge es va obtenir el vector d'expressió pET-28a(+)_H6*NeqC*-(TD)eGFP. Tot seguit i per a obtenir la diana *EcoRI* en l'extrem 5' del gen de la GB1, aquest es va amplificar utilitzant els oligonucleòtids 4 i 15 (Taula 3.1) sobre el vector pSKBAD02. El gen amplificat juntament amb el vector d'expressió pET-28a(+)_H6*NeqC*-(TD)eGFP es van digerir amb els enzims de restricció *EcoRI* i *HindIII*. El producte de la digestió es va lligar i com a resultat es va obtenir el vector pET-28a(+)_H6 *NeqC*-(TD)GB1, el qual tenia la GB1 com a exteïna C. Finalment, i per a obtenir el vector pSKBAD02_*NeqC*-(TD)GB1, es va clonar l'insert *NeqC*-(TD)GB1, escindit del vector pET-28a(+)_H6 *NeqC*-(TD)GB1 amb els enzims de restricció *NdeI* i *HindIII*, al plasmidi d'expressió pSKBAD02, el qual també havia sigut prèviament digerit pels mateixos enzims de restricció. Aquest plasmidi s'utilitzà per a obtenir l'expressió del complex *NeqN/NeqC in vivo*. El procediment d'expressió i purificació es detallarà més endavant.

NeqC-(TD)GB1H6

La seqüència és la corresponent al precursor C. La seqüència de la *NeqC* es troba indicada en negreta i els residus salvatges de l'exteïna C de *N.equitans* es troben subratllats.

MRYLGKKRVILYDLSTESGKFYVNGLVLHNTDIGSEFNGTQYKLIILNGKTLKGETTTEAVDAAT
AEKVFKQYANDNGVDGEWTYDDATKTYTVTEKLAAALEHHHHHH

Per a la construcció del plasmidi pET-22b(+)_NeqC-(TD)GB1H6, es va realitzar un clonatge i una mutagènesi dirigida. Per començar, es van digerir paral·lelament els vectors pET-22b(+) i pET28(+)-H₆-NeqC-GB1 (descriu anteriorment), amb els enzims de restricció *Nde*I i *Xho*I. Tot seguit, després de la digestió, l'insert i el vector es van lligar obtenint com a resultat el plasmidi pET-22b(+)_NeqC-(TD)GB1. Per mutagènesi dirigida amb els oligonucleòtids 13 i 14 (Taula 3.1) es va eliminar el codó stop en l'extrem 3' de la GB1.

Per la producció de NeqC-(TD)GB1H6 cèl·lules d'*E. coli* BL21(DE3) van ser transformades amb el plasmidi pET-22b(+)_NeqC-(TD)GB1H6 i es van fer créixer en 200 ml de medi LB amb ampicil·lina (100 µg/ml) a 37°C fins a una OD₆₀₀ = 0.4. Arribada a aquesta densitat de cultiu la temperatura es va baixar a 18°C i l'expressió es va induir afegint 0.5 mM IPTG durant tota la nit a 18°C. NeqC-(TD)GB1H6 es va purificar igual que H6MBP(GD)-NeqN (Annex A).

NeqC-(TD)ZHER2

La seqüència és la corresponent al precursor C. La seqüència de la NeqC es troba indicada en negreta i els residus salvatges de l'exteïna C de *N. equitans* es troben subratllats.

MRYLGKKRVILYDLSTESGKFYVNGLVLHNTDIGSEFVDNKFNKEMRNAYWEIALLPNLTNQOK
RAFIRKLYDDPSQSANLLAEAKKLNDQAQPK

Per a obtenir el vector d'expressió pSKBAD2_NeqC-(TD)ZHER2 el plasmidi pIDTSMART-AMP:NEQsc es va digerir amb els enzims de restricció *Eco*RI i *Hind*III, dianes dels quals es troben flanquejant el gen ZHER2. L'insert obtingut es va lligar en el vector pSKBAD2_NeqC-(TD)GB1, prèviament digerit amb els mateixos enzims. Aquest plasmidi s'utilitza per a obtenir l'expressió del complex NeqN/NeqC *in vivo*. El procediment d'expressió i purificació es detallarà més endavant.

H6GB1(D)-NEQsc-(TDS)ZHER2

La seqüència és la corresponent al precursor en cis, on les dues inteïnes es troben unides covalentment. Les seqüències de les inteïnes es troben indicades en negreta i els residus salvatges de l'exteïnes N i C de *N.equitans* es troben subratllats.

MGSSHHHHHSSGLVPRGSHMQYKLIILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWT
 YDDATKTYTVTEGSGYDSIMDTEIEVIENGIKKKEKLSDLFNKYYAGFQIGEKHYAFPPDLYVY
DGERWVKVYSIIKHETETDLYEINGITLSANHLVLSKGNWVKAKEYENKNNENLYFQGRYLGKK
RVILYDLSTESGKFYVNGLVLHNTDSGSEFVDNKFNKEMRNAYWEIALLPNLTNQKRAFIRKL
 YDDPSQSANLLAEAKKLNDAQAPK

Per a crear el plasmidi pSKDuet01-H6GB1(D)-NEQsc-(TDS)ZHER2, el vector de clonatge pIDTSMART-AMP:NEQsc es va digerir amb els enzims de restricció *Bam*HI i *Hind*III, les dianes dels quals flanquegen el gen que codifica per la NEQSC-(TDS)ZHER2. El vector d'expressió pSKDuet01 també es va digerir amb els mateixos enzims. Finalment, el vector i l'insert digerit es van lligar amb T4 lligasa.

Per a la producció de H6GB1(D)-NEQsc-(TDS)ZHER2 cèl·lules d'*E. coli* BL21(DE3) transformades amb pSKDuet01-H6GB1(D)-NEQsc-(TDS)ZHER2 es van fer créixer en 200 ml de medi LB amb ampicil·lina (100 µg/ml) a 37°C fins a una OD600 = 0.6. L'expressió es va induir afegint 0.5 mM IPTG durant 5 hores a 37°C. H6GB1(D)-NEQsc-(TDS)ZHER2 es va purificar igual que H6MBP(GD)-NeqN (Annex A). I el mateix que el precursor H6MBP(GD)-NeqN no es va poder determinar la massa per MALDI-TOF perquè no s'ionitzava amb prou eficiència.

Formació del complex in vivo

Per a la formació del complex NeqN/NeqC *in vivo* es va utilitzar el sistema dual d'expressió pSKDuet01/pSKBAD2. El vector pSKDuet01 codifica pel precursor N i pSKBAD2 pel precursor C. En aquesta tesi s'han realitzat tres tipus de coexpressions diferents:

- 1- Precursor N: H6MBP(GD)-NeqN; Precursor C: NeqC-(TD)GB1.
- 2- Precursor N: H6GB1(wt)-NeqN; Precursor C: NeqC-(TD)ZHER2.
- 3- Precursor N: H6GB1(wt)-NeqN S1C; Precursor C: NeqC-(TD)ZHER2.

A continuació, es detalla com es va realitzar aquesta coexpressió i la seva copurificació per cromatografia d'afinitat Ni-NTA. Totes les construccions segueixen el mateix protocol, degut a la baixa solubilitat del precursor C, primer s'indueix l'expressió del precursor N.

Per a la coexpressió del precursor N i C en cèl·lules d'*E. coli* BL21(DE3) en primer lloc aquestes es van cotransformar amb els plasmidis d'interès. Un cop transformades es van fer créixer en 200 ml de medi LB amb kanamicina (50 µg/ml) i ampicil·lina (100 µg/ml) a 37°C fins a una OD₆₀₀ = 0.6. L'expressió del precursor N es va induir afegint 0.5 mM IPTG, després de 30 minuts, es va induir l'expressió del precursor C amb 0.2% L-arabinosa. El cultiu es va deixar durant 6 hores a 37°C. La purificació del complex es va realitzar igual que la del precursor N, H6MBP-NeqN, expressat individualment (Annex A).

3.3.3. Producció i purificació de la Proteasa H6-TEV

Un litre de cultiu de cèl·lules d'*E. coli* BL21(DE3) transformades prèviament amb el plasmidi que conté la proteasa TEV (subministrat per G. Stier, EMBL, University of Heidelberg, Germany), es van fer créixer a 37°C fins a una OD₆₀₀ de 0.4. En aquest punt es va baixar la temperatura a 18°C, es va induir l'expressió afegint 1 mM IPTG i es van deixar tota la nit. Després de la producció, les cèl·lules es van recollir per centrifugació i es van resuspendre amb 30 ml de tampó de lisi (25 mM Tris-HCl, 150 mM NaCl, 1 mM PMSF, 5 mM Imidazol a pH 8.0). Seguidament, es van lisar passant-les per la premsa de French. Es va recuperar el sobrenadant i es va incubar amb 10 ml d'agarosa Ni-NTA, prèviament equilibrada amb tampó de lisi, durant dos hores a temperatura ambient amb agitació forta. A continuació, la reïna junt amb el sobrenadant es van carregar dintre de la columna, i aquesta es va rentar amb 5 VC de tampó de lisi, i 5 VC de tampó de lisi amb 50 mM imidazol. Per a la seva elució es van passar 5 volums de tampó de lisi amb 250 mM d'imidazol i 10% glicerol. La fracció amb la proteïna pura es va dialitzar enfront de 25 mM Tris-HCl a pH 8.0, 150 mM NaCl i 10% Glicerol. Finalment, la concentració de proteïna es va ajustar a una concentració final de 2 mg/ml i es va augmentar al 50% la concentració de glicerol per a guardar les alíquotes a -80°C.

Taula 3.2. Resum de les construccions utilitzades en aquest treball.

Precursor	Construcció	Vector	Funció
N	T7-NeqN	pET-21a(+)	Caracterització de la inteïna N de <i>N. equitans</i>
	TrxH6TEVT7-NeqN	pBADM-20	
	H6TEV(wt)-NeqN	pET-22b(+)	
	H6MBP(GD)-NeqN	pSKDuet01	Associació <i>in vivo</i> + <i>Trans-splicing in vitro</i>
	H6GB1(wt)-NeqN		
H6GB1(wt)-NeqN(S1C)			
C	NeqC-(TDS)H6	pET-29(+)	Caracterització del procés d'associació Caracterització del complex NeqN/NeqC
	NeqC-(TD)GB1H6	pET-22b(+)	Associació + <i>Trans-splicing in vitro</i>
	NeqC-(TD)GB1	pSKBAD02	Associació <i>in vivo</i> + <i>Trans-splicing in vitro</i>
	NeqC-(TD)ZHER2		
NEQ – única cadena polipeptídica	H6GB1(D)-NEQ _{sc} -(TDS)ZHER2	pSKDuet01	<i>Splicing</i>

3.4. Anàlisi i caracterització proteica

3.4.1. Anàlisi seqüencial de la NeqN i NpuN

Els punts isoelèctrics teòrics i les masses moleculars teòriques de cada precursor van ser calculats utilitzant l'eina ExPASy ProtParam [129]. L'alineament seqüencial de la Neq pol amb les inteïnes partides Ssp DnaE i Npu DnaE es va realitzar utilitzant el servidor T-coffee [130]. La composició aminoacídica de la NeqN i de la NpuN es va analitzar utilitzant el software Composition profiler [131], que inclou les bases de dades Disprot 3.4 (Base de dades de proteïnes intrínsecament desestructurades, [132] i PDB_Select_25. Per a la predicció de l'estat desordenat de la NeqN i la NpuN es va utilitzar l'anàlisi de la funció de distribució acumulada (CDF) del software PONDR[®], (VL-XT) [133]. L'anàlisi de la funció de distribució acumulada (CDF) prediu una puntuació ordenat-desordenat per cada un dels residus de la proteïna. Aquesta puntuació s'ha establert determinant les característiques relatives de tres grups de proteïnes (proteïnes totalment desplegadas, totalment plegades i proteïnes plegades amb regions desordenades). D'acord amb l'anàlisi CDF, proteïnes amb corbes per sota del límit donat (és a dir un percentatge alt de residus tenen una puntuació desordenada prevista elevada) es prediu que seran desordenades, i pel contrari, proteïnes predites com ordenades presentaran corbes per sobre del límit donat. En el gràfic càrrega-hidrofobicitat, es representa la càrrega neta mitjana enfront de la hidrofobicitat normalitzada en l'escala Kyte-Doolittle. La càrrega neta mitjana (R) és el valor absolut de la diferència entre el nombre de residus Arg/Lys i Asp/Glu en la seqüència dividit entre el nombre de residus totals. Els valors d'hidrofobicitat es van calcular amb una finestra de 5 residus i una escala normalitzada de 0 a 1 de Kyte-Doolittle, on 0 és molt polar i 1 molt hidrofòbic. La hidrofobicitat mitjana (H) és el resultat del valor d'hidrofobicitat de cada residu dividit entre el nombre de residus totals. La recta que separa empíricament les proteïnes plegades de les desplegadas, està definida per l'equació $R = 2.785 H - 1.151$ [134]. Les proteïnes ordenades presentaran valors de càrrega-hidrofobicitat que es trobaran per sota de la recta, al contrari, les proteïnes desordenades tindran valors per sobre de la recta.

3.4.2. Bradford

Per a la quantificació de proteïna total es va utilitzar el mètode Bradford. Aquest mètode es basa en el desplaçament del màxim d'absorció de 465 nm a 595 nm del colorant Comassie Brilliant Blue G-250 (Bio-Rad, USA) en solucions àcides quan s'uneix a la proteïna. La recta patró es prepara amb BSA (Sigma, USA) en concentracions que van dels 2.5 als 20 µg/ml.

3.4.3. Electroforesi discontinua en gel d'acrilamida amb SDS (SDS-PAGE)

Aquest mètode analític ha sigut utilitzat de manera rutinària per la determinació del contingut proteic en cada una de les fraccions cel·lulars, el seguiment dels processos de purificació, determinació del grau de puresa, i per al seguiment de les reaccions de proteòlisi i *trans-splicing* estudiades en aquesta tesi. Per a realitzar el SDS-PAGE es va utilitzar el sistema discontinu de Laemmli [135] en gels d'acrilamida. El percentatge d'acrilamida variava entre un 12% i un 15% en funció de la construcció a analitzar. Les mostres es van preparar, generalment, aplicant tampó de càrrega (X4) (50 mM Tris-HCl, 20% glicerol, 8% SDS, 0.02% bromofenol) en una relació 3:1, i es van bullir durant 10 minuts a 95°C. Per al seguiment de la reacció de *trans-splicing* catalitzada pel precursor N, H6GB1(wt)-NeqN, les mostres es van preparar amb tampó de càrrega amb urea (X2) (50 mM Tris-HCl, 1.6% SDS, 7% glicerol, 8M urea) i es van bullir a 95°C durant 1 hora. Les bandes de proteïna i els marcadors es van tenyir amb una solució de Blau de Comassie.

3.4.4. Anàlisi de solubilitat a 37°C

Cèl·lules d'*E. coli* BL21(DE3) transformades amb els vectors recombinants d'interès es van fer créixer en 10 ml de medi LB suplementat amb l'antibiòtic corresponent fins a una OD600 = 0.6 i es va induir l'expressió amb 0.5 mM IPTG. Les cèl·lules es van recollir per centrifugació (15000 rpm, 1 minut) i es van resuspendre amb tampó de lisi (20 mM fosfat a pH 8.0, 0.5 M NaCl, 1 mM EDTA). Els volums de tampó de lisi amb el qual es resuspenen les cèl·lules s'ajusta a la densitat del cultiu, per a carregar en el gel la mateixa quantitat de proteïna. Les cèl·lules es van lisar realitzant 3 cicles de 30 segons de sonicació. A continuació se separa la fracció soluble de la insoluble per centrifugació 20 minuts a 4°C, 15000 rpm. El precipitat, *fracció insoluble*, es resuspèn amb el mateix volum de tampó de lisi que es té en el sobrenedant. Les fraccions obtingudes s'analitzen en un SDS-PAGE.

3.4.5. Estudi del procés de *trans-splicing*

Totes les reaccions de *trans-splicing* i *cis-splicing* realitzades al llarg d'aquesta tesi, s'han fet amb precursors purificats prèviament per Ni-NTA a pressió atmosfèrica, i han sigut analitzades utilitzant gels d'acrilamida.

En els casos que va ser possible es van identificar bandes obtingudes després de la reacció de splicing amb Western-Blot, utilitzant anticossos AntiHis , i MALDI-TOF.

3.4.6. Western-Blot

El protocol utilitzat per a la realització d'un Western-Blot és el següent:

En primer lloc es separen les proteïnes per mida molecular utilitzant SDS-PAGE, la concentració d'acrilamida variarà en funció de la mida de les bandes a separar. A continuació, s'activa la membrana de PVDF (Millipore, USA) submergint-la amb metanol 100% durant 15 segons i després rentant-la amb H₂O Milli-Q durant 2 minuts. Un cop activada es submergeix juntament amb les esponges en tampó d'electrotransferència (25 mM Tris, 192 mM glicina, 20% metanol (v/v) a pH 8.3) mínim 5 minuts. Simultàniament, es submergeixen els papers de filtre durant 30 segons. Després es realitza l'electrotransferència a 100 V durant 2 hores a 4°C i amb agitació suau. Un cop finalitzada l'electrotransferència la membrana PVDF es submergeix en tampó TBST (10 mM Tris-HCl, 100 mM NaCl, 0.1% Tween, a pH 7.5) durant 5 minuts a temperatura ambient amb agitació suau. A continuació es realitza el bloqueig de la membrana, afegint una solució 1% BSA en TBST, i deixant-la durant mínim 1 hora a temperatura ambient i amb agitació. Després d'eliminar la solució de bloqueig s'afegeixen 10 ml d'una solució d'anticòs primari Tetra-His Antibody mouse IgG (Qiagen, Alemanya) i s'incuba durant tota la nit a 4°C. La solució d'anticòs Anti-His es prepara realitzant una dilució 1:6000 en tampó de bloqueig. Posteriorment, es renta la membrana 3 vegades amb TBST, per a eliminar l'excés d'anticòs primari, i s'afegeix l'anticòs secundari Goat Anti-Mouse IgG Peroxidase Conjugate té (Calbiochem, Alemanya) en tampó de bloqueig a una dilució 1:20000. Es deixa incubar 1 hora ambient i agitació suau.

Per al revelat del Western-Blot, es renta la membrana amb TBST 3 vegades durant 5 minuts a temperatura ambient i amb agitació, i s'afegeix la solució de revelat (2 ml de la solució A + 2 ml de solució B) del kit quimioluminiscent Immobilon™ Western (Millipore, EUA). Abans de llegir a 425 nm en el quimioluminòmetre FluorChem™ SP (Alpha Innotech, EUA)s'elimina l'excés de solució de revelat.

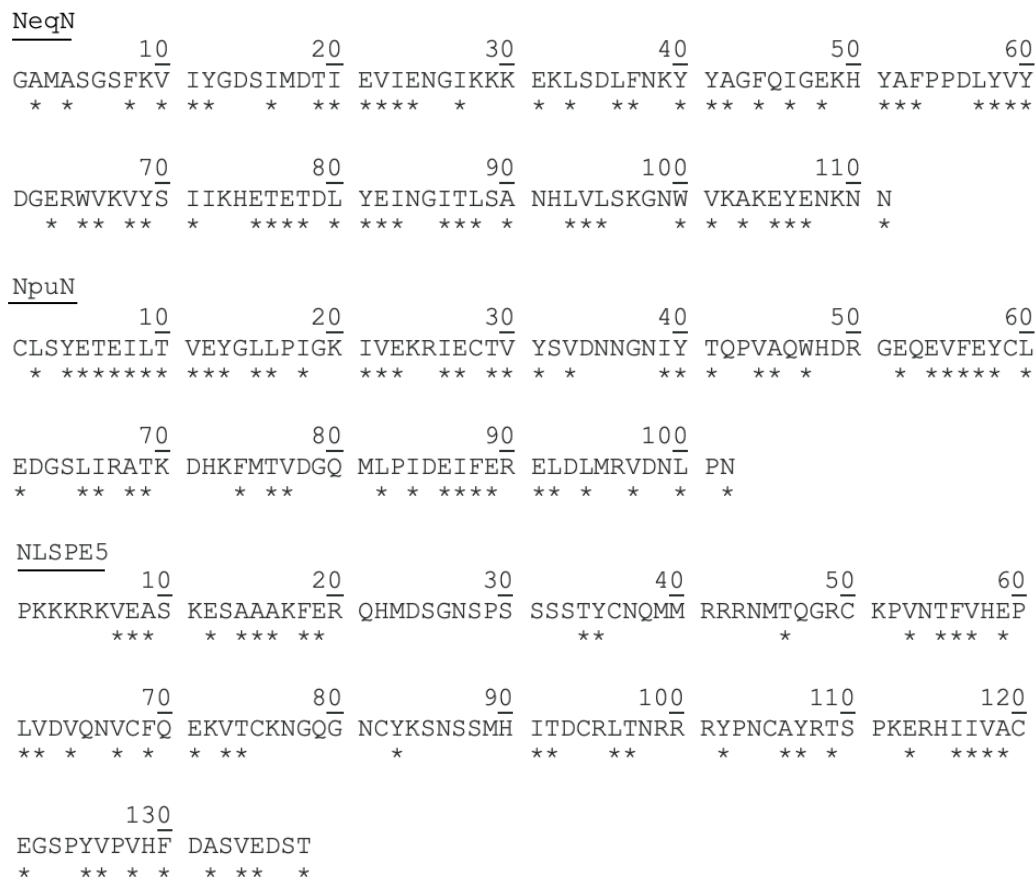
3.4.7. Proteòlisi limitada

La proteïnasa K (EC 3.4.21.14) és una serin-proteasa que talla específicament seqüències X-Y, on X és un residu alifàtic, aromàtic o hidrofòbic, i Y és qualsevol aminoàcid [136]–[138]. És un enzim molt robust, ja que es manté actiu en un ampli rang de pH (7.5-12.0, pH òptim 8.0) i de temperatures (20–60°C), sent la seva temperatura òptima 37°C. També conserva la seva activitat proteolítica en presència d'agents desnaturalitzants com SDS (0.5-2%) i urea (fins a 4M). La proteïnasa K és inactivada a temperatures >65°C, amb àcid tricloroacètic i inhibidors de serin-proteases, com el PMSF. El calci és important per a la seva termoestabilitat, però no és indispensable per a la seva catàlisi, a més a més aquest enzim és actiu amb tampons que continguin agents quelants com ara l'EDTA.

Les proteïnes purificades a una concentració de 10 µM amb 50 mM tampó fosfat, 150 mM NaCl a pH 8.0 es van tractar amb 10 µg/ml de proteïnasa k i es van incubar durant 1 hora a temperatura ambient. A diferents temps (0', 10', 1h) es van extraure alíquotes, i la reacció es va aturar afegint 1.5 µM de PMSF (Sigma-Aldrich, Germany). La cinètica de la reacció es va analitzar per 15% SDS- PAGE. Per a la validació del mètode es va utilitzar com a control positiu la inteïna NpuN aïllada, la qual no presenta estructura tridimensional definida [57], i com a control negatiu, NLS_PE5, variant de ribonucleasa pancreàtica humana, la qual presenta una estructura terciària compacta [139]. La preparació dels controls és la mateixa que per a les mostres.

Per a predir quina regió de la seqüència és resistent a la proteasa, el producte de la digestió després d'1 hora es va analitzar per espectrometria de masses MALDI-TOF (Serveis Tècnics de Recerca, Universitat de Girona).

A continuació, es mostra la seqüència en format FASTA de les proteïnes utilitzades per l'anàlisi on s'indica amb un asterisc els llocs de tall predits per la proteïnasa K (https://web.expasy.org/peptide_cutter).



L'estabilitat termodinàmica de la NeqN purificada es va analitzar per proteòlisi limitada en presència de concentracions creixents d'urea [140]. Per a realitzar aquest assaig es va utilitzar una solució de NeqN purificada a una concentració de 10 µM preparada amb 50 mM tampó fosfat, 150 mM NaCl a pH 8.0. Les solucions de proteïna es van equilibrar tota la nit a 25°C amb diferents concentracions d'urea (0M, 1M, 2M, 3M i 4M). Les mostres equilibrades es van tractar amb 10 µg/ml de proteïnasa K i es van deixar 1 hora a temperatura ambient. La reacció es va aturar afegint PMSF. La proteòlisi dependent d'urea de la inteïna NeqN es va analitzar per 15%-SDS-PAGE.

3.4.8. Determinació de l'estabilitat tèrmica per UV

Per a determinar l'estabilitat conformacional de la construcció NeqN, es va monitoritzar l'absorbància a 278 nm en un rang de temperatura entre 20°C-80°C, utilitzant un espectrofotòmetre Perkin-Elmer LS-50B (Perkin Elmer, USA [141]. L'augment de la temperatura es va realitzar amb increments de 2°C, amb un temps d'equilibrat posterior de 15 minuts. Al contrari, la monitorització del replegament es va seguir de la mateixa manera disminuint la temperatura amb decrements de 5°C amb un temps d'equilibrat de 10 minuts. Tots els

experiments es van realitzar amb una cubeta de quars a una concentració de proteïna de 20 μM preparada amb 50 mM citrat, 50 mM NaCl a pH 4.5 i 50 mM fosfat, 50 mM NaCl a pH 7.0.

3.4.9. Determinació de l'estabilitat tèrmica per Calorimetria diferencial d'escombrat (DSC)

Els experiments per a determinar l'estabilitat tèrmica per DSC es van realitzar en un microcalorímetre equipat amb cel·les capil·lars VP-DSC (MicroCal, Northampton, MA, USA). El rang de temperatura emprat va ser de 5°C-125°C amb un increment de la temperatura de 1.5°C/min. Es va treballar a pressió constant, 60 psi. La reversibilitat del desplegament tèrmic sempre es comprovava amb un segon escombrat consecutiu de la mateixa mostra. Per a equilibrar adequadament l'aparell abans de qualsevol anàlisi es realitzaven diferents línies base amb el tampó en el qual es trobava dissolt la proteïna d'estudi. La línia base s'extreia del termograma de la mostra i el temps de resposta del calorímetre es corregia. Les corbes de capacitat de calor molar parcial (C_p) es van calcular i analitzar utilitzant Origin 7.0 (OriginLab, Northampton, MA, USA).

Per l'estudi de l'estabilitat tèrmica de la NeqN i del complex NeqN/NeqC, es van preparar dues solucions de les proteïnes prèviament purificades a una concentració final de 26 μM i 16 μM , respectivament, amb 20 mM fosfat, 150 mM NaCl, 1 mM EDTA a pH 8.0.

3.4.10. Caracterització per dicroisme Circular (CD)

Les mesures de CD es van realitzar amb un espectropolarímetre Jasco J-715 (Tokyo, Japó) equipat amb un mòdul de control de temperatura. Tots els espectres es van registrar a 25°C utilitzant cubetes de quars de 0.1 cm de pas òptic. Els espectres de CD proper es van registrar en un rang de $\lambda = 250\text{-}350$ nm. Les dades són la mitjana de 20 repeticions utilitzant una amplada de banda de 1 nm i un temps de resposta de 2 segons. Les mostres, NeqN, NeqC i el complex, es varen preparar amb 20 mM tampó fosfat, 150 mM NaCl, 1 mM EDTA a pH 8.0, a una concentració de 1.6 $\mu\text{g}/\mu\text{l}$, 0.54 $\mu\text{g}/\mu\text{l}$ i 2,4 $\mu\text{g}/\mu\text{l}$. Els espectres de CD llunyà es van registrar en un rang de $\lambda = 203\text{-}260$ nm. Les dades són la mitjana de 5 repeticions utilitzant una amplada de banda de 1 nm i un temps de resposta de 1 segons. Les mostres, NeqN, NeqC i el complex, es varen preparar amb 20 mM tampó fosfat, 150 mM NaCl, 1 mM EDTA a pH 8.0, a una concentració de 0.22 $\mu\text{g}/\mu\text{l}$, 0.54 $\mu\text{g}/\mu\text{l}$ i 0.38 $\mu\text{g}/\mu\text{l}$. Els espectres a longituds d'ona més baixes

no es van poder registrar a les concentracions assajades a causa de la saturació del senyal. Els estudis de CD es van realitzar en el Departament de Química Física de la Universitat de Granada.

3.4.11. Assaig d'unió a ANS

L'espectre d'emissió de fluorescència de l'ANS amb proteïna i sense es va registrar a temperatura constant, 30°C, en un rang de 400-750 nm (amplada de finestra de 10 nm) amb una $\lambda_{exc}=370$ nm (amplada de finestra de 5 nm) en el fluorímetre LS-50B (Perking-Elmer, Waltham, MA, EUA). Es va preparar una solució de NeqN prèviament purificada a una concentració final de 25 μ M amb 50 mM fosfat, 50 mM NaCl a pH 6.9. Paral·lelament, es va preparar una solució d'ANS amb el mateix tampó emprat per a la proteïna, es va filtrar amb filtres de 0.22 μ m i es va afegir a la solució proteica amb una concentració final de 25 μ M i es va incubar durant 5 minuts a temperatura ambient abans de cada lectura. Abans de realitzar les lectures la solució de treball es va deixar durant 5 minuts. Com a control positiu, es va utilitzar l' Apop Δ Pro [142]–[144].

3.4.12. Cromatografia d'exclusió molecular (SEC)

Per dur a terme aquest estudi s'ha utilitzat una columna S75 10/300 GL (GE Healthcare, USA) a temperatura ambient, equilibrada amb el tampó en el qual es trobi la proteïna d'estudi, generalment es tracta d'un tampó 20 mM Tris-HCl/fosfat, 150 mM NaCl, 1 mM EDTA a pH 8.0. Tots els cromatogrames s'han monitoritzat utilitzant una longitud d'ona de 280 nm. Les condicions cromatogràfiques utilitzades tant per a mostres com per a patrons, són: 500 μ l d'injecció, flux de 0.5 ml/min i gradient isocràtic.

Per a determinar la massa molecular es va calibrar la columna amb una sèrie de proteïnes globulars amb massa molecular coneguda (GE Healthcare): Aprotinina (6500 Da); Ribonucleasa A (13700 Da); Anhidrasa Carbònica (29000 Da) i Ovoalbúmina (44000 Da). Tant proteïnes com patrons es van preparar amb 20 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA a pH 8.0. Es va analitzar una solució de NeqN, una de NeqC i una altra del complex NeqN/NeqC, les quals es trobaven a una concentració de 325 μ M, 76 μ M i 356 μ M, respectivament. Els patrons es van preparar a una concentració final de 3 mg/ml. Per a establir la relació quantitativa entre el temps d'elució i la massa molecular es va calcular una equació de regressió lineal ($r^2= 0.9959$). Aquesta anàlisi es va realitzar a l'Institut de Biologia Molecular de Barcelona (IBMB-CSIC).

3.4.13. Ultracentrifugació analítica

La massa molecular de la NeqN i del complex NeqN/NeqC-(TDS)H6 es va determinar per ultracentrifugació analítica amb col·laboració de l'Institut de Química-Físic Rocasolano (CSIC) de Madrid. Els assajos de velocitat de sedimentació es van realitzar utilitzant una ultracentrífuga analítica amb sistema òptic Optima XL-A i un sistema òptic d'interferència Optima XL-I (Beckman Coulter) per a la detecció de gradients de concentració. Els experiments es van realitzar amb 20 mM tampó fosfat, 150 mM NaCl, 1 mM EDTA a 45000 rpm i 20°C, monitoritzant l'absorbància a 297 nm. La NeqN es trobava a una concentració final de 2.3 µg/µl i el complex a una concentració de 4.61 µg/µl.

3.4.14. Caracterització fisicoquímica del procés d'associació entre els dos fragments de la Neq pol

La caracterització del procés d'associació entre la NeqN i NeqC es va seguir en un espectrofluorímetre Varian Cary Eclipse (Agilent Technologies, Santa Clara, CA) equipat amb una cel·la termoregulada Peltier. Tots els espectres d'emissió de fluorescència es van registrar a $\lambda_{ex}=298$ nm amb una amplada de finestra d'excitació i d'emissió de 5 nm i una $\lambda_{em} = 307-500$ nm. Les mesures es van realitzar amb una cubeta de quars amb 1 mm de pas de llum. Tant les solucions mares de la (NeqN i de la NeqC, com les mostres de fluorescència es van preparar amb 20 mM fosfat, 150 mM NaCl i 1 mM EDTA a pH 8.0. Aquest assaig es va realitzar en el Departament de Química Física de la Universitat de Granada.

Determinació de la cinètica de la reacció

Per a determinar la velocitat de la reacció es van registrar espectres d'emissió de fluorescència de la NeqN en presència de NeqC a diferents temps per cada una de les temperatures assajades (17°C, 36°C, 45°C, 55°C, 64°C i 74°C) (Annex I). Tots els espectres es van mesurar a una concentració final de 20 µM i de 30 µM per la NEQN i la NeqC, respectivament. Es va treballar amb un excés de NeqC, a causa de la seva tendència a precipitar. Abans de registrar cada espectre la NeqN s'incubava a la temperatura assajada durant 10 minuts, després s'afegia la NeqC, s'homogeneïtzava correctament, i finalment, s'iniciava la lectura de l'espectre. Les corbes d'unió es van monitoritzar seguint la disminució al llarg del temps de la fluorescència intrínseca dels triptòfans de la NeqN en presència de la NeqC (Annex I). El temps en el qual es registraven espectres variava en funció de la temperatura i depenia del temps que tardava la reacció a

arribar a l'equilibri. Abans de realitzar l'assaig es va mesurar el soroll de fons que podia introduir la NeqC. Per a demostrar que la disminució de la fluorescència depenia només del procés d'unió entre els dos fragments de la Neq pol es va mesurar la fluorescència de la construcció NeqN sola en les mateixes condicions assajades.

Per a determinar la velocitat de la reacció en cada una de les temperatures, les corbes d'unió obtingudes es van ajustar a una equació exponencial utilitzant el programa SigmaPlot:

$$A = A_0 e^{-kt}$$

on A és l'àrea, A_0 és l'ordenada a l'origen, -k és el pendent.

Amb la finalitat d'obtenir l'energia d'activació de la reacció (E_a) es va representar el $\ln k$ enfront de l'invers de cadascuna de les temperatures, i la recta obtinguda es va ajustar a l'equació d'Arrhenius (Annex I):

$$\ln k = \ln A - E_a/RT$$

k és la velocitat de la reacció, A és el factor de freqüència (s^{-1}), E_a és l'energia d'activació ($kJ mol^{-1}$), R= la constant de gasos i T la temperatura. L'estudi de la cinètica de la reacció d'associació es va realitzar en el Departament de Química Física de la Universitat de Granada.

Determinació de la termodinàmica de la reacció

Per a calcular la constant d'afinitat, K_d , i els paràmetres termodinàmics del procés d'unió entre la NeqN i la NeqC, es van registrar diferents espectres d'emissió de fluorescència a concentracions creixents de NeqC en un rang de temperatures entre 20-70°C, un cop la reacció havia assolit l'equilibri (Annex J). Per cada punt de la titulació, es va afegir a la solució NeqN un volum determinat de NeqC, es va barrejar rigorosament i es va incubar a la temperatura desitjada durant el temps necessari per arribar a l'equilibri, després es van registrar els espectres. El temps per assolir l'equilibri es va establir amb l'estudi cinètic. Les corbes d'unió es van mesurar monitoritzant la disminució de la fluorescència intrínseca dels triptòfans de la NeqN en presència de concentracions creixents de NeqC (Annex J). Per a les diferents titulacions es va treballar amb tres concentracions de NeqN: 24 μM , 20 μM i 14 μM i un rang de concentracions de NeqC entre 5 – 61 μM .

Per a calcular la constant de dissociació de la reacció, cada corba d'unió es va ajustar a la següent equació:

$$F_{obs} = F_{NeqN} + (F_{sat} - F_{NeqN}) \frac{([NeqC] + K_d + [NeqN]) - \sqrt{([NeqC] + K_d + [NeqN])^2 - 4[NeqC][NeqN]}}{2[NeqN]}$$

on $[NeqC]$ és la concentració de la inteïna C en μM , F_{obs} és la intensitat de la fluorescència, F_{sat} és la intensitat de fluorescència en la saturació, F_{NeqN} és la fluorescència de la NeqN, K_d és la constant d'afinitat en μM , i $[NeqN]$ és la concentració d'inteïna N en μM .

El càlcul de la K_d per cada temperatura va permetre mesurar les constants termodinàmiques de la reacció utilitzant l'equació de van't Hoff (Annex J):

$$\frac{d \ln K_d}{d T} = \frac{\Delta H^\circ}{RT}$$

Un cop coneixent ΔH° els paràmetres de ΔG i ΔS es calculen a partir de la següent equació:

$$\Delta G^\circ = \Delta H^\circ + T\Delta S^\circ$$

L'estudi de l'afinitat entre els dos fragments de la inteïna Neq pol es va fer en el Departament de Química Física de la Universitat de Granada.

3.4.15. Formació del complex NeqN/NeqC *in vitro*

Els fragments N-i C-terminals purs, NeqN i NeqC, es van barrejar a una concentració final de 30 μM i 20 μM , respectivament, amb 20 mM tampó fosfat, 150 mM NaCl i 1 mM EDTA a pH 8.0 i es van incubar durant 24 h a 25°C amb agitació suau, 50 rpm.

Per als estudis de cristal·lització i difracció de raig X, com a fragment C-terminal es va utilitzar un pèptid sintètic, 33-mer, el qual contenia la seqüència de la NeqC i els residus TDS de l'exteïna salvatge (NYZTech, Portugal). Per obtenir el complex NeqN/NeqC es va utilitzar el mateix protocol descrit anteriorment però en aquest cas utilitzant tampó Tris-HCl.

3.4.16. Cristal·lització i Difracció de raig X

Aquest apartat de la tesi es va realitzar en col·laboració amb l'Institut de Biologia Molecular de Barcelona (IBMB-CSIC).

Cristal·lització, recollida de dades i processament

Els cristalls de proteïna es van obtenir a partir de cribratges de cristal·lització realitzats amb plaques de 96 pous on es barrejaven 100 nL de proteïna (13 mg/ml) i 100 nL de dissolvent utilitzant un robot Cartesian Technologies (Cartesian Technologies Inc, CA, EUA). Els cristalls de NeqN van aparèixer al cap de dos dies i es van deixar créixer durant una setmana mentre que els cristalls del complex no van parar de créixer fins al cap d'un mes. Els dos cristalls NeqN i el complex van ser optimitzats en plaques de 24 pous per a obtenir grans cristalls que permetessin els experiments de difracció de raig X. Els cristalls de NeqN es van cultivar en 1.1 M de tartrat d'amoni a pH 7.0 formant part del grup espacial monocínic C2. Per al complex es van obtenir dues morfologies diferents: hexagonal i tetragonal amb grups espacials P6522 i P4212, respectivament. Els cristalls hexagonals van créixer en 25% PEG3350, 0.2 M acetat d'amoni i 0.1 M bis-Tris, pH 6.5, mentre que els cristalls tetragonals van créixer al 30% PEG3350 i 0.2 M nitrat d'amoni, pH 6.3. Per a la recollida de dades, els cristalls es van refrigerar ràpidament amb nitrogen líquid amb 30% i 20% de glicerol com a crioprotector per als cristalls de NeqN i del complex, respectivament. Les dades es van mesurar a 100 K a la línia ALBA Xaloc (Universitat Autònoma de Barcelona), en un detector Pilatus 6M, i es van indexar, integrar i escalar mitjançant el paquet XDS [145].

Determinació estructural i refinament

Per resoldre qualsevol dels tres cristalls, a causa del seu baix grau d'homologia amb la resta d'intèines descrites fins ara, es varen realitzar nombrosos intents infructuosos de substitució molecular mitjançant diferents models d'intèines disponibles al PDB. Finalment, el cristall hexagonal del complex NeqN/NeqC es va resoldre *ab initio* amb el programa ARCIMBOLDO_BORGES a una resolució de 1.2 Å mitjançant la biblioteca distribuïda de 7650 models que representen tres fulls beta de 20 residus [146] com a fragments de cerca contra objectius d'intensitat de Phaser [147] i ampliant els fragments col·locats a una solució completa mitjançant la modificació de la densitat iterativa i el seguiment automàtic de restricció beta amb SHELX [148]. Un cop resolta l'estructura del complex hexagonal aquesta va permetre resoldre

l'estructura del cristall tetragonal a 1.7 Å per Reemplaçament Molecular mitjançant el programa PHASER [149]. Tots dos cristalls del complex estaven formats per una sola molècula de NeqN/NeqC a la unitat asimètrica. El cristall monoclínic de NeqN amb quatre subunitats a la unitat asimètrica, es va resoldre mitjançant Reemplaçament Molecular aconseguint una resolució de 2.6 Å amb un model derivat de l'estructura del complex NeqN/NeqC. Alguns residus de la regió C-terminal de la NeqN de les tres estructures estan absents. Aquests residus que falten podrien correspondre a algunes densitats extraininterpretables que es troben en les tres formes cristal·lines. El perfeccionament de l'estructura es va realitzar mitjançant passos iteratius de la construcció de model manuals a Coot [150] i perfeccionament a Refmac5 [151]. El perfeccionament de cristalls monoclínic es va millorar mitjançant restriccions de simetria no-cristal·logràfica (non-crystallographic symmetry, NCS). Per als cristalls hexagonals, es van utilitzar factors de temperatura anisotròpics en els cicles finals de perfeccionament, donant factors d'acord R i RFree de 15.6 i 13.7. Les estructures de cristall monoclínic i tetragonal es van perfeccionar donant factors R i Rfree de 20.7 / 25.7 i 20,4 / 21.7, respectivament (Taula 4.5; Taula 4.9 de Resultats i Discussió). La validació d'estructures es va realitzar amb rutines de MolProbity amb un 99,4%, 100% i 100% de residus a la regió més favorable de la trama Ramachadran per als grups espacials monoclínic, tetraèdric i hexagonal respectivament [152].

4. Resultats i Discussió

4.1. Caracterització de la reacció de *trans-splicing* catalitzada per la inteïna partida Neq pol

Com ja s'ha comentat en la introducció d'aquesta tesi sembla que la reacció de *trans-splicing* catalitzada per la inteïna partida Neq pol es troba regulada per la temperatura, produint-se eficientment dintre del rang de temperatures comprès entre els 50-80°C, amb un màxim d'eficiència a 80°C. En el cas que la regulació de l'activitat de la Neq pol fos causada exclusivament per la inteïna i que l'associació dels dos fragments es donés a temperatures mesofíliques, es podria obtenir de manera estable el complex format pels dos fragments de la inteïna partida sense que es donés la reacció de *trans-splicing*, la qual cosa permetria per primer cop obtenir l'estructura tridimensional dels dos fragments d'una inteïna partida associats de manera no covalent i amb els residus catalítics nadius.

La limitació de les inteïnes per ser utilitzades àmpliament com a eines biotecnològiques, rau generalment, en el fet que la reacció de *splicing* es troba modulada pels residus situats en la proximitat dels llocs d'unió de l'exteïna i la inteïna. Normalment, per aconseguir bons rendiments de *splicing* amb exteïnes exògenes se'ls afegeix una curta seqüència d'aminoàcids (5-10 residus), la qual correspon als residus de la regió adjacent al lloc d'unió exteïna-inteïna presents en la forma natural. Cal considerar, però, que aquesta estratègia pot modificar l'estructura tridimensional de l'exteïna o la seva funció. Com s'ha comentat anteriorment, la tolerància a introduir mutacions en aquests llocs és intrínseca de cada inteïna. Per aquesta raó, actualment, s'inverteixen molts esforços a trobar una inteïna que catalitzi la reacció de *splicing* de manera eficient en un context tant natiu com exogen.

En aquest treball s'ha volgut caracteritzar la reacció de *trans-splicing* catalitzada per la inteïna partida Neq pol utilitzant exteïnes no natives. Per aconseguir aquest objectiu i coneixent que l'eficiència de la reacció es pot veure afectada en funció de l'exteïna utilitzada [55], s'han utilitzat diferents proteïnes com a exteïnes (Taula 4.1).

D'altra banda i a causa del paper modulador descrit dels residus adjacents de la regió d'unió exteïna-inteïna [33]–[37] es va considerar oportú conservar com a mínim dos dels residus de l'exteïna nativa de la DNA polimerasa B de *N. equitans* en els precursors N i C, els quals estan indicats en la Taula 4.1.

Taula 4.1. Resum dels precursors utilitzats per a catalitzar la reacció de splicing. En vermell es troben indicats els residus salvatges de l'exteïna N en la regió adjacent a la Ser1, i en groc es destaquen els residus salvatges que es conserven en les posicions +1 i +2 de l'exteïna C nativa de la inteïna Neq pol.

Precursor N	
Construcció 1n	H6-MBP(GD)-NeqN
Construcció 2n	H6-GB1(FKVIYGD)-NeqN
Construcció 2n	H6-GB1(FKVIYGD)-NeqN (Ser1Cys)
Precursor C	
Construcció 1c	NeqC- (TD) GB1-H6
Construcció 2c	NeqC- (TD) -GB1
Construcció 3c	NeqC- (TDS) -ZHER2
Precursor	
Construcció P	H6GB1(D)-NEQ- (TDS) ZHER2

Característiques de les exteïnes

La proteïna d'unió a maltosa (**MBP**), és àmpliament utilitzada com a proteïna transportadora per a la producció recombinant de proteïnes de fusió en *E. coli*. La presència de la MBP facilita la seva purificació[153], i augmenta la solubilitat de proteïnes amb tendència a agregar[154]. La MBP és una proteïna constituïda per 396 aminoàcids i un pI teòric de 5.5, la qual presenta una temperatura mitjana de desnaturalització (T_m) a pH 8.2 de 62.6°C[153].

La **GB1**, és un domini de la proteïna G de diferents espècies de Streptococcus, constituïda per 56 aminoàcids, amb una elevada estabilitat i solubilitat. S'ha demostrat que aquesta proteïna pot unir-se a altres augmentant la seva estabilitat. Aquesta proteïna presenta un plegament cooperatiu en dos estats i la seva T_m a pH 5.4 és de 87.5°C [155].

El domini **ZHER2**, és un affibody de 58 residus que té una elevada afinitat pel receptor 2 del factor de creixement epidèrmic humà (HER2). Els affibodies són petits i robusts lligands, els quals estan basats en el domini Z de la proteïna A [156]. Aquest domini Z s'ha escollit com a punt de partida per a la construcció d'affibodies, ja que presenta excel·lents propietats biofísiques (elevada T_m, plegament reversible i ràpid i és soluble en solucions aquoses) i presenta alts nivells d'expressió en *E. coli*. Existeix una àmplia llibreria de molècules ZHER2, on cada una de les variants incorpora mutacions que li aporten diferents propietats, en aquesta tesi es va decidir treballar amb la variant N23T, S33K[157]. Aquesta doble mutació augmenta la T_m d'aquesta molècula en 12°C (T_m=77°C), convertint-la en aquest cas en una excel·lent candidata per ser utilitzada com a exteïna per una inteïna provinent d'un organisme hipertermòfil, com és *N.equitans*.

4.1.1. Reacció de *trans-splicing* utilitzant com a exteïna N la MBP i com a exteïna C el domini GB1

Per tal de comprovar que els dos fragments de la Neq pol units a exteïnes no natives podien associar formant el complex NeqN/NeqC sense que es produís la reacció de *trans-splicing* a temperatures mesòfiles, es van coexpressar els dos precursors, N i C, en *E. coli* a 37°C.

En la Figura 4.1 s'esquematitzen les estructures del precursor N, H6MBP(GD)-NeqN, i del precursor C, NeqC-(TD)GB1, respectivament. El precursor N està format per 509 aminoàcids (56492.6 Da), i té un punt isoelèctric teòric de 5.5 i el precursor C està constituït per 95 residus (10626.80 Da) i presenta un pI de 5.0.



Figura 4.1. Esquema del precursor N, H6MBP(GD)-NeqN, i del precursor C, NeqC-(TD)GB1. (a) El precursor N està compost per una cua d'histidines (H6) en l'extrem N-terminal, seguida per la MBP, unida a través d'una curta regió policonnectora de 5 residus a la NeqN. (b) El precursor C està compost per la NeqC unida a través d'un linker d'11 residus a la GB1. En vermell s'assenyalen els residus salvatges que es troben en les posicions -2, -1 i +1 i +2.

Per la coexpressió es va utilitzar el sistema dual psKDuet01/pskBAD2. El precursor N, H6MBP(GD)-NeqN, es va clonar en el plasmidi psKDuet01, i el precursor C, NeqC-(TD)GB1, en el plasmidi psKBAD02. Els dos vectors es van emprar per cotransformar la soca d'*E. coli* BL21(DE3), on l'expressió del precursor N, es va induir amb 0.5M d'IPTG i la del precursor C amb 0.2% de L-Arabinosa. En el gel on s'analitza la coexpressió dels precursors s'observen dues bandes ben definides que corresponen al precursor N i C, i no s'observa, però cap banda amb una massa molecular aproximada a l'esperada pel producte de la reacció ni per cap dels productes de les reaccions secundàries no desitjades d'hidròlisi (Figura 4.2-a). Aquests resultats indiquen que la inteïna Neq pol en un context exogen a 37°C no és activa i per tant, es pot obtenir el precursor intacte.

Per demostrar que els dos fragments de la inteïna interaccionen durant la seva coexpressió en *E. coli* formant correctament el complex NeqN/NeqC, la fracció soluble de l'extracte cel·lular obtingut després de la coexpressió es va carregar en una cromatografia de Ni-NTA i les fraccions recollides es van analitzar per 12% SDS-PAGE. És d'esperar que si el complex s'ha format correctament tant el precursor N com el precursor C, que no te cua d'histidines, coelueixin a elevades concentracions d'imidazol. El primer que crida l'atenció dels resultats obtinguts (Figura

4.2-b) és que a diferència de l'expressió individual del precursor C, on la proteïna s'havia d'expressar a 18°C per a obtenir-la en la fracció soluble, la coexpressió de tots dos precursors a 37°C va permetre obtenir els dos precursors en la fracció soluble. D'altra banda, l'anàlisi de les fraccions obtingudes de la cromatografia d'afinitat indica que les dues bandes corresponents als precursors N i C coelueixen a una concentració de 25 mM d'imidazol. Aquests resultats indiquen que els dos fragments de la inteïna Neq pol interaccionen durant la seva expressió a 37°C en *E. coli* formant el complex NeqN/NeqC inactiu, ja que no s'observa cap banda corresponent al producte de *splicing* ni a cap dels productes de les reaccions d'hidròlisi, i que aquest es manté estable al llarg del procés de purificació. La qual cosa converteix a la inteïna Neq pol en una excel·lent candidata per a posteriors estudis estructurals.

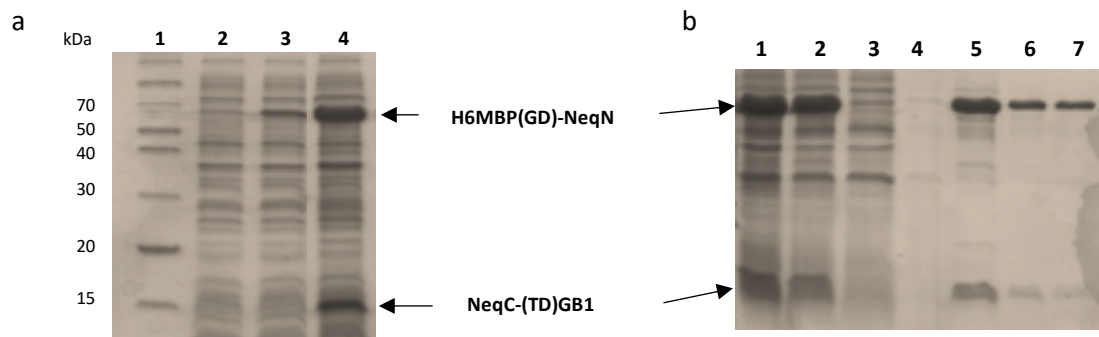


Figura 4.2. Coexpressió recombinant a 37°C i copurificació del complex format per H6MBP(GD)-NeqN i NeqC-(TD)GB1 (56492.6 Da i 10626.8, respectivament). (a) Anàlisi de la coexpressió a 37°C Carril 1: Marcadors de massa molecular; carril 2: -IPTG; carril 3: 30 minuts després de la inducció amb 0.5 mM IPTG, just abans d'induir el precursor C amb 0.2% L-Ara; carril 4: després de 6 hores de producció dels dos precursors. (b) Coexpressió a 37°C i purificació per cromatografia d'afinitat Ni-NTA. Carril 1: fracció total; carril 2: fracció soluble; carril 3: través, carril 4: rentat 1 (20 mM tampó fosfat a pH 7.4, 300 mM NaCl, 10 mM Imidazol); carril 5: rentat 2 (20 mM tampó fosfat a pH 7.4, 300 mM NaCl, 25 mM Imidazol), carril 6: elució 1, carril 7: elució 2. Les elucions es van realitzar utilitzant 20 mM tampó fosfat a pH 7.4, 300 mM NaCl, 250 mM Imidazol.

Per estudiar la reacció de *trans-splicing in vitro* es va partir d'una solució del complex prèviament purificat a una concentració final de 10 µM amb 50 mM tampó fosfat a pH 8.0, 150 mM NaCl, 1 mM EDTA. La mostra es va incubar durant 24 hores a 50°C. Per al seguiment de la reacció es van agafar al·lquotes a diferents temps (0 h, 1.5 h, 4 h, 6 h i 24 h) i es varen analitzar per 12%SDS-PAGE. Es va decidir realitzar la reacció de *trans-splicing* a 50°C, perquè és una temperatura a la qual la inteïna ja és activa [68] i es troba per sota de la T_m descrita per la MBP.

A continuació en la Figura 4.3 s'esquematitzen el producte de la reacció de *trans-splicing* així com els productes de les reaccions d'hidròlisi.

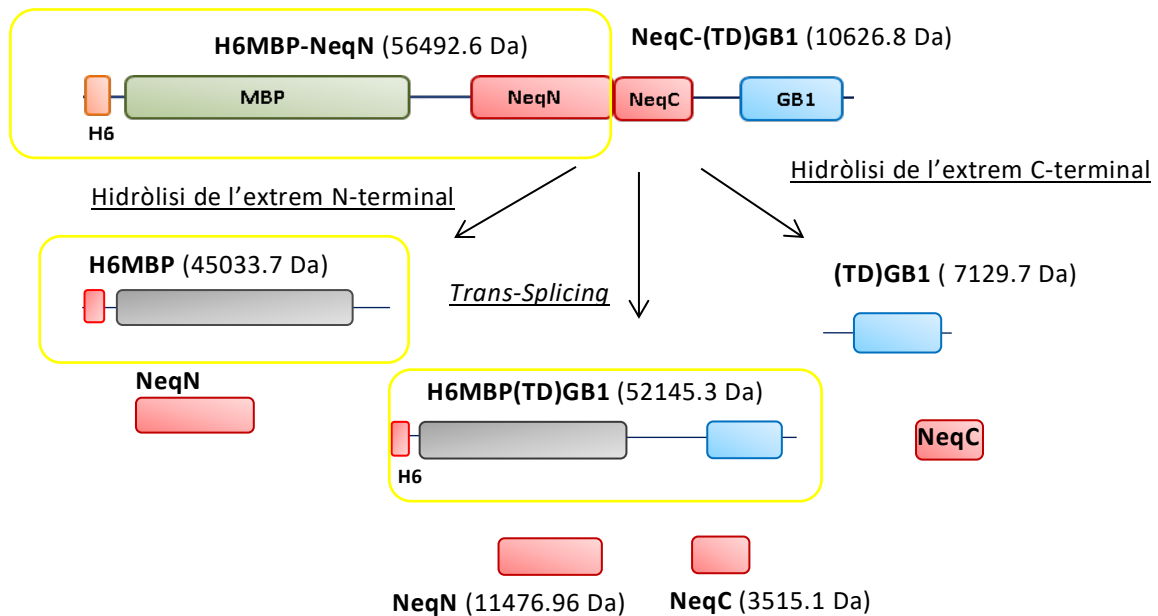


Figura 4.3. Esquema simplificat de la reacció de *trans-splicing* i de les reaccions secundàries d'hidròlisi catalitzada pels precursors H6MBP(GD)-NeqN i NeqC-(TD)GB1. En aquest esquema es representen les molècules que es poden generar amb la incubació dels dos precursors amb les seves respectives masses moleculars. En groc es troben emmarcades les molècules que es poden detectar per anticossos Anti-His.

En el SDS-PAGE on s'analitza la reacció de *trans-splicing* (Figura 4.4), s'observa la disminució gradual del precursor C, NeqC-(TD)GB1, al llarg del temps. La disminució del precursor N, H6MBP(GD)-NeqN, al llarg del temps es fa difícil d'apreciar degut a l'excés de precursor N lliure. Paral·lelament, s'observa la formació de diverses bandes, cap identificada com el producte de la reacció de *trans-splicing*, H6MBP(GD)-(TD)GB1. Les bandes formades al llarg de la reacció corresponen a H6MBP(GD) i (TD)GB1, productes d'hidròlisi de l'extrem N- i C-terminal, respectivament. Aquests resultats indiquen que la ciclació de l'Asn és dona de manera prematura en la inteïna Neq pol amb exteïnes exògenes i que la temperatura no té cap efecte sobre aquesta reacció, ja que s'observa exteïna C hidrolitzada a temps 0. D'altra banda que es produeixi la hidròlisi de l'extrem N-terminal, indica que almenys la inteïna Neq pol es capaç de catalitzar la primera etapa de la reacció de *splicing*. A diferència de la ciclació de l'Asn, sembla que aquesta reacció sí que està regulada per la temperatura, ja que a temps 0 no s'observa cap banda corresponent a la MBP.

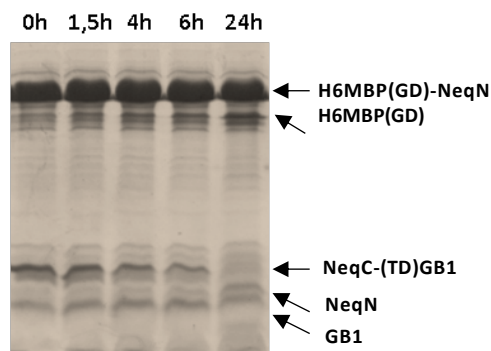


Figura 4.4. Anàlisi per SDS-PAGE de la reacció de *trans-splicing* a 50°C catalitzada pel complex format *in vivo* pel precursor N, H6MBP(GD)-NeqN (56492.6 Da), i el precursor C, NeqC-(TD)GB1 (10626.8 Da), a una concentració de 10 µM amb 50 mM tampó fosfat a pH 8.0, 150 mM NaCl, 1 mM EDTA.

El sistema dual d'expressió pSKDuet01/pSKBAD2 permet coexpressar els dos precursors de manera simultània en *E. coli*, obtenint d'aquesta manera el complex NeqN/NeqC. L'inconvenient que presenta aquest sistema és que els nivells d'expressió de cada precursor són molt diferents, obtenint alts nivells d'expressió del precursor N que s'expressa sota el promotor T7 del vector pSKDuet01, i baixos nivells del precursor C que es troba clonat en el vector pSKBAD2 sota el control del promotor d'arabinosa. A conseqüència d'aquesta expressió diferencial i tenint en compte que l'associació NeqN-NeqC segueix una relació 1:1, s'obté un excés de precursor N lliure, el qual no formarà complex i difícilment serà purificat del que forma complex amb el precursor C. Com a resultat s'obté una barreja de precursor N lliure i acomplexat que dificultarà la interpretació dels resultats de la reacció de *trans-splicing*.

Per a solucionar aquest problema i poder ajustar millor les concentracions dels precursors es va decidir expressar i purificar els dos precursors separatament.

El precursor N, H6MBP(GD)-NeqN, és el mateix que l'utilitzat anteriorment (Figura 4.1). En la Figura 4.5 es representa de manera esquematitzada l'estructura del precursor C: NeqC-(TD)GB1-H6. La massa molecular d'aquest precursor és de 12146.5 Da (108 residus) i presenta un punt isoelèctric teòric de 6.1.



Figura 4.5. Esquema del precursor C, NeqC-(TD)GB1H6. El qual està format per la intèina C, NeqC, unida a la GB1, a través d'una regió connectora, que incorpora els dos residus natus, TD, en posicions +1 i +2 (vermell) i en l'extrem C-terminal una cua d'histidines (H6).

Les masses moleculars tant dels precursors com dels possibles productes de la reacció s'indiquen en la Taula 4.2.

Taula 4.2. Masses moleculars dels productes obtinguts en la reacció de *trans-splicing* catalitzada pel precursor N, H6MBP(GD)-NeqN, i el precursor C, NeqC-(TD)GB1H6 *in vitro*. Amb un * estan marcades aquelles espècies que es poden detectar amb anticossos Anti-His.

Espècie	Massa molecular (Da)
Precursor N*	56492.6
Precursor C*	12146.5
Producte <i>Splicing</i> *	53665.0
NeqN	11477.0
NeqC	3515.1
H6MBP(GD)-*	45033.7
-(TD)GB1H6*	8649.37

La reacció de *splicing* així com les reaccions d'hidròlisi poden ser regulades pel pH del medi [158], [159]. Per estudiar l'efecte del pH en la reacció de *trans-splicing* catalitzada pels precursors H6MBP(GD)-NeqN i NeqC-(TD)GB1H6, es van preparar tres reaccions a diferent pH (6.0, 7.0 i 8.0). Els tampons utilitzats tenen la mateixa força iònica i concentració d'amortidor, l'únic que varia és el pH (Taula 4.3). Per a preparar aquestes reaccions es va partir dels precursors prèviament purificats per cromatografia d'afinitat Ni-NTA i dialitzats enfront del tampó amb del pH corresponent. Per a obtenir la reacció de *trans-splicing* pròpiament dita es van barrejar els volums adients de cada precursor al pH escollit amb l'objectiu d'obtenir una concentració final de cada un dels precursors de 2.3 μ M. Les tres reaccions de *trans-splicing* es van incubar durant 24 hores a 50°C. Per a realitzar el seguiment de la reacció es van extreure diferents alíquotes a diferents temps. La reacció de *trans-splicing* es va analitzar per 12%-SDS-PAGE.

Taula 4.3. Diferents tampons utilitzats per l'estudi de l'efecte del pH sobre la reacció de *trans-splicing*.

pH	Tampó
6.0	50 mM tampó fosfat, 150mM NaCl + 1mM EDTA
7.0	50 mM tampó fosfat, 150mM NaCl + 1mM EDTA
8.0	50 mM tampó fosfat, 150mM NaCl + 1mM EDTA

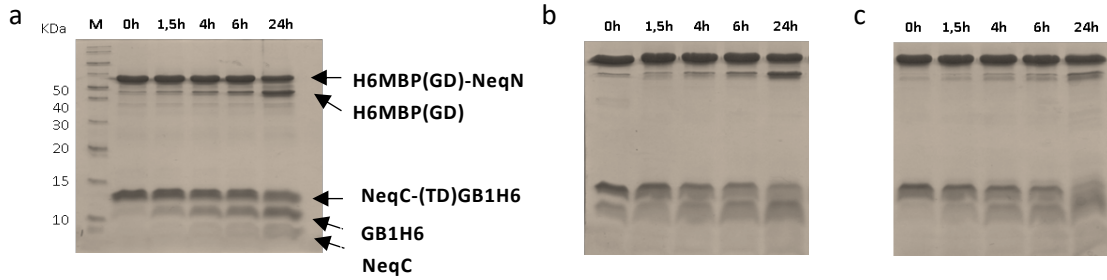


Figura 4.6. Anàlisi per SDS-PAGE de la reacció de *trans-splicing* a 50°C catalitzada pel precursor N, H6MBP(GD)-NeqN, i el precursor C, NeqC-(TD)GB1 a diferents pHs. (a) pH 6.0 (b) pH 7.0 (c) pH 8.0. (M) Marcador de massa molecular.

En cap de les reaccions assajades s'observa cap banda corresponent al producte de *trans-splicing*, en canvi, si que s'identifiquen les corresponents als productes de les reaccions d'hidròlisi (Figura 4.6; Annex C). El que indica que la inteïna partida Neq pol amb les exteïnes exògenes en un rang de pHs entre 6.0-8.0 a 50°C no és capaç de coordinar les diferents etapes de la reacció de *trans-splicing* per a formar el producte final d'aquesta reacció.

Fins ara tots els assajos descrits s'han realitzat a 50°C, límit inferior del rang de temperatures on Suk-Tae Kwon i col·laboradors varen detectar activitat de la inteïna Neq pol en un context endogen [68]. En aquest treball es va descriure que la inteïna Neq pol presentava el seu màxim d'activitat a temperatures per sobre dels 70°C. Per aquesta raó, es va voler estudiar el comportament d'aquests precursors quan s'incuben a una temperatura de 70°C.

Per a dur a terme aquest estudi, es va preparar una solució a una concentració final de 3.5 µM de precursor N i de precursor C amb 50 mM tampó fosfat, pH 8.0 + 150 mM NaCl + 1 mM EDTA i es va incubar durant 24 hores a 70°C. Per realitzar el seguiment de la reacció es van extreure alíquotes a diferents temps. La reacció de *trans-splicing* es va analitzar per 12%-SDS-PAGE.

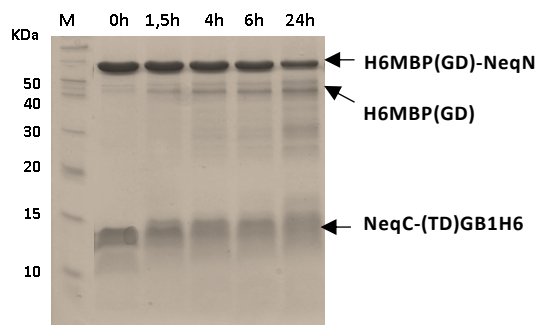


Figura 4.7. Anàlisi per SDS-PAGE de la reacció de *trans-splicing* catalitzada pels precursors N i C, H6MBP(GD)-NeqN i NeqC-(TD)GB1H6 a 70°C. M: Marcador de massa molecular.

En la Figura 4.7 es pot observar que a 70°C a mesura que avança el temps de reacció s'observa una disminució significativa en la intensitat de la banda corresponent al precursor N i una reducció en la mobilitat electroforètica del precursor C, la qual va acompanyada d'un decrement en la seva resolució. D'altra banda, també s'aprecia que a partir de les 4 hores d'incubació a 70°C, apareixen nombroses bandes inespecífiques que van augmentant en incrementar el temps de reacció. La disminució del precursor N al llarg de la reacció sense observar cap increment proporcional en el producte de la hidròlisi de l'extrem N-terminal o en el producte de *trans-splicing*, així com l'aparició de bandes inespecífiques, suggereix que el precursor N a 70°C agrega. L'agregació del precursor N a aquesta temperatura es pot explicar per què la MBP té una T_m a pH 8.2 de 62.6°C [153].

No és aconsellable treballar amb una exteïna que presenta una T_m per sota de la temperatura òptima d'activitat descrita per la inteïna partida Neq pol, com a conseqüència es va decidir substituir-la i assajar noves exteïnes per a futurs estudis.

4.1.2. Reacció de *cis-splicing* utilitzant com a exteïna N el domini GB1 i com a exteïna C el domini ZHER2

S'ha demostrat que l'efecte de les exteïnes exògenes, en la reacció de *cis-splicing* és menor que en la reacció de *trans-splicing* [55], on les exteïnes poden arribar a inhibir-la [160]. El diferent efecte que tenen les exteïnes sobre la reacció de *trans-splicing* i *cis-splicing* suggereix que probablement les exteïnes poden modular o bloquejar el procés d'associació dels dos fragments de la inteïna. Les exteïnes poden interferir en el procés d'associació de les inteïnes partides perquè poden canviar la solubilitat o la càrrega neta dels precursors [55].

Tot i que *in vivo* s'ha vist que els dos precursors de la inteïna partida Neq pol interaccionen correctament formant el complex NeqN/NeqC, es coneix poc sobre el procés *in vitro*. Per minimitzar la interferència de les exteïnes en la reacció de *splicing*, es van fusionar els gens que codifiquen els dos fragments de la inteïna Neq pol a través d'una seqüència que codifica per la diana de la proteasa TEV (Figura 4.8).

A conseqüència dels problemes d'agregació obtinguts amb el precursor N, H6MBP(GD)-NeqN, a elevades temperatures, es va realitzar un canvi d'exteïnes. En la nova construcció la inteïna es troba unida per l'extrem N-terminal al domini GB1 i per l'extrem C-terminal al affibody ZHER2.

En la Figura 4.8 s'esquematitza l'estructura del precursor en cis H6GB1(D)-NeqSC-(TDS)ZHER2. En aquest precursor els dos fragments de la inteïna partida Neq pol es troben units covalentment a través de la seqüència diana de la proteasa TEV, la qual permetria obtenir per digestió proteolítica els dos precursors en *trans*, H6GB1(D)-NeqN i NeqC-(TDS)ZHER2. Aquest precursor està compost per 280 residus (PM= 31997.7 Da) i presenta un pI de 6.27.



Figura 4.8. Esquema del precursor en cis H6MBP(D)-NeqSC-(TDS)ZHER2, el qual està compost per una cua d'histidines en l'extrem N-terminal, seguida del domini GB1 unit a la NeqN a través d'una curta seqüència polipeptídica, a continuació es troba la diana per la proteasa TEV i la NeqC unida mitjançant un curt linker a l'exteïna C, ZHER2. En vermell s'assenyalen els residus nadius en les posicions flanquejants als dos fragments de la inteïna.

A la Figura 4.9 s'esquematitza la reacció de *splicing* catalitzada pel precursor en cis H6GB1(D)-NeqSC-(TDS)ZHER2, així com les reaccions no desitjades d'hidròlisi.

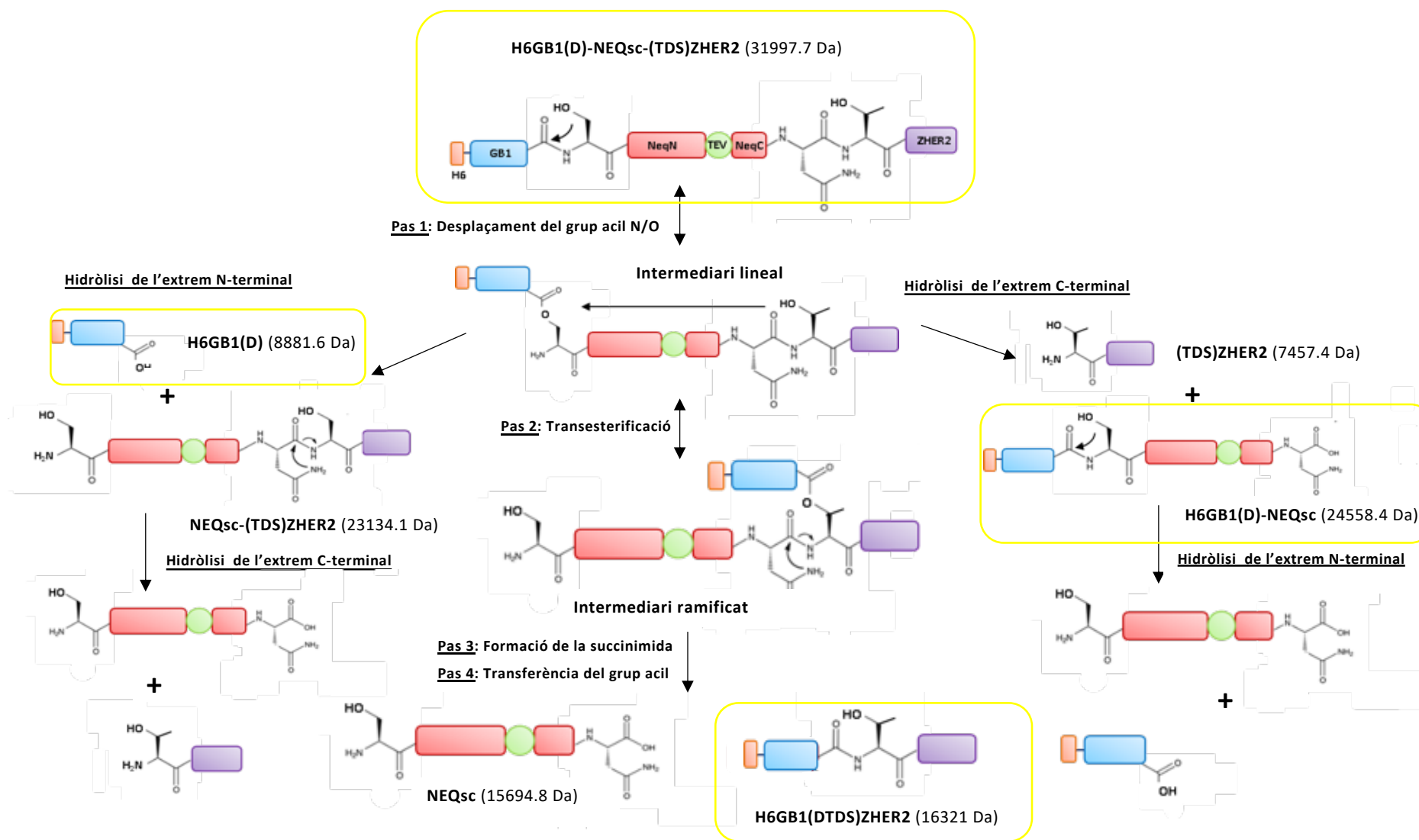


Figura 4.9. Esquema de la reacció de cis-splicing catalitzada pel precursor H6GB1(D)-NeqSC-(TDS)ZHER2, i les possibles reaccions no desitjades d'hidròlisi. Entre parèntesi es troben les masses moleculars de les diferents espècies que es poden formar al llarg de la reacció. En groc es troben emmarcades les molècules detectables amb anticossos AntiHis.

Per tal d'estudiar l'efecte de la temperatura, el precursor H6GB1(D)-NeqSC-(TDS)ZHER2 prèviament purificat per cromatografia d'afinitat Ni-NTA a una concentració de 5 μ M amb 20 mM tampó fosfat a pH 8.0, 50 mM NaCl es va aliquotar i cada fracció es va incubar durant 24 hores a diferents temperatures (40°C; 50°C; 60°C; 70°C). Per al seguiment de la reacció a diferents temps (0 min, 10 min, 30 min, 1h, 4h, 24h) es van extreure diferents alíquotes de cada mostra per a posteriorment analitzar-les per 15% SDS-PAGE.

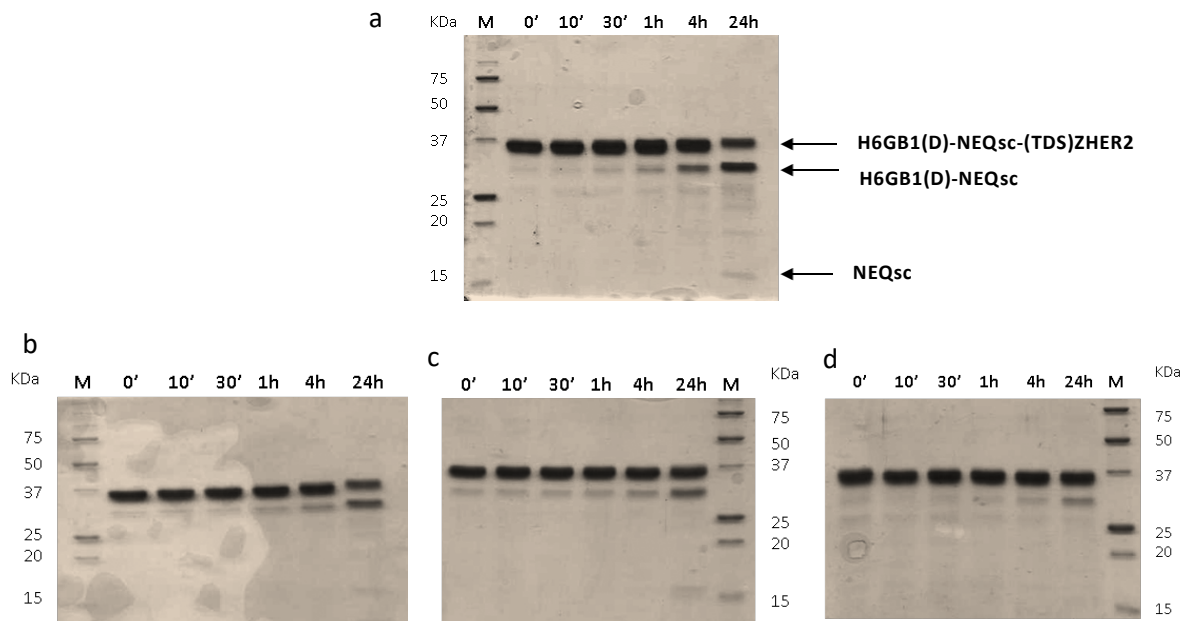


Figura 4.10. Anàlisi de la reacció de *cis-splicing* per SDS-PAGE catalitzada pel precursor precursor H6GB1(D)-NeqSC-(TDS)ZHER2 a pH 8.0 en un rang de temperatures entre els 40°C i 70°C. (a) 70°C (b) 60°C (c) 50°C (d) 40°C. El precursor ha estat preparat a una concentració final de 5 μ M amb 20 mM tampó fosfat a pH 8.0, 50 mM NaCl. M: Marcadors de massa molecular.

Els resultats obtinguts s'observen en la Figura 4.10. En aquests es pot veure com el precursor en *cis* H6GB1(D)-NeqSC-(TDS)ZHER2 es manté soluble en tot el rang de temperatures estudiat, la qual cosa indica que les exteïnes escollides són bones candidates per a ser utilitzades en reaccions a elevades temperatures. D'altra banda, no s'observa cap banda corresponent al producte de *splicing* H6GB1(D)-(TDS)ZHER2 en cap de les temperatures assajades. Com era d'esperar tenint en compte els resultats obtinguts anteriorment (Figura 4.6), s'observa la ciclació de la cadena lateral de l'Asn de manera prematura, donant lloc a la formació dels productes H6GB1(D)NeqSC i (TDS)ZHER2. Els productes de la hidròlisi de l'extrem C-terminal han sigut identificats per Western Blot, utilitzant anticossos AntiHis i MALDI-TOF. La comparació dels resultats obtinguts a les diferents temperatures, suggereix que aquesta reacció no està regulada per la temperatura, però sí que la seva cinètica es veu incrementada a mesura que augmenta la

temperatura. En els gels d'acrilamida també es veu l'aparició a temps llargs, >4 h, de la *cis* inteïna NeqSC, la qual va acompanyada de l'aparició de l'exteïna N, H6GB1, la presència de la qual va ser identificada per MALDI-TOF. L'aparició al llarg de la reacció d'aquestes dues espècies indica que també es dona la hidròlisi de l'extrem N-terminal. En aquest cas i a diferència de la ciclació de l'Asn sembla que aquesta reacció estaria regulada per la temperatura, ja que a 40°C no s'observa cap banda corresponent a la inteïna. Aquest resultat és consistent amb els resultats de *splicing* obtinguts en el treball de Suk-Tae Kwon i col·laboradors [68]. A més a més, l'aparició tardana de la inteïna suggereix que les reaccions d'hidròlisi no es troben coordinades, dit d'una altra manera sembla que la hidròlisi de l'extrem N-terminal és sobre el producte de la hidròlisi de l'extrem C-terminal, H6GB1NeqSC, el que indica que la hidròlisi de l'extrem N-terminal es dona des de l'intermediari lineal. Per últim, s'ha de dir que el patró de bandes obtingut, tot i que s'utilitzen diferents exteïnes, s'assembla molt a l'obtingut per la reacció de *trans-splicing* catalitzada pels precursors H6MBP(GD)-NeqN i NeqC-(TD)GB1H6 (Figura 4.6), la qual cosa suggereix que les exteïnes no interfereixen significativament en el procés d'associació dels dos fragments de la inteïna partida.

Per estudiar l'efecte del pH en la reacció de *cis-splicing* el precursor en *cis* prèviament purificat es va preparar a una concentració final de 5 µM en dues solucions amortidores amb pHs diferents, 20 mM tampó acetat a pH 5.0, 50 mM NaCl i 20 mM tampó fosfat a pH 7.2, 50 mM NaCl. Alíquotes de cada una de les solucions es va incubar a una temperatura diferent (40°C, 50°C, 60°C i 70°C) durant 24 hores. Pel seguiment de la reacció es varen agafar mostres a diferents temps i es van analitzar per 15% SDS-PAGE.

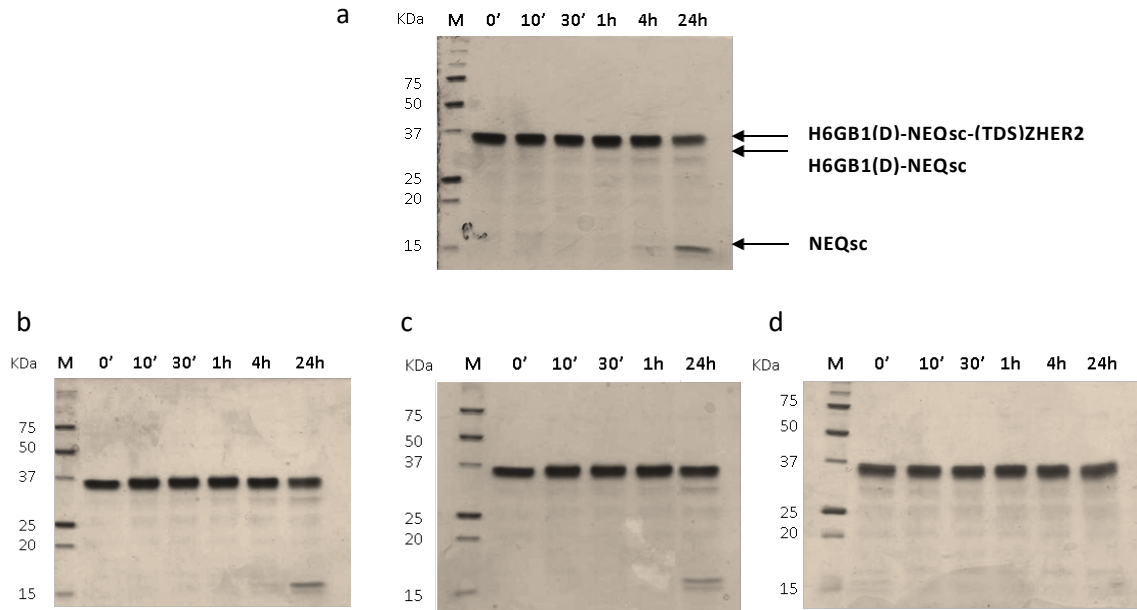


Figura 4.11. Anàlisi per SDS-PAGE de la reacció de cis-splicing catalitzada pel precursor H6GB1(D)-NeqSC-(TDS)ZHER2 a pH 5.0 en un rang de temperatures entre 40°C i 70°C. (a) 70°C (b) 60°C (c) 50°C (d) 40°C. M: Marcadors de massa molecular.

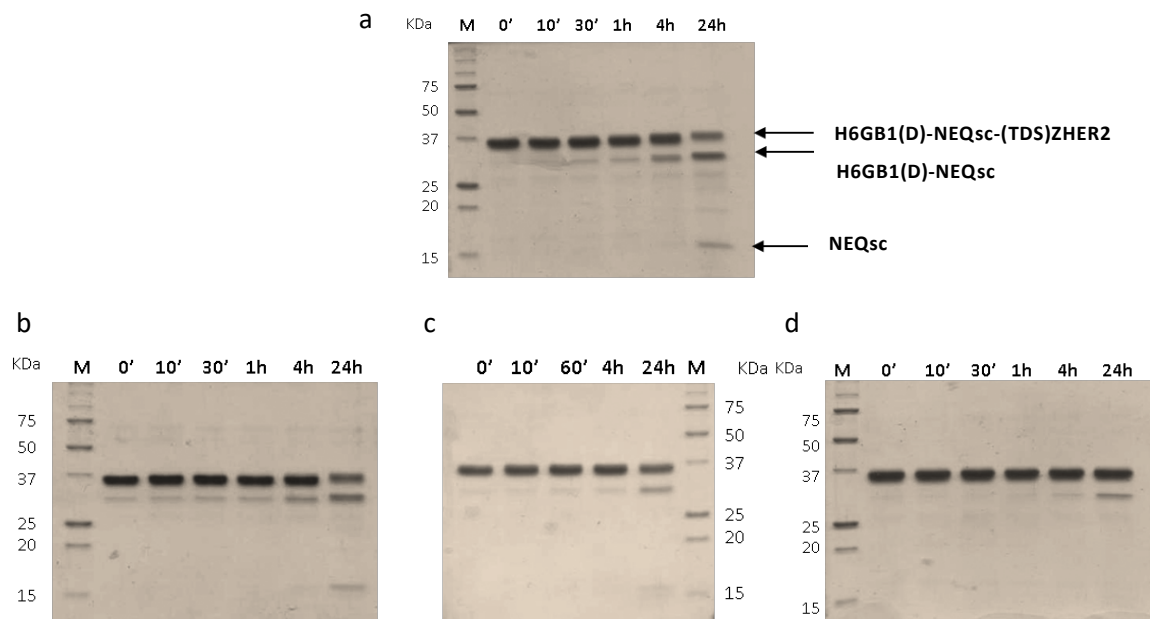


Figura 4.12. Anàlisi per SDS-PAGE de la reacció de cis-splicing catalitzada pel precursor H6GB1(D)-NeqSC-(TDS)ZHER2 a pH 7.2 en un rang de temperatures entre 40°C i 70°C. (a) 70°C (b) 60°C (c) 50°C (d) 40°C. M: Marcadors de massa molecular.

Amb els resultats obtinguts a pH 5.0 i 7.2 (Figura 4.11 i Figura 4.12) sembla que el patró de bandes obtingut de la reacció de *splicing* als diferents pH és bastant similar a l'obtingut a pH 8.0 (Figura 4.10), és a dir, en cap de les condicions de pH i temperatura s'observa producte de *splicing*, i en tots els pHs amb major o menor grau s'observen els productes de la hidròlisi de l'extrem C- i N-terminal.

Comparant els diferents pHs (5.0, 7.2, 8.0) s'observa que en tots ells independentment de la temperatura es dona la ciclació de la cadena lateral de l'Asn. Aquest resultat reforça els resultats obtinguts anteriorment per la reacció de *trans-splicing* catalitzada pels precursors H6MBP(GD)-NeqN i NeqC-(TD)GB1H6 on la ciclació de la cadena lateral de l'Asn no estava regulada per la temperatura. Per a facilitar la interpretació dels resultats s'assumeix que la hidròlisi de l'extrem N-terminal es dona un cop s'ha produït la hidròlisi de l'extrem C-terminal, i per tant tot el precursor que desapareix al llarg de la reacció és hidrolitzat per la ciclació de l'Asn, convertint-se en H6GB1(D)NeqSC + (TDS)ZHER2. Comparant la disminució de la intensitat de la banda corresponent al precursor després de 24 hores a 70°C (temperatura a la qual la ciclació es dona més eficientment) sembla que la velocitat d'hidròlisi de l'extrem C-terminal és més elevada a pHs bàsics, pH 8.0. Per contra, la velocitat d'hidròlisi de l'extrem N-terminal és més elevada a pH àcids, ja que la intensitat de la banda corresponent a NEQsc és més gran a pH 5.0. Aquest increment de la velocitat en la hidròlisi de l'extrem N-terminal a pHs àcids es pot explicar per què en l'equilibri entre l'enllaç amida (precursor) i l'enllaç èster (intermediari lineal) l'enllaç amida es veu afavorit a alts pHs i en canvi l'enllaç èster a baixos pHs [161]. Com ja s'ha comentat anteriorment, a cap dels pH assajats es dona hidròlisi de l'extrem N-terminal a temperatures per sota dels 50°C, fet que reforça la idea que aquesta reacció està modulada per la temperatura.

4.1.3. Reacció de *trans-splicing* utilitzant com a exteïna N el domini GB1 i com a exteïna C el domini ZHER2

A causa de la baixa expressió i solubilitat del precursor C, NeqC-(TD)ZHER2 (Annex D), i a la dificultat de purificar-lo, tot i els seus inconvenients, es va decidir coexpressar els precursors en *E. coli*, i així copurificar el complex NeqN/NeqC, el qual es manté soluble i estable.

D'altra banda, degut a la manca de producte de *splicing* i a la importància descrita dels residus natis de les exteïnes en les regions adjacents als llocs d'unió exteïna-inteïna, es va incorporar una petita seqüència de 7 residus de l'exteïna N nativa en la regió connectora entre la GB1 i la NeqN. Aquesta curta seqüència salvatge s'assenyala com wt en la Figura 4.13 .

Les estructures dels precursors N: H6GB(wt)-NeqN i C: NeqC-(TDS)ZHER2 estan esquematitzades en la Figura 4.13. El precursor N està format per 183 aminoàcids (20846.2 Da), i té un punt isoelèctric teòric de 5.9, i el precursor C està constituït per 95 residus (10626.8 Da) i presenta un pI de 5.04.



Figura 4.13. Esquema de precursor N, H6GB(wt)-NeqN, i C, NeqC-(TDS)ZHER2. (a) Precursor N compost per una cua d'histidines (H6) en l'extrem N-terminal de la proteïna de fusió, seguida de l'exteïna N, GB1, la qual es troba unida a través d'un linker a la NeqN. (b) Precursor C està format per la inteïna C, NeqC, unida a ZHER2, a través d'una regió connectora. Els residus assenyalats en vermell són els residus nadius de l'exteïna de *N. equitans*.

El precursor N, proteïna cinèticament estable

L'estat natiu de les proteïnes és un estat dinàmic, el qual està format per una població de molècules amb diferents conformacions estructurals que van fluctuant entre elles. Les molècules de SDS requereixen tenir accés al nucli hidrofòbic de la proteïna per a unir-se, i així poder desplegar-la. Per aquesta raó, la unió del SDS es dona en la transició entre les diferents conformacions de l'estat natiu. Aquesta *transició* es considera un moment de debilitat, en el qual la proteïna és susceptible a la unió d'aquestes molècules [138]. Un cop les molècules de detergent s'uneixen, aquesta proteïna queda atrapada en l'estat desplegat.

Quan es parla de proteïnes cinèticament estables es fa referència a proteïnes que presenten una inusual rigidesa en la seva estructura nativa, i per tant, el nombre de conformacions estructurals que formen l'estat natiu és menor. Aquesta rigidesa dificulta la unió de les molècules de SDS a la proteïna, com a conseqüència s'obté una població heterogènia de molècules plegades, parcialment plegades i totalment desplegadas que es tradueix en una elevada heterogeneïtat de bandes en els gels d'acrilamida. Per a solucionar aquest problema, generalment els protocols d'anàlisi de proteïnes per SDS, incorporen un pas de preparació de la mostra on aquesta es sotmet a un xoc tèrmic, bullint la mostra durant 10 minuts a 95°C. Quan es treballa amb proteïnes termoestables, com és el cas de la Neq pol, aquest protocol preparatiu no funciona, ja que a la temperatura a la qual es bullis la mostra no desnatura les espècies parcialment o totalment plegades, per tant, la complexitat es manté.

Com es pot observar en la Figura 4.14, la NeqN per si sola també presenta aquesta heterogeneïtat de bandes, per tant, aquests resultats indiquen que l'elevada heterogeneïtat del precursor N és deguda a la NeqN exclusivament, el que a la vegada indica que aquest fragment és cinèticament estable. En el cas dels precursors, H6MBP-NeqN i H6GB1(D)-NEQsc-(TDS)ZHER2, no s'observa aquesta heterogeneïtat de bandes en SDS-PAGE, segurament, perquè la complexitat aportada per la NeqN queda emmascarada per la gran mida del 'cargó' que tenen aquests precursors. Per tant, aquests resultats indiquen que al NeqN és una proteïna cinèticament estable.

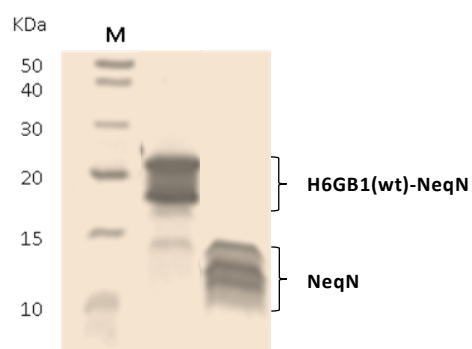


Figura 4.14. Anàlisi per SDS-PAGE de la complexitat del precursor N, H6GB1(wt)-NeqN i la NeqN (20846.2 Da i 12861.5 Da, respectivament). El precursor N i la proteïna N s'han preparat a una concentració final de 20 μ M amb 50 mM Tris-HCl a pH 8.0 i 50 mM NaCl. M: Marcadors de massa molecular

Aquesta heterogeneïtat de bandes originada per la proteïna N dificulta enormement la interpretació dels resultats obtinguts en l'anàlisi per SDS-PAGE de la reacció de *trans-splicing*. Per aquesta raó es va cercar un protocol que permetés desnaturalitzar totalment la NeqN i reduir l'heterogeneïtat de bandes en l'anàlisi electroforètica. Per tal de trobar el protocol, es va preparar una solució de precursor N aïllat, H6GB1(wt)-NeqN, i una altra formant complex amb el precursor C, NeqC-(TD)ZHER2, a una concentració final de 20 μ M, amb 50 mM Tris-HCl, pH 8.0 i 50 mM NaCl. Es varen barrejar volums iguals de les solucions proteïques amb tampó de càrrega 2X amb urea (50 mM Tris-HCl, pH 6.8 + 1.6% SDS + 7% Glicerol + 8M urea) i es van incubar durant 3 hores a 95°C. Per al seguiment de la reacció de desnaturalització es varen agafar alíquotes a diferents temps, les quals es van analitzar per 15% SDS-PAGE.

A continuació, es mostren els gels d'acrilamida al 15% obtinguts de l'anàlisi de desnaturalització (Figura 4.15). Com es pot observar, per desnaturalitzar totalment el precursor N, sigui aïllat o formant complex, és necessari incubar la mostra durant un mínim de 2h en presència de 4M d'urea.

D'aquest estudi s'extreu el protocol per al tractament de mostres amb l'objectiu d'analitzar mitjançant SDS-PAGE la reacció de *trans-splicing* catalitzada pels precursors H6GB1(wt)-NeqN/ H6GB1(wt)-NeqN(S1C) i NeqC-(TDS)ZHER2, el qual consisteix en barrejar volums igual de la reacció amb el tampó de càrrega x2 amb urea (50 mM Tris-HCl, pH 6.8 + 1.6% SDS + 7% Glicerol + 8M urea), seguidament incubar durant 2 hores a 95°C, i carregar al gel SDS-PAGE.

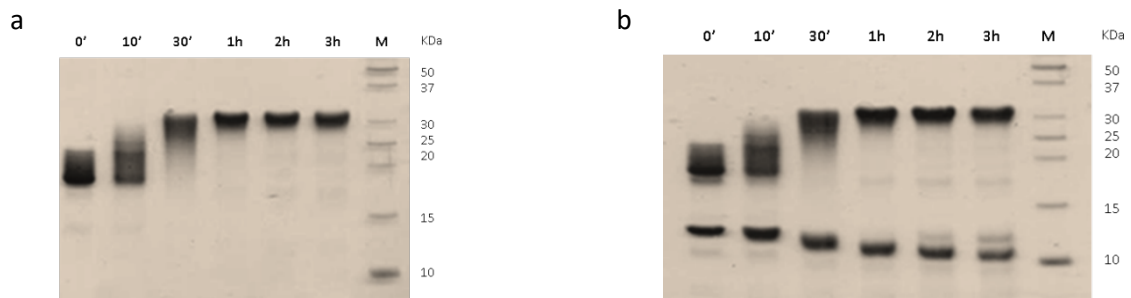


Figura 4.15. Cinètica del procés de desnaturalització a 95°C amb 4 M d'urea analitzada per SDS-PAGE del precursor H6GB(wt)-NeqN i del complex format per H6GB(wt)-NeqN i NeqC-(TDS)ZHER2 . (a) Precursor N aïllat (b) Precursor N formant complex amb el precursor C M: Marcadors de massa molecular.

Estudi de la reacció de *trans-splicing*

Amb l'objectiu de trobar alguna condició en la qual s'observés el producte de *trans-splicing*, H6GB1(wt)-(TDS)ZHER2, la reacció catalitzada pels precursors: H6GB1(wt)-NeqN i NeqC-(TDS)ZHER2 es va estudiar en un rang de pH entre 3.5 i 9.0. Per l'anàlisi de les reaccions per SDS-PAGE es va utilitzar el protocol descrit anteriorment per evitar l'heterogeneïtat de bandes observada en el precursor N i la NeqN.

En la Figura 4.16 es representa de manera esquemàtica la reacció de *trans-splicing* estudiada, així com les reaccions secundàries no desitjades d'hidròlisi.

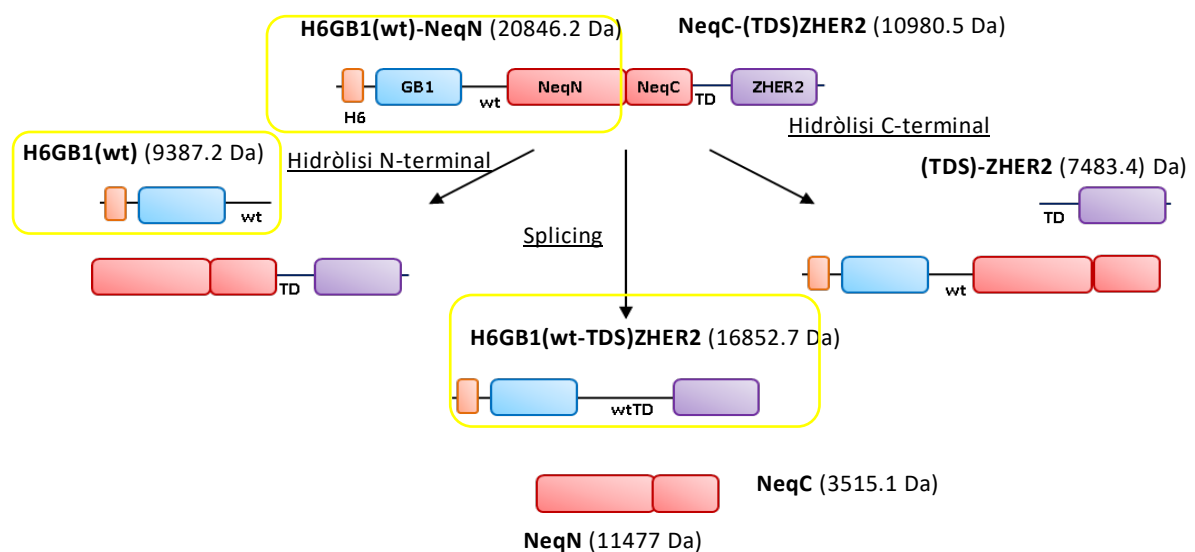


Figura 4.16. Esquema de la reacció de trans-splicing i de les reaccions secundàries del precursor N, H6GB1(wt)NeqN, i el precursor C, NeqC-(TDS)ZHER2. En aquest esquema es representen les diferents espècies que es poden generar durant la reacció amb les seves masses moleculars. En groc es troben emmarcades les molècules que es poden detectar per anticossos Anti-His.

Per dur a terme l'assaig, es va preparar una solució del complex NeqN/NeqC prèviament purificat a una concentració final de 20 μ M a cada un dels pHs. Per als pHs entre 3.5-6.0 es va utilitzar una solució amortidora 50 mM tampó citrat, 50 mM NaCl, el pH de la qual es va ajustar barrejant diferents volums d'una solució 0.1M d'àcid cítric monohidratat amb una 0.1M de citrat trisòdic dihidratat [162]; i per als pHs entre 7.0 i 9.0 es va utilitzar una solució amortidora 50 mM Tris-HCl, 50 mM NaCl, ajustant el pH amb àcid clorhídric. Cada una de les solucions es va incubar durant 24 hores a 70°C. Es va escollir incubar les mostres a 70°C, ja que era la temperatura a la qual s'ha observat la inteïna més activa (Apartat 4.1.2 de Resultats i Discussió). En aquest cas es van agafar dues alíquotes una a temps 0 i una altra a les 24 hores d'incubació, les quals es van analitzar per SDS-PAGE¹.

¹ Per l'anàlisi per SDS-PAGE les mostres es van preparar utilitzant el protocol descrit en la pàgina 96.

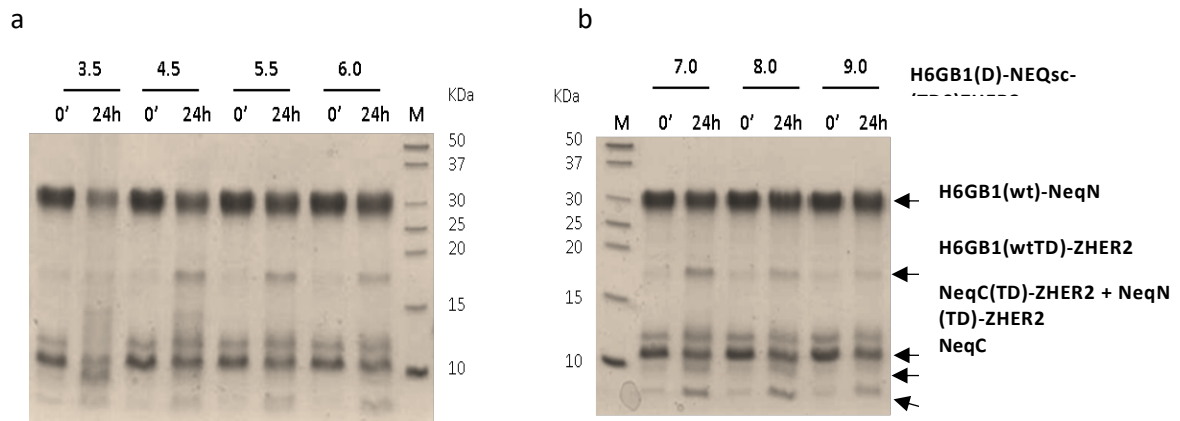


Figura 4.17. Anàlisi per SDS-PAGE de la reacció de *trans-splicing in vitro* a 70°C catalitzada pels precursors: H6GB1(wt)-NeqN i NeqC-(TDS)ZHER2 en un rang de pH entre 3.5 i 9.0. (a) Rang pH 3.5 a 6.0; (b) Rang de pH: 7.0 a 9.0. M: Marcadors de massa molecular.

Els resultats (Figura 4.17) mostren que després d'incubar els dos precursors durant 24 a 70°C en un rang de pHs entre 4.5 i 8.0 apareixen tres bandes: una banda amb una mobilitat electroforètica entre els marcadors de massa molecular de 20 i 15 kDa, la qual podria ser el producte de *splicing*, H6GB1(wt)-(TDS)ZHER2; una altra banda just per sota del precursor C, segurament es tracti de la hidròlisi de l'extrem C-terminal, (TD)ZHER2; i la tercera banda en el límit inferior del gel, amb una mobilitat electroforètica molt per sota del marcador de massa molecular de 10 kDa, seria la inteïna C. La formació del producte de *trans-splicing* es va confirmar per Western-Blot utilitzant anticossos Anti-His (Annex C).

Aquests resultats mostren que amb la introducció dels set residus nadius en la regió C-terminal de l'exteïna N, s'aconsegueix que la inteïna Neq pol catalitzi la reacció de *trans-splicing* de manera coordinada, obtenint d'aquesta manera el producte final d'aquesta reacció, el que indica que aquests residus juguen un paper clau en la reacció de *trans-splicing*. De fet si s'observa l'estructura cristal·logràfica del complex NeqN/NeqC (5oxz) s'observa que aquests residus estableixen interaccions específiques amb residus de la NeqC i de l'exteïna C, probablement contribuint a la correcta orientació del centre catalític de la inteïna.

Aquesta reacció de *trans-splicing* està lluny de considerar-se eficient, ja que la cinètica de reacció és lenta i la quantitat de producte obtingut és molt baix en tots els casos, el que suggereix que la inteïna Neq pol és molt restrictiva a canvis d'exteïnes. D'altra banda, aquests resultats indiquen que la reacció de *trans-splicing* es dona en un ampli rang de pH, entre 4.5 i 8.0, obtenint el rendiment més gran a pH 7.0. El rang de pH en el qual es dona la reacció de *trans-splicing* amb exteïnes exògenes és diferent del rang en el qual es dona la reacció en un context endogen (rang de pH entre 6.0 i 9.0) [68]. Aquesta diferència segurament ve definida pel diferent pl de les exteïnes. Finament, cal ressaltar que amb la introducció dels residus nadius de l'exteïna N en la

regió adjacent a la Ser1 en cap dels pH s'observa hidròlisi de l'extrem N-terminal i la ciclació de la cadena lateral de l'Asn s'ha reduït molt.

Per tal d'estudiar l'efecte de la temperatura sobre la reacció de *trans-splicing* catalitzada per aquests precursors es va preparar una solució del complex NeqN/NeqC prèviament purificat a una concentració final de 20 μ M amb 20 mM Tris-HCl a pH 7.0, 50 mM NaCl. Aquesta solució es va fraccionar en diferents alíquotes cada una de les quals es va incubar a una temperatura durant 24 h. Per l'anàlisi de la reacció es van agafar mostres a diferents temps i es van analitzar per SDS-PAGE².

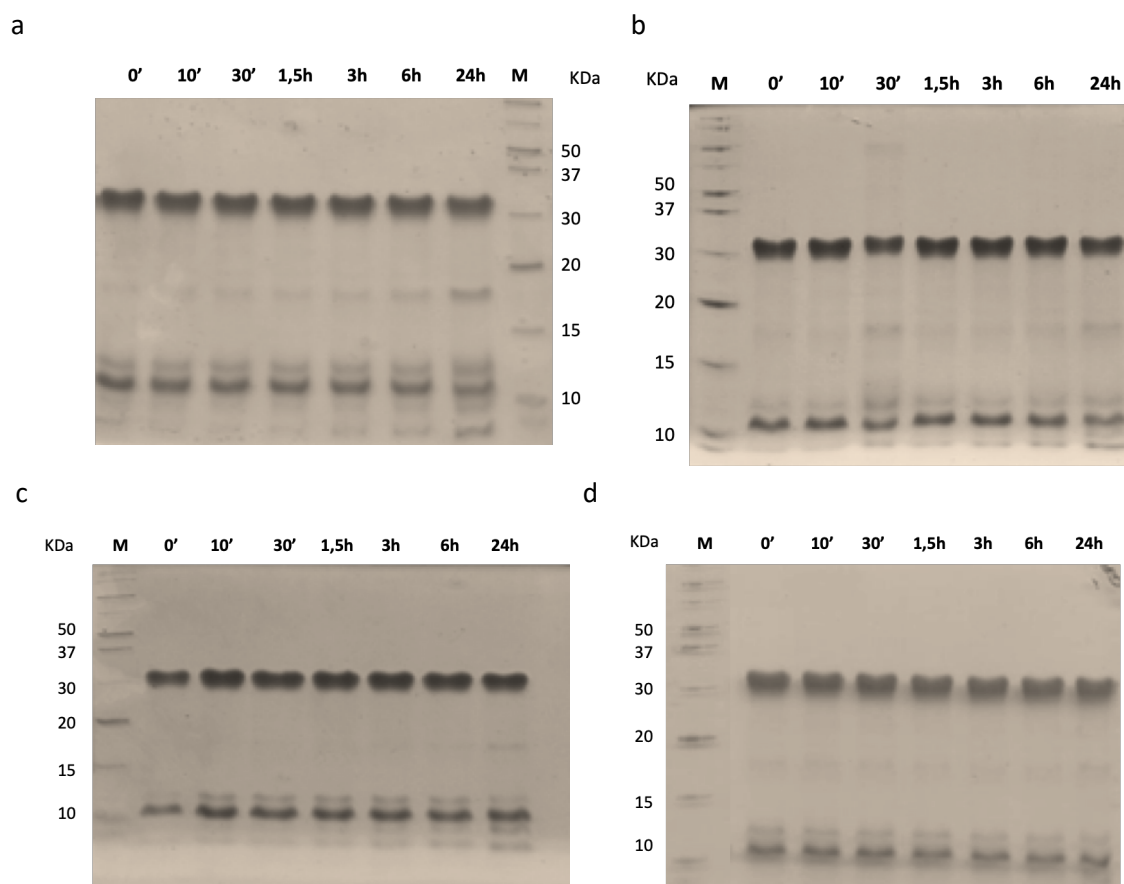


Figura 4.18. Anàlisi per SDS-PAGE de la reacció de *trans-splicing in vitro* catalitzada pels precursors: H6GB1(wt)-NeqN i NeqC-(TDS)ZHER2 en un rang de temperatures entre els 40°C i 70°C. (a) 70°C; (b) 60°C; (c) 50°C; (d) 40°C. M: Marcadors de massa molecular.

² Per l'anàlisi per SDS-PAGE les mostres es van preparar utilitzant el protocol descrit en la pàgina 96

Els resultats obtinguts Figura 4.18 mostren que efectivament la reacció de *trans-splicing* catalitzada per la Neq pol es troba regulada per la temperatura, ja que no només s'observa el producte d'aquesta a temperatures per sobre dels 50°C.

4.1.4. Efecte de la Ser1 en la reacció de *trans-splicing* catalitzada pels precursors H6GB1(wt)-NeqN i NeqC-(TDS)ZHER2

En la introducció d'aquesta tesi s'ha vist que la majoria de les inteïnes descrites fins ara tenen en la posició 1 una Cys. Per tal d'estudiar l'efecte que té la mutació Ser1Cys sobre la reacció de *trans-splicing*, es van utilitzar els precursors descrits anteriorment, precursor N: H6GB1(wt)-NeqN i NeqC-(TDS)ZHER2, però en aquest cas amb una Cys en la posició 1. Aquests precursors van ser coexpressats i copurificats seguint el mateix protocol que en l'apartat anterior, l'única diferència és que el tampó de càrrega contenia un 2% β ME. La incorporació de l'agent reductor en el tampó de càrrega és per tal de prevenir la formació d'estructures oligomèriques d'elevada massa molecular durant la coexpressió dels precursors a causa de l'establiment de ponts disulfur entre les Cys1 del precursor N (Annex A).

Per la reacció de *trans-splicing* es va preparar una solució dels precursors coexpressats i purificats a una concentració final de 20 μ M amb 20 mM Tris-HCl a pH 7.0 i 50 mM NaCl. Aquesta solució es va incubar durant 24 hores a 70°C. Pel seguiment de la reacció es van agafar alíquotes a diferents temps, les quals es van analitzar per 12% SDS-PAGE³. Paral·lelament, i com a control, es va utilitzar una solució dels precursors coexpressa i purificats, on el precursor N presentava la Ser salvatge en posició 1. Com a control addicional i per a demostrar que realment el precursor N, H6GB1(wt)-NeqN(S1C), en absència del precursor C no és capaç de catalitzar la reacció de desplaçament del grup acil N/O, s'incuba sota les mateixes condicions que el complex, una solució de H6GB1(wt)-NeqN(S1C) a una concentració 10 μ M.

Les masses moleculars tant dels precursors com dels possibles productes de la reacció s'indiquen en la Taula 4.4.

³ Per l'anàlisi per SDS-PAGE les mostres es van preparar utilitzant el protocol descrit en la pàgina 96

Taula 4.4. Masses moleculars dels productes obtinguts en la reacció de *trans-splicing* catalitzada pel precursor N, H6GB1(wt)-NeqN (S1C), i pel precursor C, NeqC-(TDS)ZHER2. * espècies que es poden detectar amb anticossos Anti-His.

Espècie	Massa molecular (Da)
Precursor N (S1C)*	20862.2
Precursor C	10980.5
Producte Splicing*	16852.7
NeqN (S1C)	11493.0
NeqC	3515.1
H6GB1(wt)-*	9387.2
-(TDS)ZHER2	7483.4

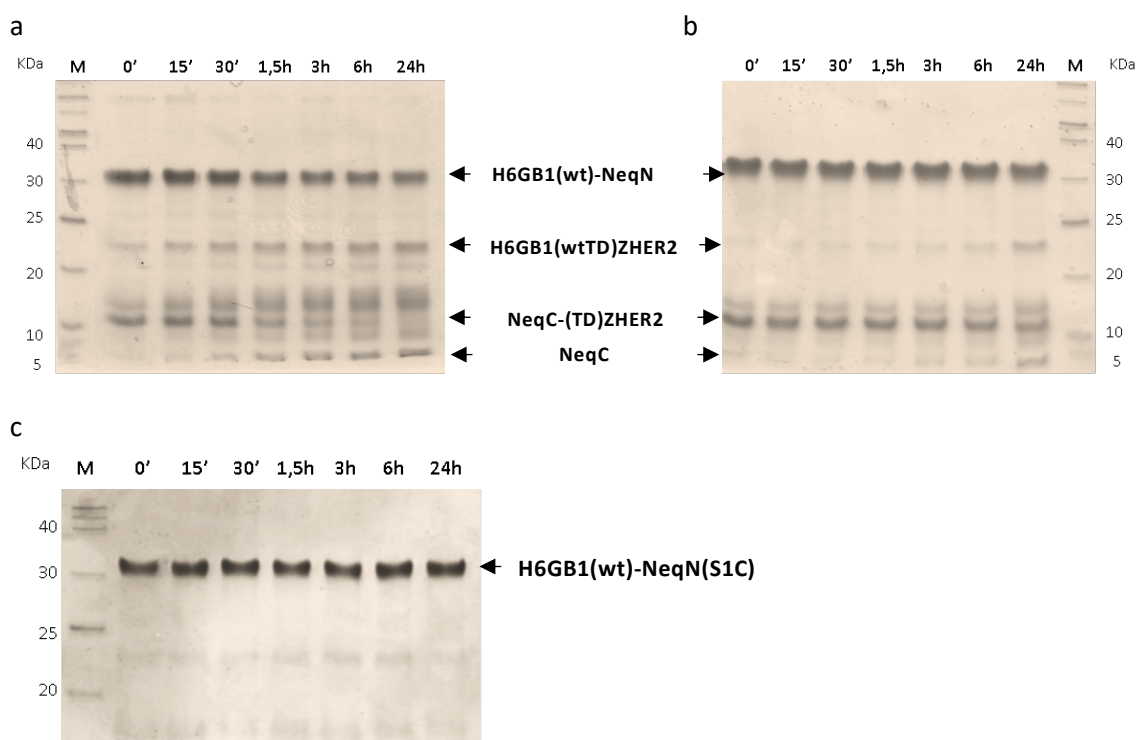


Figura 4.19. Anàlisi per SDS-PAGE de la reacció de *trans-splicing* catalitzada pel complex format *in vivo* H6GB1(wt)-NeqN(S1C)/NeqC-(TDS)ZHER2 i estudi de l'estabilitat del precursor N, H6GB1(wt)-NeqN(S1C) a pH 7.0 i 70°C. (a) Reacció del complex H6GB1(wt)-NeqN(S1C)/NeqC-(TDS)ZHER2; (b) Reacció del complex H6GB1(wt)-NeqN(S1)/NeqC-(TDS)ZHER2. (c) Estabilitat del precursor N, H6GB1(wt)-NeqN(S1C). M: Marcadors de massa molecular.

Els resultats obtinguts (Figura 4.19) indiquen que la mutació Ser1Cys fa que la reacció de *trans-splicing* sigui més eficient, ja que tant la cinètica com el rendiment de la reacció són superiors. De fet sembla que la reacció arriba al seu màxim a les 3 hores, probablement a conseqüència de que el precursor C s'exhaureix (Figura 4.19-a). En aquesta reacció a diferència de la reacció catalitzada pel precursor N amb la Ser1 (Figura 4.19-b), s'observa que per sota del producte de *trans-splicing* apareix una altra banda de menor intensitat al llarg del temps, que probablement es tracta del producte d'hidròlisi de l'extrem N-terminal, H6GB1(wt). D'altra banda, s'observa

què el precursor N amb la mutació Cys1Ser a 70°C durant 24 hores es manté estable en absència del precursor C tot i presentar la Cys1 (Figura 4.19-c), la qual cosa indica que no es duu a terme la primera etapa de desplaçament N/C acil del procés de *splicing*.

Finalment, per a veure l'efecte de la temperatura sobre la reacció de *trans-splicing* catalitzada pel complex amb la mutació Ser1Cys, es va preparar dues solucions a una concentració final de 20 µM amb 20 mM Tris-HCl a pH 7.0 i 50 mM NaCl, cada una de les solucions es van incubar 24 hores a 70°C i 40°C. Per l'anàlisi de la reacció es van extreure mostres a diferents temps i es van analitzar per SDS-PAGE⁴.

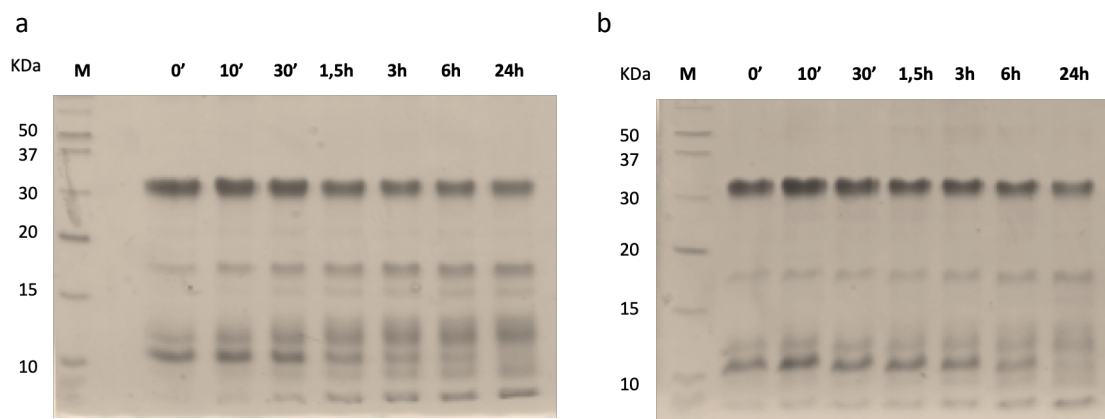


Figura 4.20. Anàlisi per SDS-PAGE de la reacció de *trans-splicing* catalitzada pel complex format *in vivo* H6GB1(wt)-NeqN(S1C)/NeqC-(TDS)ZHER2 a pH 7.0 i temperatures de 70°C i 40°C. (a) 70°C; (b) 40°C. M: Marcadors de massa molecular.

En els resultats mostrats a la Figura 4.20 s'observa que la reacció de *trans-splicing* es produeix a temperatures per sota dels 50°C, tot i que amb un baix rendiment i amb una velocitat més baixa. Un factor que influeix en la diferent reactivitat observada entre la Cys i la Ser podria venir definida pels pKas dels grups tiol (8,5) i hidroxil (13) de les cadenes laterals d'aquests residus. Tot plegat, suggereix que la Ser en posició 1 és un element clau en la regulació per temperatura de la reacció de *trans-splicing* catalitzada per la inteïna Neq pol però no l'únic, ja que si no s'esperaria obtenir un patró de bandes similar a l'obtingut amb l'anàlisi de la reacció a 70°C.

⁴ Per l'anàlisi per SDS-PAGE les mostres es van preparar utilitzant el protocol descrit en la pàgina 96

4.2. Caracterització de la inteïna N de *N.equitans*, NeqN

Expressió i purificació de la NeqN

Per a la caracterització fisicoquímica del fragment N-terminal de la inteïna partida Neq pol, NeqN, es van optimitzar els protocols de producció i purificació d'aquesta proteïna. En material i mètodes es descriuen amb més detall les construccions exposades a continuació. En un primer intent, el gen que codifica per la NeqN es va clonar en el plasmidi pET-21-a(+) amb els enzims de restricció *Bam*HI i *Hind*III, obtenint T7-NeqN. Com es pot observar a la Figura 4.21 la T7-NeqN s'expressa a 37°C com una proteïna soluble. La realització d'un xoc tèrmic a elevades temperatures després del lisat cel·lular és un pas bastant habitual en protocols de purificació de proteïnes termòfiles, ja que amb aquest pas precipiten un gran nombre de proteïnes i la proteïna d'estudi es manté en la fracció soluble [163], [164]. Després de sotmetre durant 10 minuts la fracció soluble a 50°C i a 70°C s'observa, com és d'esperar, que la T7-NeqN es manté soluble en totes dues condicions, obtenint un grau de puresa superior en el tractament a 70°C (Figura 4.21). Per tal d'aconseguir un grau de puresa superior es va realitzar una cromatografia de bescanvi aniònic a pH 8.0. Tot i treballar a pHs molt llunyans al pI teòric, pI= 5.9, la NeqN s'unia dèbilment i inespecíficament a la columna, com a conseqüència les punxades no eren reproduïbles i el grau de puresa obtingut era baix (Figura 4.22)

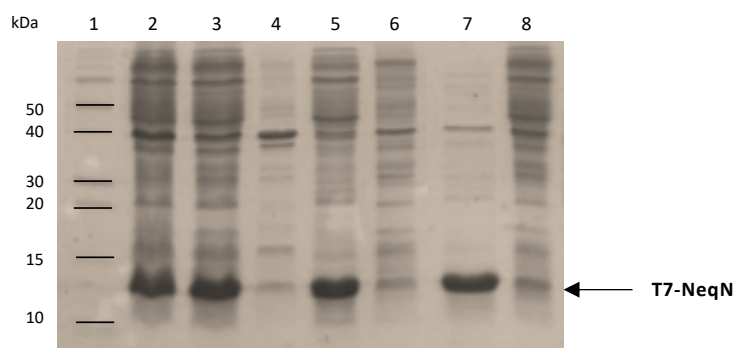


Figura 4.21. Anàlisi per SDS-PAGE de solubilitat de la T7-NeqN (12770.4 Da) expressada a 37°C. (1) Marcadors de massa molecular. (2) Extracte cel·lular després de 5 hores de la inducció amb 0.5 mM IPTG. (3) Fracció soluble després de la lisi cel·lular. (4) Fracció insoluble després de la lisi cel·lular. Anàlisi del xoc tèrmic realitzat durant 10 minuts a 50°C (5) Fracció soluble. (6) Fracció insoluble. Anàlisi del xoc tèrmic realitzat durant 10 minuts a 70°C (7) Fracció soluble (8) Fracció insoluble. El tampó utilitzat per a realitzar aquest assaig és: 20 mM fosfat, 0.5 M NaCl, a pH 8.0.

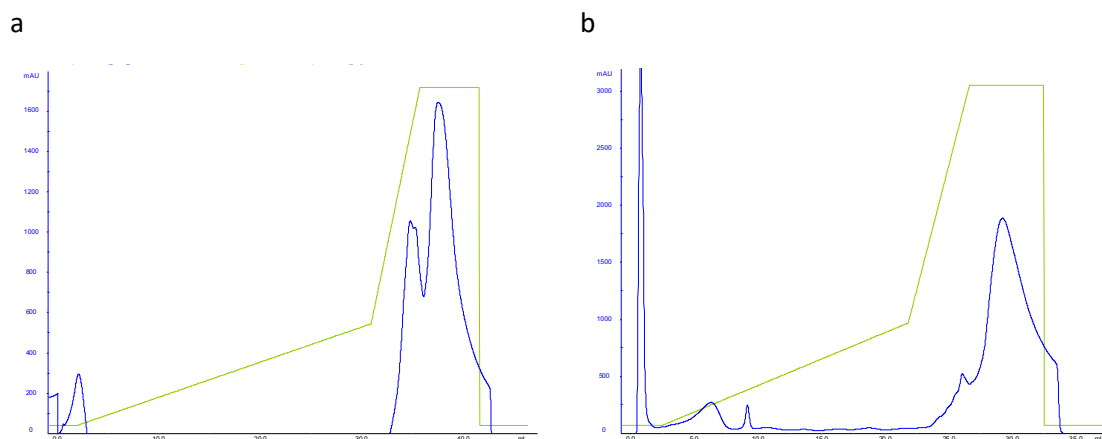


Figura 4.22. Purificació per cromatografia de bescanvi aniònic de la construcció T7-NeqN. Per dur a terme aquesta purificació es van utilitzar columnes HiTrap® Q HP (GE Healthcare) equilibrades prèviament amb 20 mM fosfat, 1mM EDTA, a pH 8.0. La fracció soluble després del xoc tèrmic a 70°C es va dialitzar enfront de 20 mM tampó fosfat, 1 mM EDTA a pH 8.0, es va filtrar i es va injectar a la FPLC. Per a l'elució de la proteïna es va realitzar un gradient lineal de 0 a 600 mM NaCl en 30 minuts. (a-b) Perfils d'absorbància a 280 nm obtinguts de dues punxades de T7-NeqN diferents realitzades en les mateixes condicions on es pot observar la manca de reproductibilitat.

Per intentar facilitar el procés de purificació, la NeqN es va expressar com una proteïna de fusió juntament amb la proteïna Trx en el seu extrem N-terminal, seguida d'una cua d'histidines (H6), de la diana de la proteasa del virus del mosaic del tabac (TEV) i una regió connectora de 13 residus, la qual incorpora la seqüència per l'epítot T7. La cua d'histidines permet purificar la proteïna per cromatografia d'afinitat Ni-NTA i la diana TEV permet eliminar fàcilment la Trx i la H6, les quals podrien interferir en estudis fisicoquímics. Per aconseguir aquesta construcció, el gen de la NeqN es va subclonar en el vector pBADM-20. La T7-NeqN es va purificar a partir de la fracció soluble, seguint el protocol descrit en Material i Mètodes. En aquest protocol i de manera resumida es va realitzar una primera purificació del precursor, TrxH6TEVT7-NeqN, mitjançant una cromatografia d'afinitat Ni-NTA, seguida de la digestió per la proteasa TEV, i finalment, una segona cromatografia d'afinitat Ni-NTA on s'esperava separar la T7-NeqN del material no desitjat. Com es pot observar en el gel d'acrilamida no es va obtenir un grau de puresa suficientment alt de la T7-NeqN, ja que tant la Trx com el material de partida coelueixen amb aquesta (Figura 4.23).

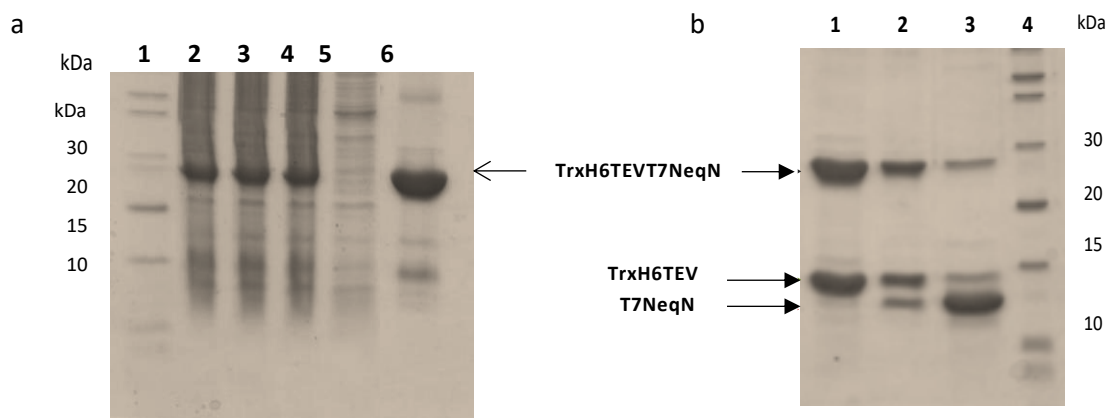


Figura 4.23. Anàlisi per SDS-PAGE del procés de purificació de la T7-NeqN per cromatografia d'afinitat Ni-NTA. (a) Anàlisi de la solubilitat de TrxH6TEVT7-NeqN 27149.7 Da expressada a 37°C i seguiment de la primera cromatografia d'afinitat Ni-NTA, per la qual es va utilitzar una columna HisTrap™ FF FPLC column, GE Healthcare, Chicago, ILL, USA. Les proteïnes es van separar amb un gradient de 5-250 mM d'imidazol en 30 minuts (tampó A: 20 mM tampó fosfat, 500 mM NaCl, 5 mM imidazol a pH 7.2; Tampó B: 20 mM tampó fosfat, 500 mM NaCl, 250 mM imidazol a pH 7.2). (1) Marcadors de massa molecular. (2) Extracte cel·lular després de 5 hores de la inducció amb 0.2% L-Arabinosa. (3) Fracció soluble. (4) Fracció insoluble. (5) Traves. (6) Elució. (b) Seguiment de la cromatografia d'afinitat Ni-NTA realitzada a pressió atmosfèrica després de la digestió amb la proteasa TEV. Per aquesta cromatografia s'utilitza la resina Ni-NTA equilibrada amb 20 mM tampó fosfat, 500 mM NaCl, 20 mM imidazol a pH 8.0. L'elució del precursor no digerit i de la TrxH6TEV es realitza rentant la columna amb 20 mM tampó fosfat, 500 mM NaCl, 250 mM imidazol a pH 8.0 (1) Elució. (2) Rentat. (3) Traves. (4) Marcadors de massa molecular.

Atès que no es va obtenir el grau de puresa desitjat, es va optar per expressar la NeqN com una proteïna de fusió, però en aquest cas sense la Trx. Per a obtenir aquesta construcció es va subclonar el gen que codifica per H6TEVT7-NeqN del plasmidi pBADM20 al vector d'expressió pET-22b(+). Com s'ha comentat anteriorment en la introducció d'aquesta tesi són nombroses les evidències que demostren que els residus que es troben a la vora dels llocs d'unió exteïna-inteïna tenen un paper important en l'activitat de les inteïnes [33]–[37]. Per aquesta raó es van introduir 7 residus de l'exteïna N salvatge en la regió flanquejant a la Ser1. Per aconseguir la NeqN amb un elevat grau de puresa i sense la cua d'histidines, es va millorar el protocol de purificació de la T7-NeqN a partir del precursor TrxH6TEVT7-NeqN. De manera resumida, la millora en el protocol de purificació de la NeqN va consistir a augmentar el grau de puresa de la proteïna i el seu rendiment. Per augmentar el grau de puresa es van modificar diferents punts del protocol: 1- augmentar la concentració inicial d'imidazol en la solució de resuspensió, disminuint d'aquesta manera interaccions inespecífiques; 2- disminuir la velocitat del gradient de la cromatografia d'afinitat Ni-NTA; 3- Disminuir la concentració de sals en el tampó de purificació per evitar l'elució de part del precursor no digerit o de H6-TEV. Per tal de millorar el rendiment, es va optimitzar la digestió per part de la proteasa TEV, afegint una petita

>NeqN

GAMASGSFKVIYGDSIMDTEIEVIENGIKKKEKLSDLFNKYYAGFQIGEKHYAFPPDLYVYDGERWV
 KVYSIIKHETETDLYEINGITLSANHLVLSKGNWVKAKEYENKNN

Figura 4.25. Seqüència polipeptídica de la construcció NeqN. La NeqN (vermell) en el seu extrem N-terminal presenta 7 residus nadius de l'exteïna N de *N.equitans* (taronja) i set residus més residuals producte del clonatge.

Estudi in silico de la NeqN

Per analitzar *in silico* la probabilitat de què la NeqN es comportés com una IDP es van utilitzar una sèrie d'algoritmes predictius. Per a validar els mètodes de predicció es va creure adient analitzar simultàniament la inteïna NpuN, perquè existeixen evidències experimentals que demostren que no presenta un plegament ben definit [57].

En primer lloc, es va estudiar la composició aminoacídica de la NeqN (Figura 4.26-a). Es coneix que les IDPs tenen una composició d'aminoàcids característica, contenen més residus que promouen el desordre (Lys, Ser, Asp, Pro i Gln), que no pas residus que promouen l'ordre (Cys, Trp, Ile, Tyr, Phe, Leu, His, Val, Asn i Met) [165]. La comparació de la composició aminoacídica de la NeqN amb la base de dades Disprot 3.4 indica que la NeqN no pertany al grup de proteïnes desestructurades. D'altra banda, com s'observa en la Figura 4.26-b la corba de la funció de distribució acumulada (CDF) obtinguda per la NeqN mostra que aquesta es troba en el rang de proteïnes ordenades. Per últim, el gràfic de càrrega-hidrofobicitat (Figura 4.26-c), on es representa la hidrofobicitat normalitzada Kyte-Doolittle enfront de la càrrega neta mitjana per a classificar les proteïnes en ordenades o desordenades, també localitza la NeqN com una proteïna ordenada.

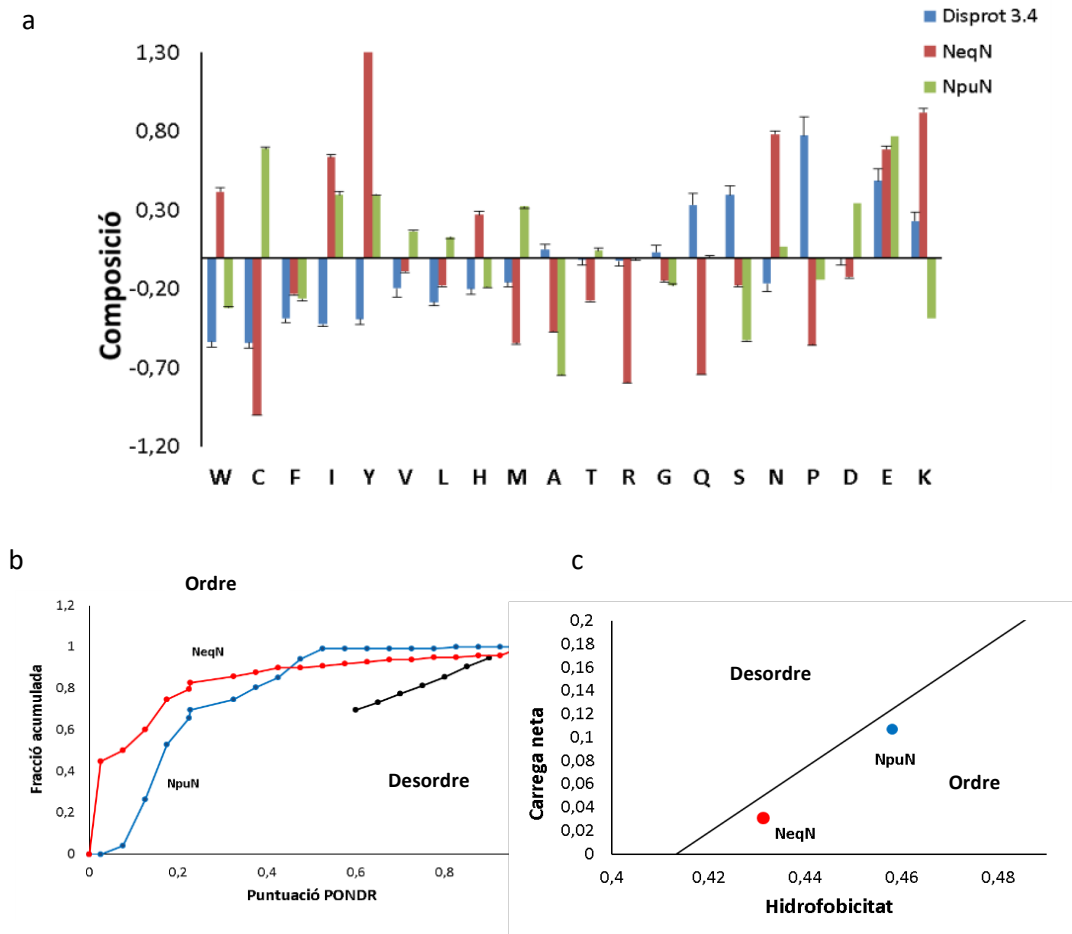


Figura 4.26. Anàlisi bioinformàtica de la NeqN i comparació amb la NpuN. (a) Comparació de la composició d'aminoàcids de la NeqN (Barres vermelles), NpuN (barres verdes) i Disprot 3.4 (Barres blaves) amb el banc de dades PDB_Select_25 utilitzant Composition Profile. Els valors de cada aminoàcid segueixen l'escala de flexibilitat Vihinen's, els valors positius indiquen enriquiment respecte a la base de dades PDB_Select_25, en canvi, els valors negatius indiquen una disminució. (b) Anàlisi de la funció de distribució acumulada (CDF). La línia negra indica el límit entre proteïnes plegades (part de dalt) i desplegadas (part de sota). Tant la NeqN (línia vermella) com la NpuN (línia blava) es troben en la regió plegada. (c) Gràfic de càrrega-hidrofobicitat on es representen la NeqN (vermell) i la NpuN (blau). La línia negra ($\langle R \rangle = 2.785$ i $\langle H \rangle = 1.151$) separa el gràfic en dues regions, la part superior que correspon a les proteïnes desordenades i la part inferior que correspon a les proteïnes ordenades.

Els resultats obtinguts són prometedors, ja que tant la composició d'aminoàcids, com l'anàlisi CDF i el gràfic de càrrega-hidrofobicitat, convergeixen en què la NeqN presenta estructura. Però si s'observen els resultats obtinguts per la NpuN, tot i que presenten petites diferències, es pot veure que l'estudi *in silico* també prediu que aquesta es troba plegada, en canvi, com ja s'ha comentat anteriorment els resultats experimentals demostren just el contrari.

Per a obtenir més informació sobre l'estat de plegament de la NeqN es va realitzar un estudi fisicoquímic.

Estudi de la solubilitat in vivo

Es coneix que l'estat plegat d'una proteïna la protegeix de l'agregació [166], i que en la majoria dels casos la formació d'agregats és dona a partir d'intermediaris parcialment o totalment desplegats, els quals actuen com a nuclis d'agregació [167]. Com a conseqüència, com a normal general, la solubilitat d'una proteïna aporta informació de l'estat natiu d'aquesta.

Com s'ha pogut veure anteriorment, en la purificació del precursor H6TEV(wt)-NeqN, aquest s'expressa en *E. coli* com una proteïna soluble, el que suggereix que probablement es troba correctament plegat. Per demostrar que la solubilitat és un bon indicador del grau d'estructura, es van comparar les solubilitats de H6TEV(wt)-NeqN i de H6TEV-NpuN, després de la seva expressió a 37°C (Figura 4.27). Com era d'esperar i amb concordança amb els resultats obtinguts per Muir i col·laboradors [57], la construcció H6TEV-NpuN sota les mateixes condicions d'expressió que la H6TEV(wt)-NeqN, es troba en la fracció insoluble formant cossos d'inclusió. En canvi, la construcció que conté la NeqN es troba en la fracció soluble. Aquest resultat suggereix que la NeqN es troba correctament plegada.

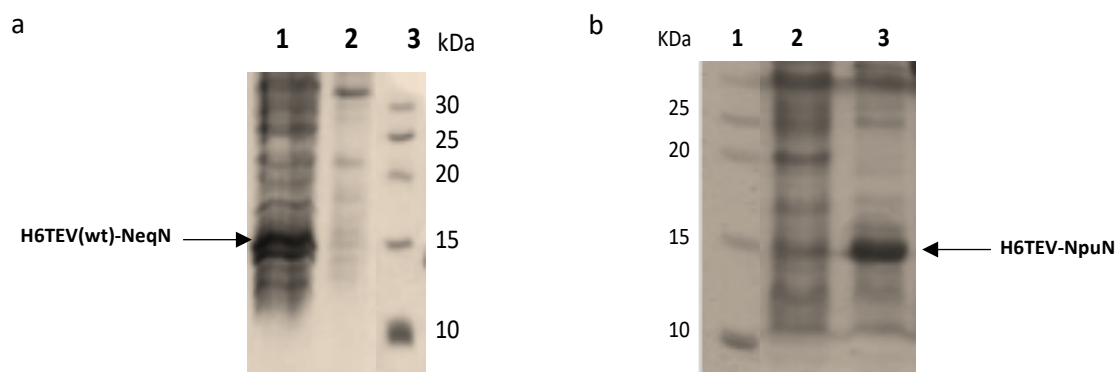


Figura 4.27. Anàlisi per SDS-PAGE de la solubilitat de les construccions H6TEV(wt)-NeqN i H6TEV-NpuN (14841.6 Da i 15258.15 Da, respectivament). Després de l'expressió a 37°C, les cèl·lules es lisen, per centrifugació es separa la fracció soluble de la insoluble (a) Estudi de solubilitat de H6TEV(wt)-NeqN. (1) Fracció soluble; (2) Fracció insoluble després de la lisi cel·lular (3) Marcadors de massa molecular. (b) Estudi de solubilitat de H6TEV-NpuN. (1) Marcadors de massa molecular; (2) Fracció soluble (3) Fracció insoluble després de la lisi cel·lular.

Proteòlisi limitada

La proteòlisi limitada és un mètode per a determinar la fracció de proteïnes plegades i desplegadas sota unes determinades condicions. Aquest mètode es basa principalment en la diferent susceptibilitat de l'estat plegat i desplegat de les proteïnes a la proteòlisi. Perquè la proteòlisi sigui eficient es requereix que el substrat es trobi totalment o parcialment desplegat, o en una regió exposada a solvent. De manera que les proteïnes, en funció del seu grau de plegament tenen diferents susceptibilitats a la proteòlisi [140].

Per estudiar mitjançant aquesta aproximació, el grau de plegament de la NeqN es va determinar la seva resistència a la digestió per part de la proteïnasa K sota condicions fisiològiques. Per a la validació de l'assaig es varen utilitzar controls interns: com a control negatiu, una proteïna amb una estructura ben definida i plegada com la variant de la ribonucleasa pancreàtica humana NLSPE5, i com a control positiu la NpuN la qual no presenta una estructura terciària ben definida [57]. En la Figura 4.28 es pot observar com tant la inteïna NeqN com NLSPE5 presenten una resistència parcial a l'acció de la proteasa al llarg del temps. Al contrari, i com és d'esperar, la inteïna NpuN és totalment degradada. El patró de bandes obtingut després de la digestió de la NeqN suggereix que aquesta tot i tenir majoritàriament una estructura ben definida presenta una petita regió desestructurada, probablement als extrems. L'anàlisi per MALDI-TOF va permetre estimar que la regió desplegada estava composta per 19 residus de l'extrem N-terminal (14 residus de l'exteïna i 5 de la inteïna) i 21 residus de l'extrem C-terminal de la construcció (Figura 4.29). La manca d'estructura de l'extrem N-terminal podria tenir un rol important en els primers passos de la reacció de *trans-splicing*, ja que confereix flexibilitat. Per una altra banda, la falta d'estructura de l'extrem C-terminal dividiria la NeqN en dues regions o lòbuls, com en el cas de la NpuN [57]. En aquest cas, el lòbul 2 desestructurat seria molt més petit aproximadament un 20%, que en el cas de la NpuN, on la regió desestructurada correspondria al 50% de la seqüència.

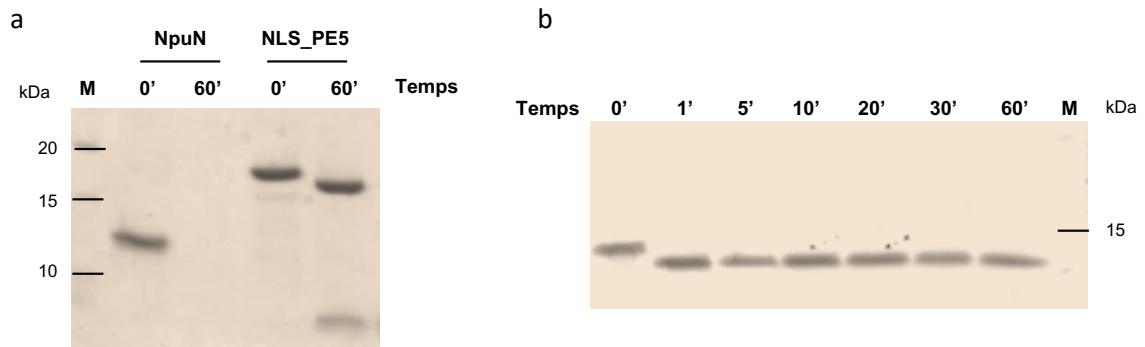


Figura 4.28. Anàlisi per SDS-PAGE de la resistència de la NeqN a la digestió per part de la proteïnasa K. (a) Proteòlisi de la NpuN (11856.4 Da) i de NLS-PE5 (15722.8 Da). (b) Cinètica de la proteòlisi de NeqN. M: Marcadors de massa molecular. 10 μ M de substrat (NeqN, NpuN o NLS-PE5) s'incuba amb 10 μ g/ml de proteïnasa K a temperatura ambient. Totes les mostres es van preparar amb 50 mM fosfat, 150 mM NaCl a pH 8.0.

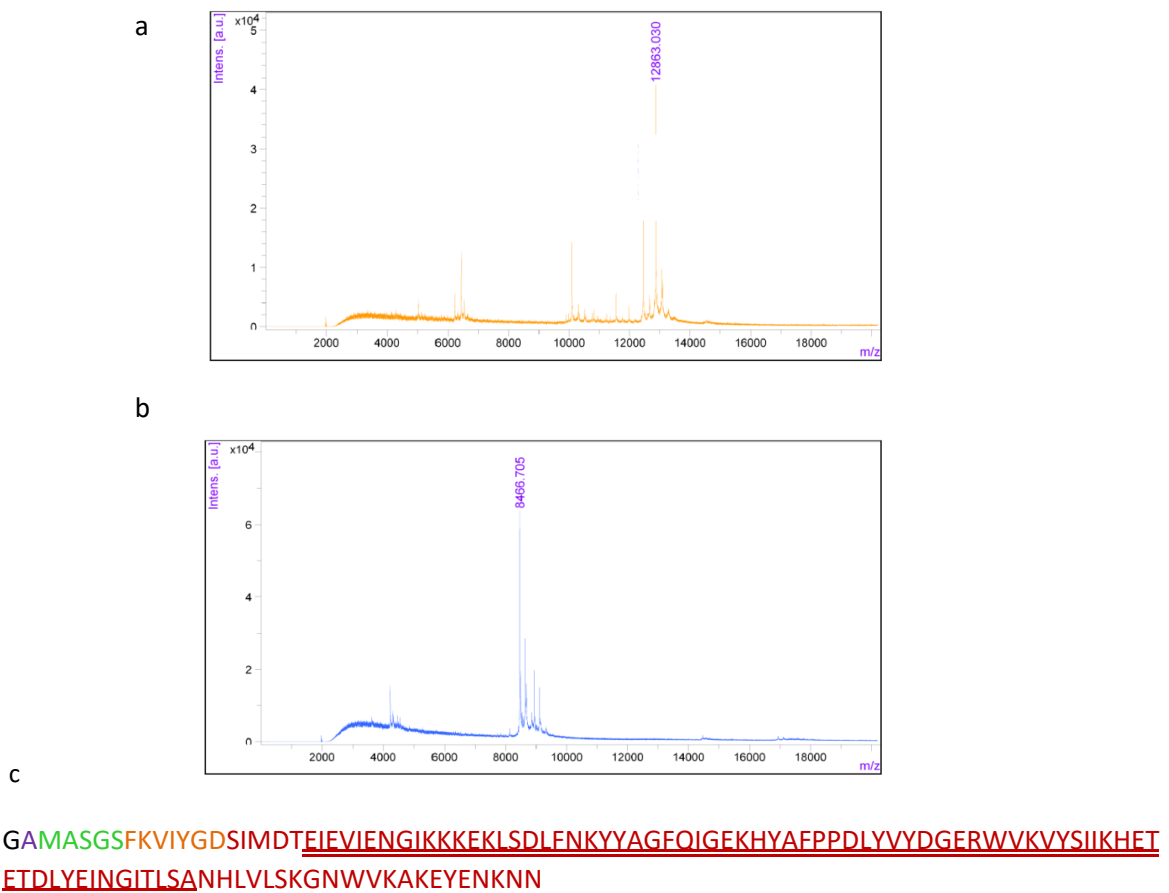


Figura 4.29. Anàlisi per MALDI-TOF de la digestió de la NeqN. (a) Temps 0 (b) i 1 h de proteòlisi (c) Seqüència NeqN substrat s'indica la regió resistent a la proteòlisi.

La proteòlisi limitada, d'altra banda, també permet determinar de manera senzilla l'estabilitat termodinàmica de les proteïnes. Mitjançant electroforesi SDS-PAGE s'analitza la susceptibilitat a la digestió proteolítica de la proteïna a diferents concentracions d'urea.[140], [168].

Per a realitzar aquest mètode diferents fraccions de la NeqN pura es van deixar en presència de concentracions creixents d'urea a 25°C durant tota la nit, posteriorment es van digerir durant 1 hora amb proteïnasa K. i van ser analitzades per 15% SDS-PAGE (Figura 4.30). Com a control positiu es va utilitzar la proteïna NLSPE5, la qual es va tractar de la mateixa manera que la mostra problema, però en aquest cas, només es va realitzar la proteòlisi a la concentració superior d'urea, 4M. La presència d'urea en el gel provoca eixamplament i difusió de les bandes de NeqN, implicant una quantificació acurada i per tant, un càlcul precís de la concentració d'urea necessària per a tenir la meitat de les molècules desplegadas (C_m). Malgrat això, en la Figura 4.30-a s'observa que el patró de bandes obtingut després de la digestió és el mateix al llarg de tot el rang d'urea assajat. Aquest resultat indica que la NeqN presenta una gran part de la seva seqüència correctament plegada, inaccessible a la proteasa, i resistent a la desnaturalització en presència de 4M urea. Aquesta resistència a la desnaturalització per agents caotròpics és un tret característic de moltes proteïnes hipertermofíliques [120].

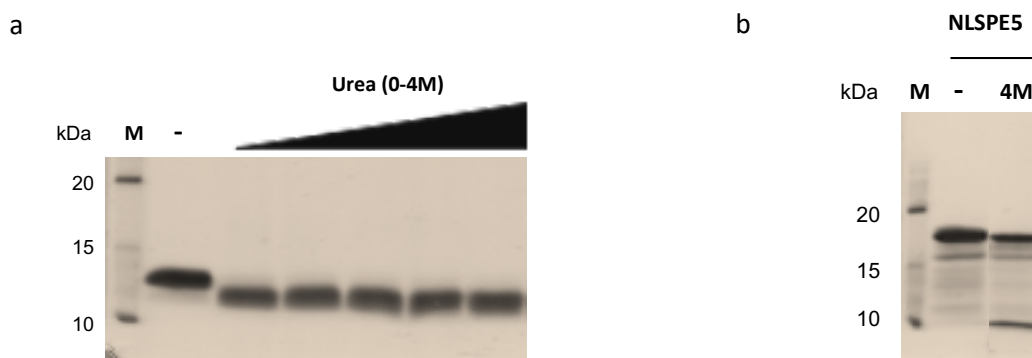


Figura 4.30. Anàlisi per SDS-PAGE *in vitro* de l'estabilitat conformacional de la proteïna partida NeqN mitjançant proteòlisi limitada. La NeqN pura va ser equilibrada tota la nit amb diferents concentracions d'urea (rang de 0 M a 4 M d'urea) i posteriorment digerida amb proteïnasa K. La mostra (-) es tracta de la proteïna d'estudi en absència d'urea i sense proteïnasa K. La proteïna NLSPE5 es van utilitzar com a control. (a) Proteòlisi limitada de la NeqN. (b) Proteòlisi limitada de NLSPE5.

Determinació de l'estabilitat tèrmica

En un primer intent de determinar l'estabilitat tèrmica de la construcció NeqN es varen registrar les corbes de desnaturalització tèrmica a pH 7.0 i pH 4.5 mesurant el canvi d'absorbància a 280 nm. En cap de les dues condicions s'observa una transició entre l'estat plegat i el desplegat (Figura 4.31). La qual cosa suggereix, tenint en compte els resultats obtinguts, que per desnaturalitzar la NeqN s'ha d'incubar a temperatures per sobre dels 86°C, que és la temperatura màxima que es podia assolir amb el bany.

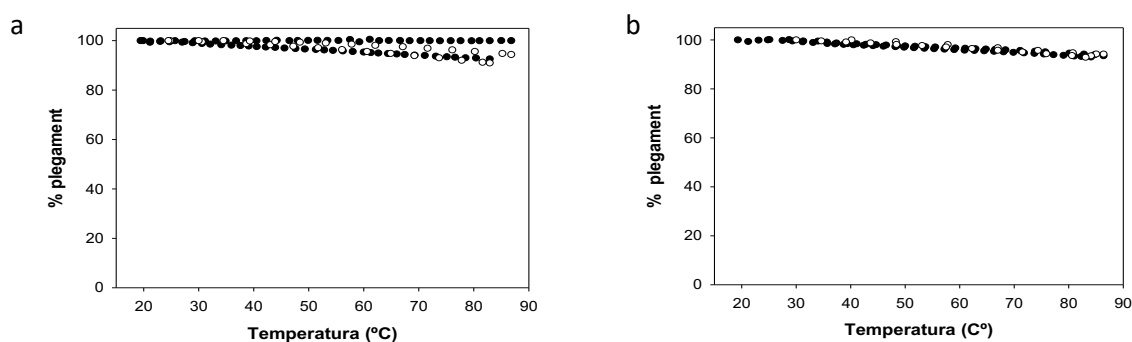


Figura 4.31. Corbes normalitzades de desplegament (●) i replegament (○) per temperatura de la NeqN a diferents pHs. (a) Corbes realitzades pH 7.0 (b) Corbes realitzades a pH 4.5. Els espectres es van obtenir en un rang de temperatures entre 20-80°C amb increments de 2°C per a les corbes de desnaturalització i amb decrements de 5°C per a obtenir les corbes de replegament. La concentració de proteïna en cada assaig és de 20 μM. Per a realitzar l'estudi d'estabilitat a pH 7.0 es va utilitzar 20 mM fosfat, 50 mM NaCl a pH 7.0 i 50 mM citrat, 50 mM NaCl a pH 4.5 per a l'estudi a pH 4.5.

Com a conseqüència, es va optar per estudiar l'estabilitat tèrmica per Calorimetria diferencial d'escombrat (DSC). Aquesta tècnica permet escalfar una solució aquosa de manera gradual i treballar a temperatures per sobre del punt d'ebullició de l'H₂O (>100°C). El termograma obtingut (Figura 4.32) mostra que NeqN presenta un desplegament induït per temperatura irreversible a pH 8.0, de manera que no es poden calcular els paràmetres termodinàmics, ΔH , ΔS i ΔG , en canvi, però sí que es pot estimar la T_m que es trobaria al voltant de 105°C.

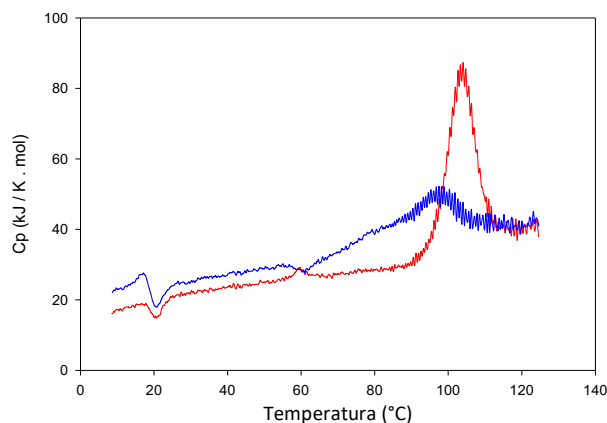


Figura 4.32. Termograma obtingut per DSC del procés de desnaturalització de la NeqN. Experiments realitzats a una concentració de 26 μM de proteïna amb 20 mM tampó fosfat, 150 mM NaCl, 1 mM EDTA a pH 8.0 a una velocitat d'escombrat de 1.5°C/min. Línia vermella: traça del primer escalfament; Línia blava: traça del segon escalfament.

Caracterització per dicroïsm circular (CD)

En la Figura 4.33-a es mostra l'espectre de dicroïsm circular (CD) obtingut de la NeqN en la regió de l'UV-llunyà (190-250 nm). Amb la concentració de proteïna utilitzada per a realitzar l'assaig el senyal de l'instrument es saturava a longituds d'ona curtes. Com a conseqüència, el senyal de CD només es va poder mesurar fins a 200 nm. Les hèlixs α tenen un espectre de CD molt característic en la regió de l'UV llunyà, constituït per un doble mínim a 222 nm i a 208-210 nm i un màxim més intens a 191-193 nm. Les estructures tipus β solen presentar un únic pic negatiu i un únic pic positiu amb intensitats inferiors que les hèlixs α . El mínim es troba normalment entre 210 i 225 nm i el pic positiu, més intens que el negatiu, es troba entre 190 i 200 nm. Les proteïnes sense estructura, normalment, presenten una forta banda negativa a prop de 200 nm. Tot i ser incomplet l'espectre de l'UV-llunyà obtingut mostra que la proteïna NeqN es troba correctament plegada. En l'espectre obtingut (Figura 4.33-a) s'observa un mínim d'el·lipticitat entre 220 nm i 200 nm característic de proteïnes que contenen estructura β , α i una part desordenada. Amb l'espectre obtingut en la regió del llunyà i utilitzant l'algoritme K2D [169], [170] es va poder estimar el contingut d'estructura secundària de la NeqN, 21% d'hèlix α , un 33% en beta β i un 46% en random coil. Un tret característic d'aquest espectre és la banda positiva que s'observa a una $\lambda \sim 230$ nm, la qual és característica de proteïnes que presenten els aminoàcids aromàtics en un ambient ordenat [171].

Els espectres CD en la regió de l'UV-proper (250-350 nm), són sensibles a canvis en estructura terciària. A aquestes longituds d'ona els cromòfors són els aminoàcids aromàtics i els enllaços disulfurs quan es troben immobilitzats en un entorn estructural asimètric, normalment l'estat

natiu de les proteïnes. El Trp presenta una banda d'absorció molt intensa a 282 nm, la qual pot desplaçar-se cap al vermell a causa de la formació d'un pont d'hidrogen en el qual participa el grup NH del grup indol. En la Figura 4.33-b, es mostra l'espectre de CD en la regió del proper obtingut per la NeqN. En aquest, es mostra una banda positiva molt intensa propera als 300 nm, que indica que la NeqN presenta una estructura nativa compacta on almenys un dels dos triptòfans es troba en un ambient ordenat probablement establint ponts d'hidrogen.

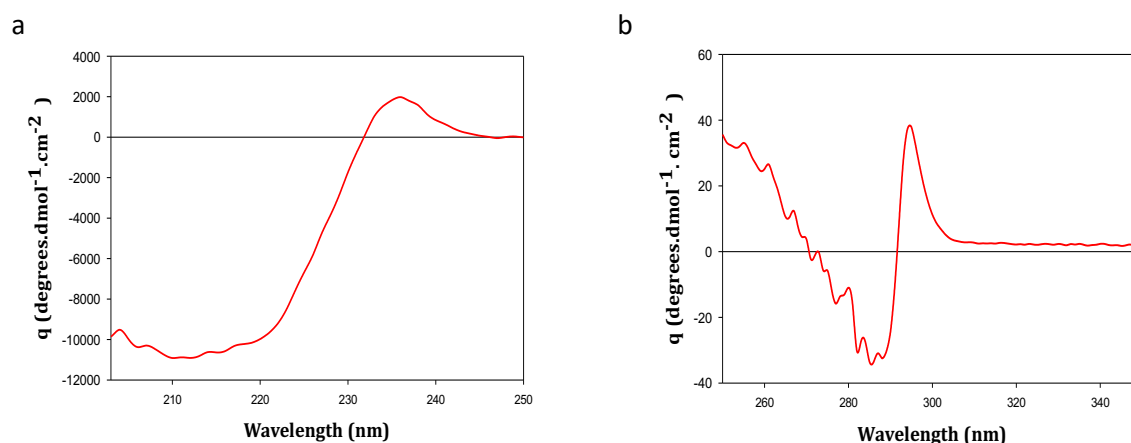


Figura 4.33. Espectres de CD de la proteïna NeqN. (a) Espectre de CD en la regió del llunyà (200-250 nm) obtingut a partir d'una solució de proteïna a una concentració final de 125 μM . (b) Espectre en la regió del proper (250-300 nm), utilitzant una solució preparada a una concentració de 18 μM de proteïna. En tots els casos la proteïna ha està dissolta amb 20 mM tampó fosfat, 150 mM NaCl, 1 mM EDTA a pH 8.0.

Unió a ANS

L'ANS s'utilitza per a detectar regions hidrofòbiques exposades en l'estat natiu de les proteïnes o en intermediaris de plegament, ja que el grup no polar anilino-naftale del ANS s'uneix a les regions hidrofòbiques exposades al solvent [172], [173]. Per estudiar la unió de l'ANS a la NeqN, es va preparar una solució d'aquesta a una concentració final de 25 μM amb 50 mM tampó fosfat, 50 mM NaCl a pH 7.0. Com a control positiu es va utilitzar l'Apop Δ Pro, la qual s'ha demostrat que uneix ANS a causa de l'exposició de regions hidrofòbiques a solvent [142]–[144]. Com es pot observar en la Figura 4.34, l'espectre d'emissió de la NeqN és el mateix que el de l'ANS lliure, la qual cosa indica que la NeqN no exposa regions hidrofòbiques al solvent.

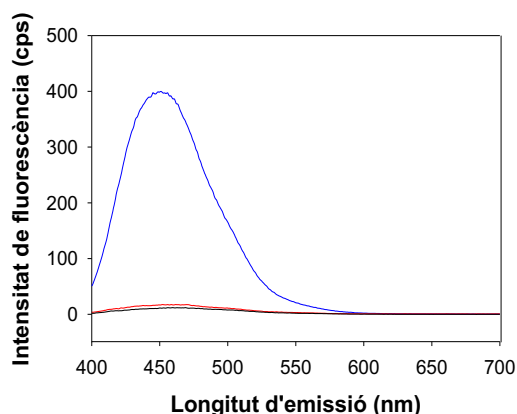


Figura 4.34. Espectre d'unió a ANS de la NeqN. En negre s'indica l'espectre obtingut de l'ANS lliure; en vermell l'espectre obtingut de la NeqN + ANS i en blau l'espectre de l'ApopΔPro + ANS (Control positiu).

SEC i ultracentrifugació analítica

La NeqN pura s'elueix, com un únic pic en una columna de cromatografia d'exclusió molecular Sephadex-75 HR-10/30, amb un volum d'elució (V_e) que varia entre 10–11.7 ml en funció de la concentració de proteïna (Figura 4.35). Existeix una relació proporcionalment inversa entre la concentració de proteïna i el volum d'elució, és a dir a concentracions més altes de proteïna el volum d'elució és més petit, i a l'inrevés. Aquesta variació del volum d'elució en funció de la concentració de proteïna, suggereix que la NeqN interacciona de manera específica i dinàmica amb ella mateixa donant lloc a espècies oligomèriques.

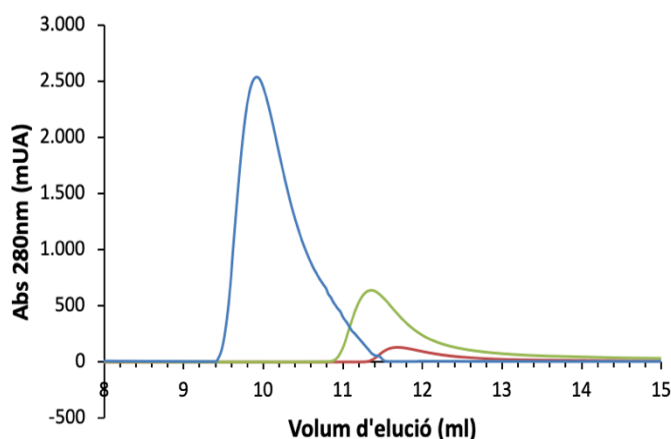


Figura 4.35. Caracterització de la NeqN per cromatografia d'exclusió molecular (SEC). Sobreposició dels diferents perfils d'absorbància a 280 nm obtinguts de l'anàlisi per SEC de la NeqN a diferents concentracions: 168 μ M, 62 μ M i 13 μ M. Els diferents perfils es representen en blau, verd i vermell respectivament.

Per determinar la naturalesa del pic corresponent a una solució de NeqN preparada amb 20 mM tampó Tris-HCl, 150 mM NaCl, 1 mM EDTA a una concentració de 325 μM (concentració aproximada utilitzada posteriorment per a l'obtenció de cristalls), es va estimar la seva massa molecular per SEC (Figura 4.36-a). Després de calibrar la columna de cromatografia d'exclusió molecular Sephadex-75 HR-10/30 amb diferents proteïnes globulars de massa molecular coneguda, es va observar que un volum d'elució d'aproximadament 10 ml corresponia a una massa molecular estimada de 28700 Da (Annex F). Paral·lelament, la massa molecular de la NeqN a una concentració de 182 μM (2.34 mg/ml) es va determinar per ultracentrifugació analítica. En la Figura 4.36-b es mostra el perfil de sedimentació obtingut on es veu un pic majoritari amb un coeficient de sedimentació corregit a aigua i a 20°C ($S_{20,w}$) de 2.41 S i una altra espècie minoritària que sedimenta a 4.37S. Les masses moleculars estimades són 27700 Da per a l'espècie principal i 67700 Da per la segona.

Les dues tècniques convergeixen en què la massa molecular estimada per una solució de NeqN és d'uns 28000Da aproximadament, aquesta massa s'ajusta a la massa teòrica calculada per un dímer format per dues molècules de NeqN (12861.5 Da). S'ha de tenir en compte que per a fer els càlculs per a obtenir la massa molecular experimental, aquestes tècniques assumeixen que la NeqN és una proteïna globular. Probablement, la NeqN, com tots els membres de la família d'inteiïnes, presentarà un plegament en ferradura majoritàriament format per cadenes β , el qual no és tant compacte com una proteïna globular, i per tant, té un radi més gran, el que pot donar lloc a una sobreestimació la massa real de la inteïna NeqN.

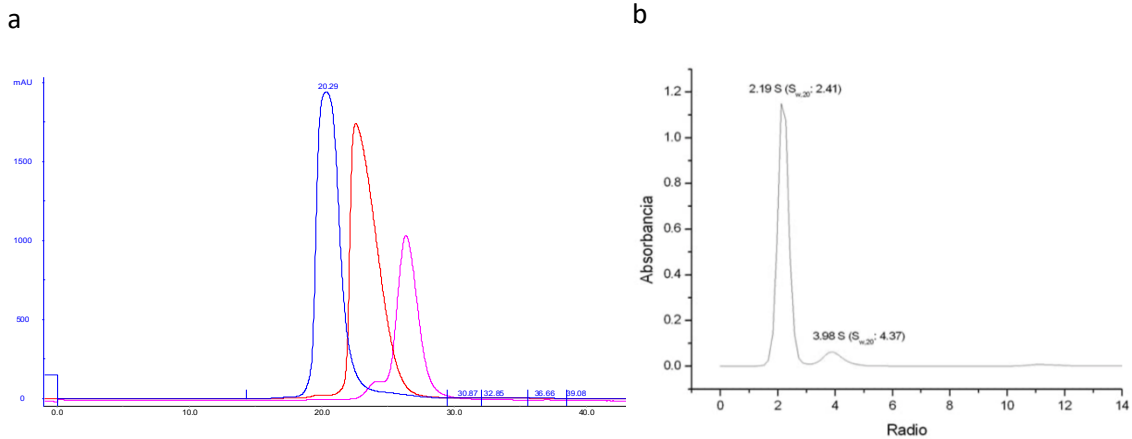


Figura 4.36. Determinació experimental de la massa molecular de la NeqN. (a) Anàlisi per SEC de la inteïna N, NeqN (línia blava), de la inteïna C, NeqC29 (línia rosa) i del complex (línia vermella). La NeqN (12861.5 Da) tot i tenir una massa molecular més petit que el complex (19001.3 Da) presenta un temps d'elució, T_e , inferior ($T_e = 20.3$ min) Aquest temps d'elució correspon a un massa molecular de 28700 Da. Per estimar la massa molecular la columna ha estat prèviament calibrada amb proteïnes globulars de massa molecular coneguda. (b) Dades de l'equilibri de sedimentació amb una corba ajustada per a la solució de NeqN a una concentració de $182 \mu\text{M}$ a 45000 rpm i a 20°C . La massa molecular estimada de la NeqN és de 27700 Da per l'espècie majoritària.

Fins aquí, totes les tècniques emprades per a la caracterització fisicoquímica de la NeqN indiquen que, a diferència de la NpuN, es troba correctament plegada. Per una altra banda, els resultats obtinguts per SEC i ultracentrifugació assenyalen que la NeqN aïllada a elevades concentracions es troba majoritàriament en un estat dimèric, l'equilibri del qual es troba fortament influenciat per la concentració de proteïna. Assumint que la NeqN en solució es troba en equilibri entre dos estats, monomèric i dimèric, la T_m calculada a una concentració de $26 \mu\text{M}$, correspondria a la temperatura mitjana de desplegament mesurada en una solució heterogènia, monòmer-dímer. Com s'ha dit abans, l'equilibri entre les dues poblacions està fortament influenciat per la concentració de proteïna, de manera que a més concentració de proteïna més dímer. En la Figura 4.37, es mostren els termogrames obtinguts per DSC a concentracions creixents de NeqN. En el cas que el dímer estabilitzés la NeqN, seria d'esperar que a concentracions més elevades de proteïna, la transició de l'estat plegat al desplegat es desplaçés cap a temperatures més altes. Els termogrames obtinguts mostren que la transició entre l'estat plegat i desplegat, la T_m , no es veu afectada per l'increment de concentració, la qual cosa indica, que el dímer no estabilitza l'estructura de la NeqN.

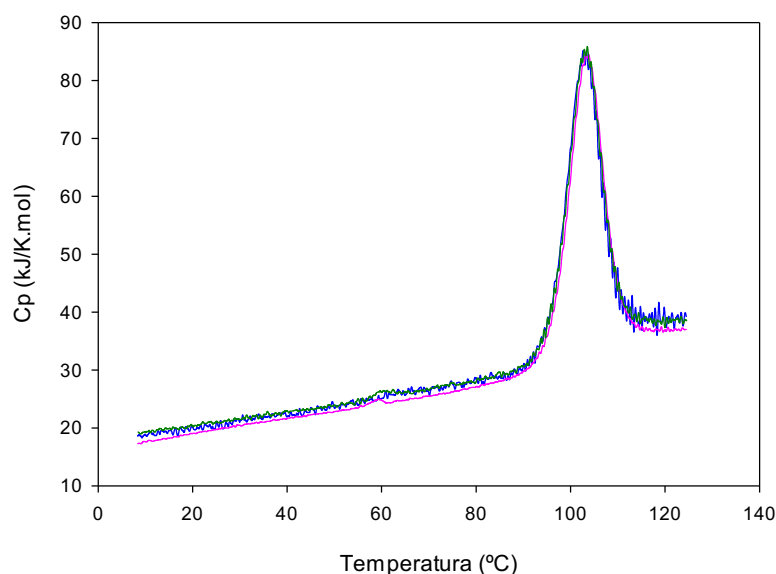


Figura 4.37. Estudi de l'efecte de la concentració en la transició de l'estat plegat-desplegat de la NeqN per DSC. Totes les concentracions es varen preparar amb 20 mM tampó fosfat, 150 mM NaCl, 1 mM EDTA a pH 8.0. Termograma blau: 26 μ M; Termograma verd: 54 μ M; Termograma rosa: 140 μ M.

4.2.2. Resolució de l'estructura tridimensional per cristal·lografia i difracció de raig X

L'estructura tridimensional de la NeqN es va resoldre per cristal·lografia i difracció de raig X. Es van fer intents previs per resoldre l'estructura per RMN però la manca d'estructura dels extrems N- i C- terminals demostrada per proteòlisi i l'equilibri existent entre la forma monomèrica i dimèrics de la NeqN, podrien explicar la dificultat d'obtenir espectres de ressonància interpretables.

La purificació de la NeqN per a obtenir cristalls es va realitzar seguint el mateix protocol que per a la seva caracterització però amb dues petites modificacions. La primera consistia a canviar el tampó fosfat per Tris, ja que el tampó fosfat cristal·litza i interfereix en l'obtenció de cristalls de proteïna. Aquest canvi, no afecta l'estat natiu de la NeqN (Annex G). L'altra modificació va ser que després del procés de purificació es van introduir dos passos més, un de purificació, en el qual es passava la NeqN per cromatografia d'exclusió molecular, i un de concentració de la proteïna.

S'han resolt dos tipus d'estructures de la NeqN. D'una banda l'estructura de la NeqN sola i de l'altra la NeqN formant complex amb la NeqC. En aquest punt es parlarà de l'estructura de la NeqN aïllada, més endavant es discutirà l'estructura obtinguda per la NeqN formant complex. L'estructura de la NeqN es va resoldre amb una resolució de 2.6 Å (Taula 4.5) a partir de cristalls monocíncics per reemplaçament molecular utilitzant com a model l'estructura del complex

NeqN/NeqC, la qual es va resoldre amb antelació utilitzant el programa ARCIMBOLDO_BORGES amb una resolució de 1.2 Å (Apartat 4.3.3 de Resultats i discussió).

Taula 4.5. Col·lecció de dades i estadístics del refinament.

Col·lecció de dades ^a	NeqN
Grup espacial	C2
Dimensions de la cel·la	
a,b,c (Å)	109.93, 67.72, 77.84
α, β, γ (°)	90.00, 129.91, 90.00
Reflexions úniques	13,172 (970)
Resolució (Å)	51.01-2.61 (2.68-2.61)
Longitud d'ona	1.1478
R _{sym} (%) ^b	0.05 (0.86)
$\langle I/\sigma I \rangle$	15.4 (1.8)
CC _{1/2} ^c	0.99 (0.72)
Completesa	98.90 (98.70)
Redundància	3.0 (3.2)
Estadístics de refinament	
Resolució	51.01-2.61
No. de reflexions	12.591
R _{crist} (%) ^d	20.79
R _{lliure} (%) ^e	25.74
No. de residus	180
Molècules de solvent	7
Contingut de solvent (%)	55
Àtoms no H	2.850
Mitjana factor B (Å ²)	61.84
Coordinada d'error (Å) ^f	0.29
RMSD bond (Å)	0.02
RMSD angles (°)	1.95

^a Els valors entre parèntesis corresponen a la cel·la de resolució més elevada

^b $R_{sym} = \frac{\sum_{hkl} \sum |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum I_i(hkl)}$, on $I_i(hkl)$ és la intensitat d'una observació i $\langle I(hkl) \rangle$ és el valor mitjà d'observacions per a una única reflexió.

^c CC_{1/2} és el percentatge de la correlació entre intensitats de bases de dades aleatòries.

^d $R_{crist} = \frac{\sum_h |F_o(h) - F_c(h)|}{\sum_h |F_o(h)|}$, on F_o i F_c són les amplituds del factor estructura observades i calculades, respectivament.

^e R lliure va ser calculat amb el 5% de les dades, les quals van ser excloses del refinament.

^f es basa en la màxima probabilitat

Cada cristall conte quatre monòmers per unitat asimètrica que estan organitzats en dos dímers amb una simetria cíclica similar a C2 (Figura 4.38). D'acord amb els resultats obtinguts per cromatografia d'exclusió molecular i ultracentrifugació analítica, el dímer és l'estat dominant de la NeqN en solució a les concentracions de treball (Figura 4.36). Anàlisis DSSP i pro-origami [145], [174] demostren que els quatre monòmers tenen la mateixa tipologia, excepte per la presència de l'hèlix α (residus 76-85), la qual només s'ha pogut traçar en els monòmers C i D. Pels quatre monòmers la densitat electrònica de l'extrem C-terminal és pobre, com a conseqüència 21 residus dels monòmers A i B i 12 dels monòmers C i D no es van poder modelar. Aquests resultats corroboren els obtinguts per proteòlisi (Figura 4.29).

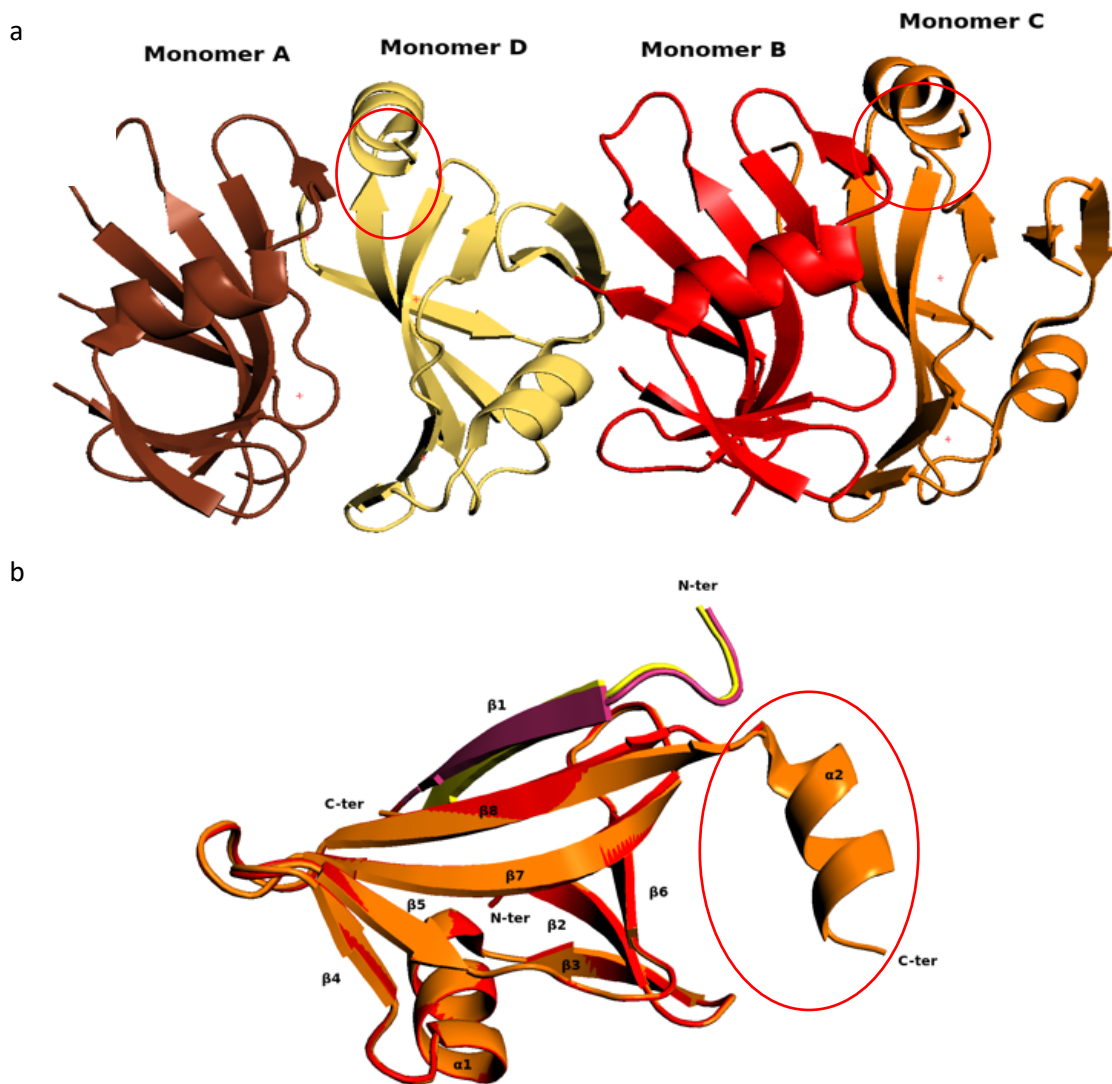


Figura 4.38. Estructura tridimensional obtinguda per cristal·lografia de la NeqN (5oxw). (a) Representació dels quatre monòmers per unitat asimètrica obtinguts per cristal·lografia. S'observa que presenten gairebé la mateixa topologia excepte per l'hèlix α de l'extrem C-terminal (cercle vermell), la qual només es troba present en el monòmers C i D. (b) Detall de la sobreposició del monòmer B (vermell) amb el monòmer C (taronja), on s'identifiquen els diferents elements d'estructura secundària, així com els extrems N-i C-terminals de l'exteïna N i de la inteïna. En rosa i groc es representen les regions corresponents a les exteïnes N del monòmer B i del monòmer C, respectivament.

L'estructura de cada monòmer NeqN està composta fonamentalment per estructures secundàries regulars de tipus β . Concretament, està format per vuit cadenes β ($\beta 1$ - $\beta 8$), les quals es troben formant una única fulla β antiparal·lela, i una o dues hèlixs α ($\alpha 1$ - $\alpha 2$) per als monòmers A i B o C i D, respectivament (Figura 4.39).

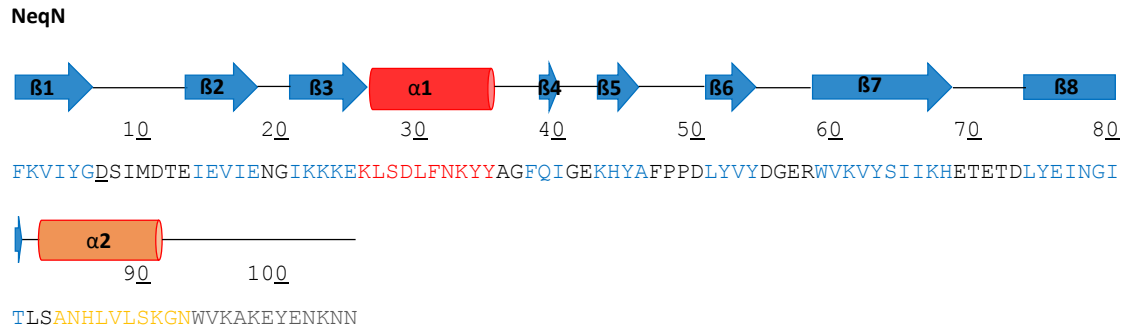


Figura 4.39. Diagrama de l'estructura secundària de la NeqN aïllada. En fletxes blaves es representen les cadenes β i en cilindres vermells les hèlixs α . L'hèlix $\alpha 2$ es troba representada en color taronja per a indicar que només es troba en els monòmers C i D. La localització de cada element d'estructura secundària es senyalitza marcant la seqüència amb el mateix color. En gris s'assenyala la regió de la proteïna que no s'ha pogut modelar, en el cas dels monòmers A i B aquesta regió no modelada també inclouria els residus marcats en taronja.

La superposició de tots els àtoms C α dels dos dímers dona una Desviació Standard (RMSD) de només 0.34 Å. Per tant, es pot considerar que el monòmer A és equivalent al B i el C al D (Figura 4.40).

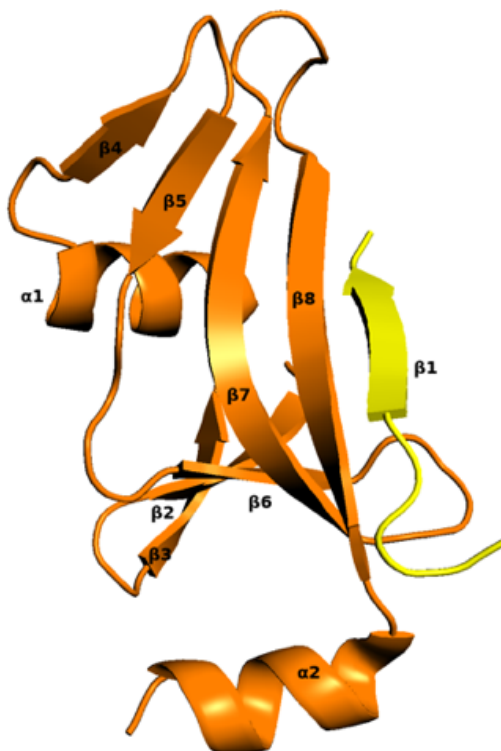


Figura 4.40. Superposició de les estructures del dímer AD i del dímer CB (RMSD=0.34Å). En taronja està representat el dímer AD i en vermell el dímer BC.

Com s'observa en l'estructura tridimensional (Figura 4.38-b; Figura 4.39) els residus de l'exteïna N (els residus -7 a -1) es troben ben definits, formant la primera cadena β . En canvi, els primers sis residus de la inteïna N, inclosa la Ser1, presenten una baixa densitat electrònica, per la qual cosa no es van poder modelar.

L'estructura nativa de la NeqN es troba estabilitzada per nombrosos ponts d'hidrogen intramoleculars i intermoleculars, aquests últims estabilitzen la cadena β_1 de l'exteïna N i la cadena β_8 de la inteïna (Figura 4.41; Annex H) i els monòmers que conformen el dímer (Figura 4.42; Annex H). A més a més, cal destacar els diferents ponts salins que s'estableixen entre les cadenes laterals dels residus: Glu11-Lys16, Glu11-Lys27, Lys-6-Asp-48 i Asp43-Lys84, els quals probablement són essencials per al manteniment de l'estructura nativa de la NeqN. Per últim, però no menys important, cal dedicar un breu espai per parlar de les interaccions π , les quals es donen entre les cadenes laterals dels residus aromàtics (interaccions π - π) o bé entre les cadenes laterals entre residus aromàtics i residus bàsics (interaccions catió- π) que s'ha demostrat que poden jugar un paper clau en la termoestabilitat de proteïnes hipertermòfiles. En l'estructura de la NeqN monomèrica s'han trobat els següents grups d'interaccions: Phe40-Phe31-Tyr38, Tyr45-Lys54; Tyr47-Trp52; i en la interfície del dímer: Phe40-Lys54; Phe-7 i Phe-7. Convé destacar que aquestes dues interaccions, Tyr45-Lys54; Tyr47-Trp52, són exclusives dels monòmers C i D. Així com s'han descrit ponts salins en altres inteïnes partides, no s'ha observat en cap de les estructures resoltes fins ara interaccions π . El fet que la inteïna NeqN sigui l'única inteïna partida, resolta fins ara, per la qual s'han identificat aquest tipus d'interaccions i tenint en com té el paper termoestabilitzador que aquestes poden tenir [105],[122]–[124], es pot suggerir que les interaccions entre les cadenes laterals dels residus aromàtics probablement desenvolupen un paper clau en la termoestabilitat d'aquesta proteïna.

a



b

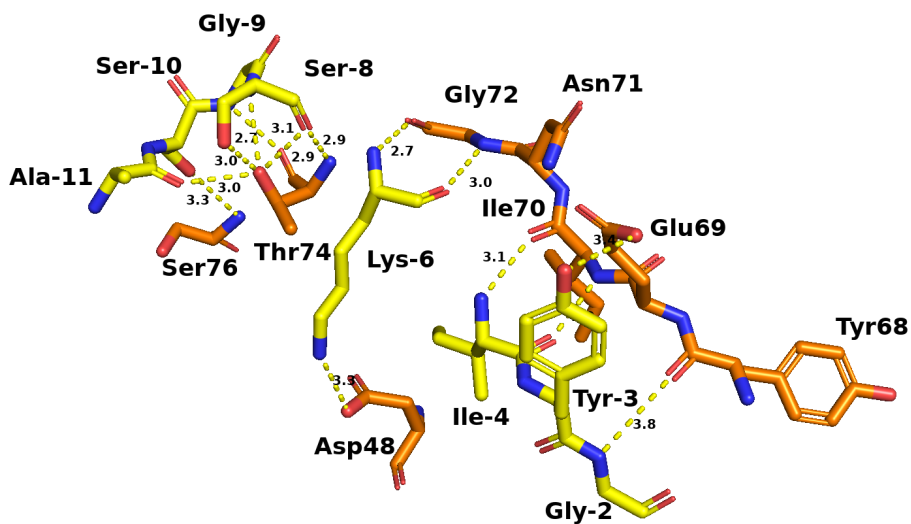


Figura 4.41. Interaccions entre l'exteïna N i la NeqN. (a) Estructura del monòmer C de la NeqN amb l'exteïna N; (b) Detall de les interaccions per pont d'hidrogen entre l'exteïna i la NeqN. L'exteïna N i la NeqN estan representades en groc i taronja, respectivament. Els ponts d'hidrogen es representen amb línies discontinües grogues. Les distàncies de les interaccions en Å.

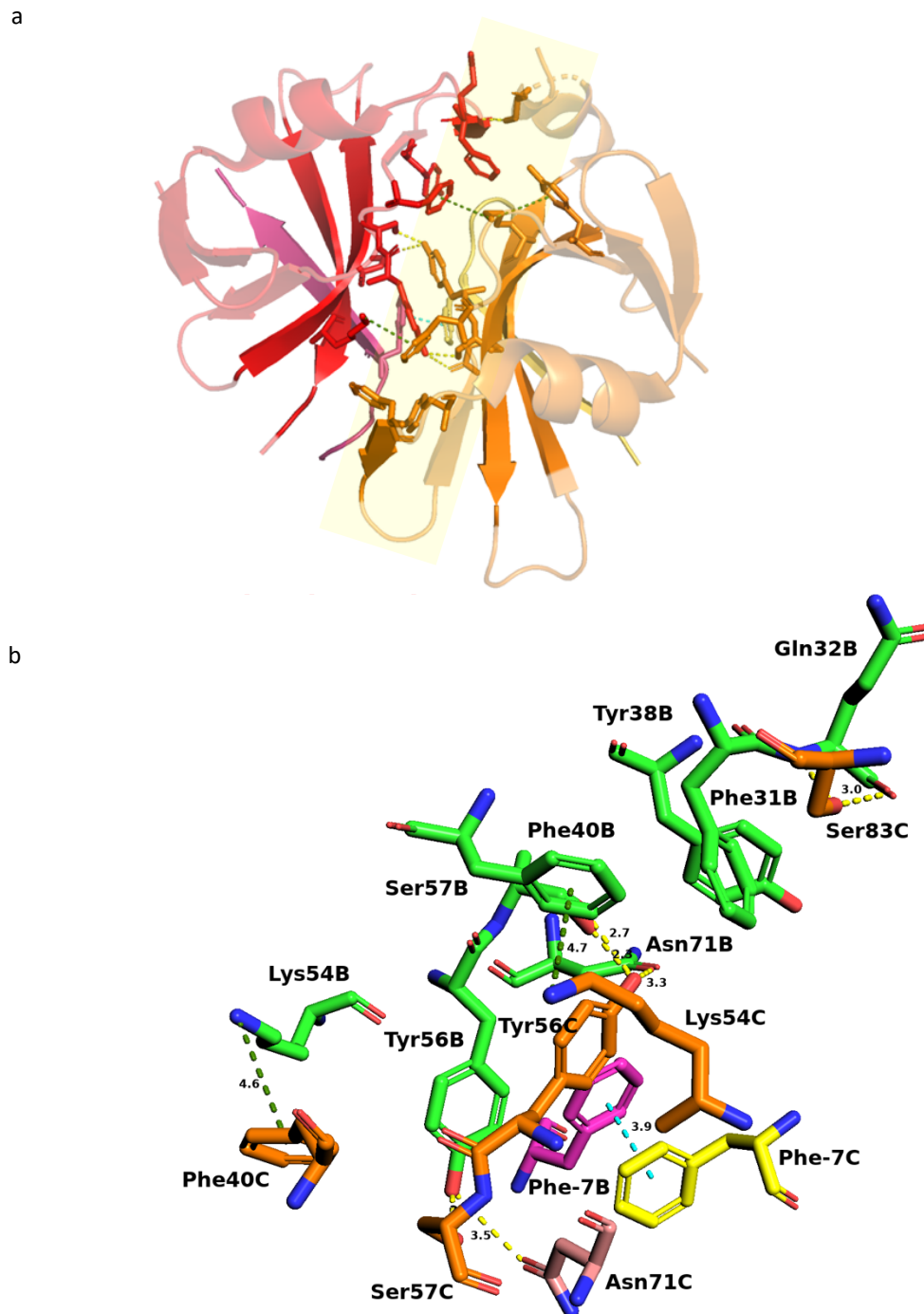


Figura 4.42. Interaccions presents en la interfície dimèrica. (a) Estructura tridimensional del dímer format pels monòmers B i C de la NeqN. La NeqN i l'exteïna N del monòmer B estan representats en vermell i rosa, respectivament. La NeqN i l'exteïna N del monòmer C estan representats en taronja i groc, respectivament. En un fons groc s'indica la interfície d'interacció entre els monòmers (b) Detall de les interaccions establertes en la interfície del dímer. Els residus de l'exteïna N i de la NeqN del monòmer B s'indiquen en rosa i verd, respectivament, i en taronja i groc estan indicats els residus de la NeqN i de l'exteïna N del monòmer C. Els residus estan identificats pel nom abreujat, posició i monòmer, del qual formen part. Els ponts d'hidrogen es representen amb línies discontinües grogues i les interaccions π amb línies verdes discontinües. Les distàncies mesurades entre àtoms es troben en Å.

La Npu, és la inteïna partida millor caracteritzada actualment [33], [64], [73], [75], [175], [176]. Com bé s'ha explicat en la introducció d'aquesta tesi, en la NpuN associada amb la NpuC es distingeixen dos lòbuls, NpuN1 i NpuN2, els quals presenten diferents propietats electroestàtiques i d'hidrofobicitat. La comparació estructural entre la NpuN i la NeqN revela que la NeqN presenta certa equivalència amb NpuN1, però cap amb NpuN2 (Figura 4.43). Aquest resultat indica que la NeqN aïllada, a diferència de la NpuN, no està dividida amb dos lòbuls, fet que explica el caràcter electroestàtic i hidrofòbic d'aquesta molècula (Figura 4.44). Aquestes evidències suggereixen que la NeqN associaria amb la NeqC seguint un mecanisme diferent del proposat per la inteïna NpuN.

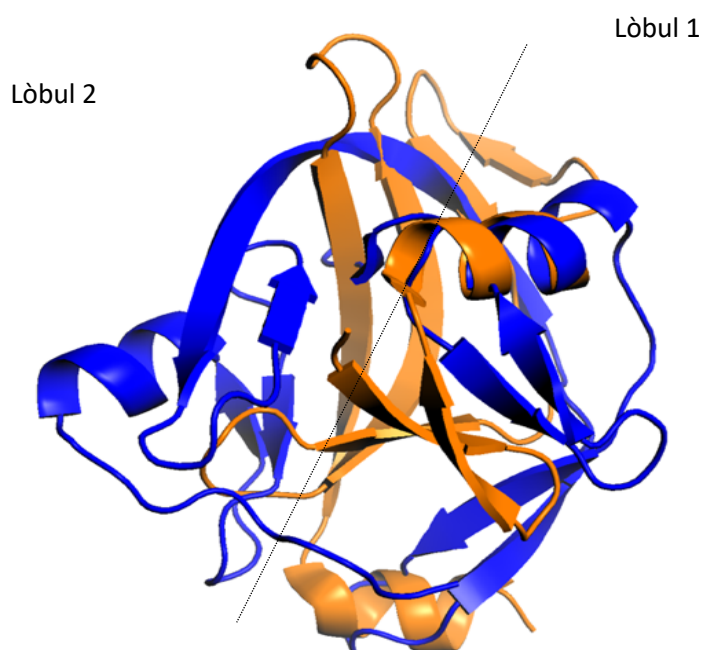


Figura 4.43. Comparació de l'estructura de la NeqN aïllada (5oxw) i la NpuN formant el complex NpuN DnaE (4lx3). La NeqN es representa en taronja i la NpuN en blau.

SIMDTEIEVIENGIKKKEKLSDLFNKYYAGFQIGEKHYAFPPDLVYVDGE

Carrega neta: -4

Hidrofobicitat: 0.43

RWVKVYSIIKHETETDLYEINGITLSANHLVLSKGNWVKAKEYENKNN

Carrega neta: +1

Hidrofobicitat: 0.44

Figura 4.44. Representació del caràcter electroestàtic i hidrofòbic de les dues meitats de NeqN assumint el model de dos lòbuls proposats per la inteïna Npu. En vermell s'indiquen els residus àcids i en blau els bàsics. La hidrofobicitat i la càrrega neta van ser calculades seguint el protocol descrit a Material i Mètodes.

Efecte de l'exteïna en el procés de dimerització de la NeqN

Es va realitzar un estudi preliminar per determinar el paper de l'exteïna N en el procés de dimerització. Per dur-ho a terme es va expressar i purificar la construcció H6TEVT7-NeqN seguint el mateix protocol descrit per a la construcció NeqN. La T7-NeqN pura i la NeqN un cop purificades es van preparar a una concentració final de 13 μM amb 50 mM Tris-HCl a pH 8.0, 100 mM NaCl i 1 mM EDTA. Paral·lelament, es van preparar tres solucions de T7-NeqN a diferents concentracions, 200 μM , 40 μM i 13 μM en el mateix tampó. Totes les mostres es van analitzar per SEC.

Com es pot observar en la Figura 4.45 la formació d'espècies dimèriques per part dels dos precursors suggereix que és un procés que depèn tant de l'exteïna (Figura 4.45-a) com de la pròpia inteïna (Figura 4.45-b). En la Figura 4.45-a s'observa com la substitució de la seqüència nativa de l'exteïna N disminueix considerablement la tendència a dimeritzar del precursor. Aquest resultat concorda amb l'estructura cristal·logràfica, on s'observa que el dímer es troba estabilitzat per la interacció entre Phe(-7) (Figura 4.42).

Aquesta tendència de la inteïna N a dimeritzar potser un dels mecanismes adoptats per la proteïna per augmentar la seva termoestabilitat. Com s'ha pogut veure en l'estructura tridimensional la interfície del dímer està formada principalment per interaccions intermoleculares entre les cadenes $\beta 7$ i $\beta 8$, que corresponent als residus 51-76. Aquesta regió presenta un marcat caràcter hidrofòbic, 0.47 en l'escala de Kyte-Doolittle, i és la responsable, com es veurà més endavant, d'interaccionar amb la NeqC. Com a conseqüència, en absència de la inteïna C i per reduir l'àrea de superfície hidrofòbica exposada a solvent aquesta dimeritzaria.

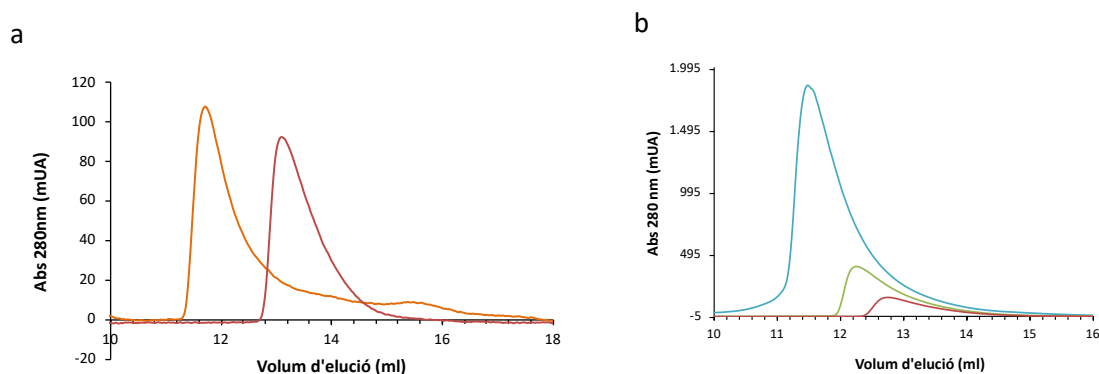


Figura 4.45. Anàlisi de dimerització de la NeqN per SEC. (a) Importància de l'exteïna N en el procés de dimerització de la NeqN. Perfil taronja: NeqN; Perfil vermell: T7-NeqN. Les dues construccions es van preparar a una concentració final de 13 μM . (b) Estudi de la variació del volum d'elució de la T7-NeqN, en funció de la concentració. Influència de la concentració de proteïna. Perfil blau: 200 μM ; Perfil verd: 40 μM ; Perfil vermell: 13 μM . Totes les preparacions es van realitzar amb 50 mM Tris-HCl, 150 mM NaCl a pH 7.4.

4.3. Caracterització del procés d'associació entre els dos fragments de la Inteïna partida Neq pol i del complex NeqN/NeqC

Simultàniament a la caracterització estructural de la NeqN, es va estudiar des d'un punt de vista cinètic i termodinàmic el procés d'associació entre els dos fragments de la inteïna partida Neq pol, NeqN i NeqC, i d'altra banda es va caracteritzar el complex NeqN/NeqC.

Expressió i purificació de la NeqN i NeqC

Per a l'expressió i purificació de la inteïna N es va utilitzar la construcció NeqN, i per la de la inteïna C la construcció NeqC. La construcció NeqN clonada en el vector pET-22b(+) va ser

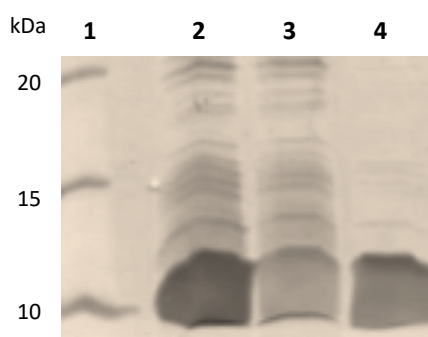


Figura 4.46. Anàlisi per SDS-PAGE de la solubilitat del precursor C, NeqC. Les cèl·lules obtingudes de la producció a 37°C de la NeqC-(TDS)H6 es resuspenen amb 20 mM Fosfat, 0.5 mM NaCl i 20 mM imidazol a pH 8.0, i es lisen. La fracció soluble es separa de la insoluble per centrifugació. Carril 1: Marcadors de massa molecular; Carril 2: Fracció total; Carril 3: Fracció soluble; Carril 4: Fracció insoluble després de la lisi cel·lular.

produïda i purificada seguint el mateix protocol que per a la seva caracterització individual. Per a obtenir el precursor C, NeqC-(TDS)-H6, la NeqC va ser clonada al vector pET-29a(+). Com era d'esperar, el precursor C a 37°C es trobava majoritàriament en la fracció insoluble formant cossos d'inclusió (Figura 4.46). Com s'ha comentat en l'apartat anterior, probablement la NeqC sigui una proteïna desestructurada, fet que fa que sigui susceptible a la degradació i a l'atac per part de proteases cel·lulars. La purificació a partir de cossos d'inclusió redueix considerablement tant la degradació com la proteòlisi [177]. Per aquesta raó, es va decidir purificar el precursor C, NeqC, a partir de cossos d'inclusió. Els quals es van solubilitzar amb una solució 8 M d'urea, i la proteïna recombinant es va purificar utilitzant cromatografia d'afinitat Ni-NTA sota condicions desnaturalitzants. Per a eliminar l'agent desnaturalitzant de manera gradual i mantenir la NeqC soluble, les fraccions pures obtingudes de la cromatografia es van barrejar i es van dialitzar en diferents passos fins a obtenir la proteïna en aigua Milli-Q.

4.3.1. Caracterització fisicoquímica del procés d'associació entre els dos fragments de la inteïna partida Neq

Caracterització cinètica

Per estudiar exhaustivament l'efecte de la temperatura sobre el procés d'associació entre les dues inteïnes es va dissenyar un assaig analític on es va mesurar la disminució de la fluorescència intrínseca emesa pel fragment NeqN (Trp52 i Trp87) a mesura que aquesta s'unia amb el fragment NeqC. Com era d'esperar tractant-se d'una proteïna hipertermofílica, la velocitat d'associació entre les dues inteïnes augmenta gradualment amb la temperatura (Taula 4.6), a excepció de la velocitat calculada a 74°C. En l'espectre de fluorescència obtingut a 74°C s'observa una reducció de la fluorescència de la NeqN en l'equilibri de tan sols un 8%, mentre que en les altres temperatures s'observa una reducció del 30-40% (Annex I). Aquesta reducció anòmala suggereix que la velocitat de la reacció d'associació entre els dos fragments a aquesta temperatura és molt elevada i no es pot determinar per aquesta tècnica. D'altra banda, els resultats mostren que la velocitat d'associació dels dos fragments de la inteïna partida Neq pol a 64°C es produeix en fraccions de segons (Taula 4.6), coherents amb resultats obtinguts prèviament per la inteïna partida Ssp DnaE [76].

Taula 4.6. Constants de velocitat de la reacció d'associació entre la NeqN i la NeqC a diferents temperatures.

Temperatura (°C)	K_{obs} (s ⁻¹) ^a
17	$2.3 \times 10^{-3} \pm 0.0001$
25	$5.5 \times 10^{-3} \pm 0.0003$
36	$1.8 \times 10^{-2} \pm 0.0011$
46	$5.5 \times 10^{-2} \pm 0.0022$
55	$0,1647 \pm 0.0107$
64	$0,3655 \pm 0.0218$
74	$7.6 \times 10^{-2} \pm 0.0171$

^a k_{obs} es va calcular ajustant les corbes d'unió a una equació de primer ordre.

A partir de les velocitats obtingudes a les diferents temperatures es va determinar l'energia d'activació necessària per formar el complex NeqN/NeqC (Taula 4.7). L'elevada energia d'activació obtinguda, 89.5 KJ/mol, és característica de proteïnes molt estables, les quals necessiten trencar enllaços per associar. L'elevada energia d'activació del procés probablement es pot explicar tenint en compte que la NeqN lliure dimeritza, ja que perquè aquesta pugui interaccionar amb la seva companya la NeqC s'hauran d'hidrolitzar interaccions que estableixen el dímer.

Taula 4.7. Energia d'activació de la reacció d'associació entre la NeqN i la NeqC.

Energia d'activació (Ea)	A
89.545 KJ/mol	30

Caracterització termodinàmica

Per mesurar la constant d'afinitat, Kd, del procés d'associació igual que per estudiar la cinètica del procés, es va seguir la disminució de la fluorescència intrínseca de la NeqN en presència de concentracions creixent de NeqC. Les construccions emprades per aquest estudi van ser NeqN i NeqC. Com era d'esperar, els dos fragments de la inteïna Neq pol associen amb una elevada afinitat, presentant valors de Kd de l'ordre nM, 18.6 nM a 25°C (TAULA 4.8), consistents amb altres resultats obtinguts per la inteïna NpuDnaE i SspDnaE, prèviament [3],[4]. Posteriorment, per obtenir els paràmetres termodinàmics del procés d'associació, es va calcular cada Kd per a cada temperatura en un rang de temperatures entre els 20°C i 60°C. Coneixent que la Kd és depèn de temperatura (Annex J), es va obtenir un valor d'entalpia de van't Hoff de 2.36 ± 1.5 kcal/mol i un valor d'entropia de $14.33 \text{ cal mol}^{-1}$. Amb aquests paràmetres i seguint la relació $\Delta G(T) = -RT \ln Kd(T)$, es va poder estimar la variació d'energia lliure (ΔG) del procés d'interacció entre els dos fragments de la inteïna, $\Delta G(25^\circ\text{C}) = -1.73 \text{ kcal/mol}$.

Taula 4.8. Constant d'afinitat i paràmetres termodinàmics del procés d'associació entre la els dos fragments de la inteïna, NeqN i NeqC.

Kd (nM)	ΔH° (kcal/mol)	$T\Delta S^\circ$ (kcal/mol) (25°C)	$\Delta G^\circ(25^\circ\text{C})$ (kcal/mol)
18.55	+2.36	+4.27	-1.73

S'ha demostrat que la interacció entre la SspN i la SspC és un procés espontani, el qual està dirigit per l'entalpia ($\Delta H = -29.1 \text{ kcal/mol}$) i que presenta un elevat cost entròpic desfavorable ($T\Delta S = -18.9 \text{ kcal/mol}$) [77]. Aquest fenomen és coneix com a compensació entàlpica-entropica [178] i freqüentment, es troba en aquelles reaccions on l'associació comporta un guany d'estructura, com és el cas d'aquesta inteïna. La variació favorable d'entalpia és conseqüència de la formació de noves interaccions entre els dos fragments. I la variació desfavorable d'entropia probablement és conseqüència de la pèrdua de llibertat conformacional que comporta el procés d'associació [179].

Igual que per la inteïna partida Ssp DnaE, el procés d'associació entre la NeqN i la NeqC és un procés espontani ($\Delta G < 0$). Al contrari, aquest experiment demostra que la interacció entre els dos fragments és desfavorable des d'un punt de vista entàlpic ($\Delta H > 0$) i favorable des d'un punt

de vista entròpic ($\Delta S > 0$). La qual cosa indica que el procés d'associació entre NeqN i NeqC segueix un mecanisme diferent del proposat per Liu i col·laboradors per a la inteïna mesofílica [77].

Per a la interpretació d'aquests resultats s'ha tingut en compte que la variació d'entalpia mesurada amb l'experiment, és el resultat del sumatori de l'entalpia conformacional, de l'entalpia de la interacció i de l'entalpia de solvatació [180].

$$\Delta H = \Delta H_{\text{Conformació}} + \Delta H_{\text{Interacció}} + \Delta H_{\text{Solvatació}}$$

L'entalpia conformacional està associada a canvis en l'estructura secundària, en el cas que l'associació comporti un guany d'estructura s'espera que aquesta variació sigui exotèrmica ($\Delta H < 0$), i en el cas que suposi una pèrdua d'estructura s'espera que aquesta sigui endotèrmica ($\Delta H > 0$). D'altra banda, l'entalpia d'interacció es pot considerar exotèrmica, ja que s'assumeix que la interacció entre dues espècies comporta intrínsecament la formació de noves interaccions no covalents. Per últim, i en el cas de l'entalpia de solvatació aquesta pot adoptar un fort caràcter endotèrmic, ja que el procés d'associació pot provocar l'alliberació de molècules d'aigua altament ordenades de la superfície que forma la interfície d'associació (hidrofòbica). Aquesta reacció on s'alliberen molècules d'aigua és entàlpicament desfavorable, ja que es trenquen interaccions iòniques i ponts d'hidrogen [181].

Els resultats obtinguts d'entalpia juntament amb els resultats obtinguts per CD i per cristal·lografia, permeten postular el següent mecanisme d'associació entre els dos fragments: Abans de formar-se el complex NeqN/NeqC, la NeqN es trobaria en un estat dimèric i la NeqC desplegada. En un primer contacte entre els dos fragments, la NeqC desestabilitzaria la forma dimèrica de la NeqN. Aquest primer contacte estaria governat principalment per interaccions de caràcter electroestàtic. Seguidament es formaria el complex NeqN/NeqC actiu, en el qual la NeqC adoptaria la seva estructura nativa.

Tenint en compte el mecanisme postulat es podria dir que:

- L'entalpia conformacional és exotèrmica, ja que la NeqC adquireix estructura secundària secundària.
- L'entalpia d'interacció no és significativa en aquest cas, ja que existeix un equilibri entre les interaccions que es trenquen al desestabilitzar el dímer i les que es formen en l'associació NeqN-NeqC.
- Per tant, ja que l'entalpia mesurada és endotèrmica, l'entalpia conformacional ha de ser compensada per valors alts positius d'entalpia de solvatació.

D'altra banda l'entropia positiva associada a l'associació de la NeqN amb la NeqC segueix el clàssic efecte hidrofòbic [182]. La NeqC presenta un fort caràcter hidrofòbic, $H=0,48$, com a

conseqüència l'associació d'aquesta amb la NeqN probablement comporta l'alliberament d'un elevat nombre de molècules d'aigua ordenades, el que explicaria els valors positius d'entropia. D'aquesta manera es pot concloure que tant els valors favorables d'entropia com els desfavorables de l'entalpia es poden explicar per l'alliberament de molècules d'aigua ordenades que provocaria l'associació de les dues inteïnes.

4.3.2. Caracterització fisicoquímica del complex NeqN/NeqC

A partir dels fragments N- i C- purs es va formar el complex in vitro seguint el protocol descrit en Material i Mètodes.

SEC i ultracentrifugació analítica

Per verificar la identitat del complex NeqN/NeqC, amb 20 mM Tris-HCl, 150 mM NaCl, 1mM EDTA a pH 8.0, aquest es va injectar a una concentració final de 356 μ M en una columna de cromatografia d'exclusió molecular Sephadex-75 HR-10/3 prèviament calibrada amb diferents proteïnes globulars de massa molecular coneguda. El complex NeqN/NeqC s'elueix com un únic pic amb un volum d'elució (V_e) aproximat de 11.3 ml (Figura 4.47-a), el qual correspon a una massa molecular estimada de 19393 Da (Annex F). La massa estimada per SEC del complex NeqN/NeqC s'ajusta a la massa molecular predita pel complex format per una molècula de cada precursor (19001.3 Da). Paral·lelament, la massa molecular del complex NeqN/NeqC es va confirmar per ultracentrifugació analítica (Figura 4.47-b). El perfil de sedimentació obtingut per ultracentrifugació analítica a una concentració de 245 μ M (4.62 mg/ml) mostra un pic majoritari amb un coeficient de sedimentació corregit a aigua i a 20°C ($s_{20,w}$) de 2.30 S i una altra espècie minoritària que sedimenta a 4.41 S que podria tractar-se d'un contaminant. Les masses moleculars estimades són de 21600 Da per a l'espècie principal i de 57600 Da per la segona. La massa molecular estimada per l'espècie principal s'aproxima a la massa molecular estimada pel complex NeqN/NeqC format per una molècula de NeqN i una molècula de NeqC. La petita variació en la massa molecular podria ser deguda al fet que aquesta tècnica assumeix que la proteïna adopta una estructura globular. En el cas del complex NeqN/NeqC, aquest probablement adopta un plegament en forma de ferradura característic de les inteïnes, el qual no és tan compacte com una proteïna globular. Com a conseqüència el radi de la proteïna és més gran, la qual cosa pot donar lloc a una sobreestimació de la massa real.

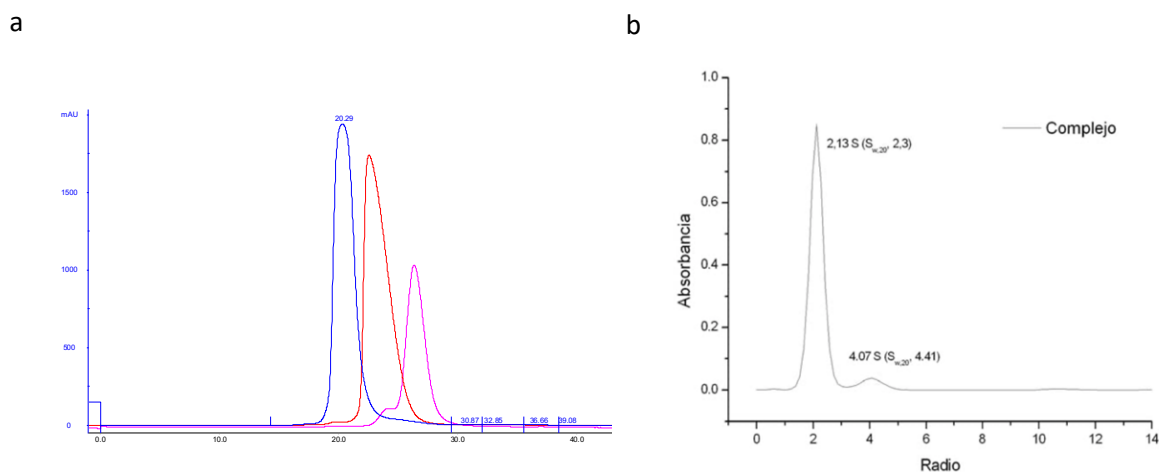


Figura 4.47. Determinació de la massa molecular del complex *NeqN/NeqC*. (a) Anàlisi per SEC de la *NeqN* (línia blava), de la *NeqC* (línia rosa) i del complex *NeqN/NeqC* (línia vermella). (b) Dades de l'equilibri de sedimentació amb una corba ajustada per la solució de complex *NeqN/NeqC* a una concentració 245 μ M a 45000 rpm i a 20°C. La massa molecular estimada de la inteiña N és de 27700 Da per a l'espècie majoritària.

Caracterització per dicroisme circular (CD)

A continuació, es mostren els espectres obtinguts per dicroisme circular de la *NeqN*, *NeqC* i del complex *NeqN/NeqC* en la regió de l'UV-llunyà (190-250 nm) i la regió UV-proper. Igual que per la *NeqN* aïllada, pels espectres en la regió de l'UV-llunyà no es van poder mesurar a longituds d'ona per sota de 200 nm, ja que a les concentracions de proteïna emprades l'instrument es saturava. Amb els espectres obtinguts per dicroisme circular en la regió del llunyà dels dos fragments aïllats i del complex (Figura 4.48-a), i utilitzant l'algoritme K2D [169], [170] es va poder estimar el contingut d'estructura secundària de cadascuna de les espècies. Per la *NeqN*, com ja s'ha comentat anteriorment, es va estimar un 21% d'hèlix α , un 33% en beta β i un 46% en random coil. Per la *NeqC* un 42% de full β , un 8% d'hèlix α , i un 50% de random coil. I per últim, pel complex *NeqN/NeqC* es preveu un 24% d'hèlix α , un 40% de full β i un 36% en random coil. Aquests resultats suggereixen que l'associació dels dos fragments de la inteiña no va associada a grans canvis estructurals, com passa amb altres membres de la mateixa família [57], [77]. En canvi, si que sembla que l'associació dels dos fragments provoca un canvi conformacional en la *NeqN*. Els espectres obtinguts del complex tant en la regió de l'UV-llunyà com en l'UV-proper mostren que la banda positiva a 295 nm característica de l'espectre de la *NeqN* desapareix (Figura 4.48-b). La pèrdua d'aquesta banda indica que algun o ambdós triptòfans de la *NeqN* passen d'estar en un entorn quiral ben estructurat en la forma aïllada a un entorn més flexible quan la *NeqN* es troba acomplexada amb la *NeqC*. Així doncs es pot dir que els resultats

obtinguts per dicroisme circular indiquen que l'associació dels dos fragments no provoca grans canvis estructurals però sí conformacionals, ja que la NeqN perd rigidesa en algunes regions de la seva estructura nativa.

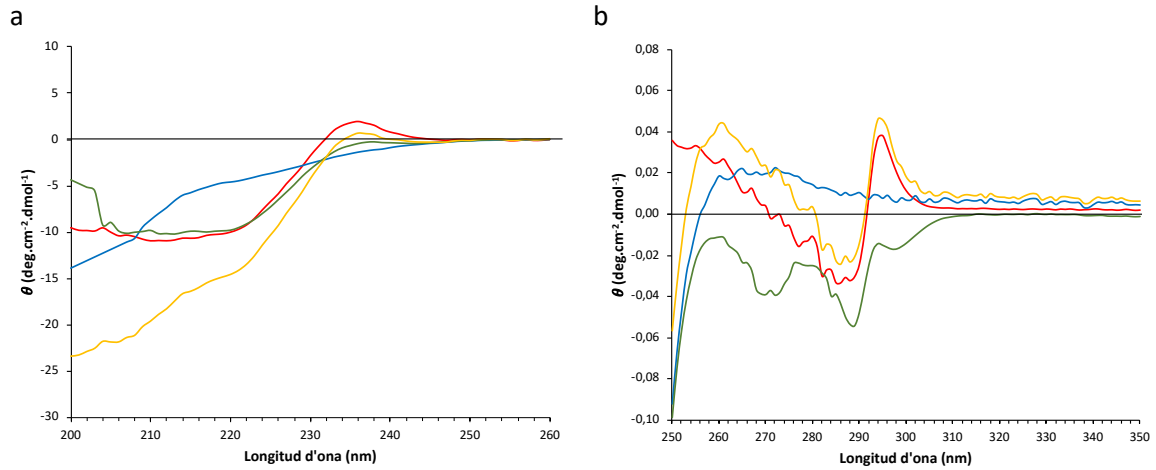


Figura 4.48. Espectres de dicroisme circular de la NeqN, de la NeqC i el complex NeqN/NeqC. (a) Espectre CD en la regió del llunyà (200-250 nm); (b) Espectre CD en la regió del proper (350-250 nm). Perfil vermell: NeqN; Perfil blau: NeqC; Perfil verd: complex NeqN/NeqC; Perfil groc: Sumatori dels espectres NeqN i NeqC. Totes les mostres s'han preparat amb 20 mM tampó fosfat, 150 mM NaCl, 1 mM EDTA a pH 8.0.

Determinació de l'estabilitat tèrmica

La T_m del complex NeqN/NeqC obtinguda a partir de l'anàlisi per DSC és de 105 °C igual que per a la NeqN aïllada. Aquest fet dona suport als resultats obtinguts per CD, on no s'observen grans canvis estructurals després de l'associació dels fragments de la proteïna Neq pol (Figura 4.49).

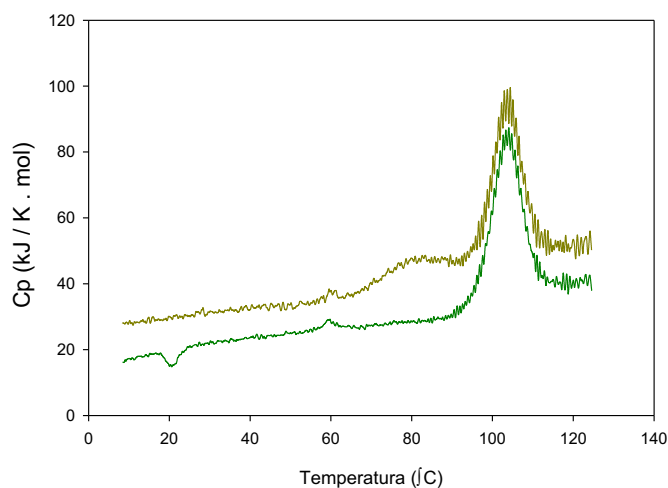


Figura 4.49. Termograma obtingut per DSC del procés de desnaturalització de la NeqN i del complex NeqN/NeqC. Experiments realitzats a una concentració de 26 μ M de proteïna amb 20 mM tampó fosfat, 150 mM NaCl, 1 mM EDTA a pH 8.0 a una velocitat d'escombrat d e90K/min. Línia verda: NeqN; Línia caqui: complex NeqN/NeqC

4.3.3. Resolució de l'estructura tridimensional per cristal·lografia

El complex NeqN/NeqC purificat per cromatografia d'exclusió molecular va cristal·litzar en dues formes de cristalls diferents, tetragonal i hexagonal (50xx i 50xz). Els cristalls hexagonals es van resoldre des del començament amb Arcimboldo [183] amb una resolució d'1.2 Å. Posteriorment, es van resoldre els cristalls tetragonals per reemplaçament molecular utilitzant com a model l'estructura del complex determinada en els cristalls hexagonals (Taula 4.9). Les unitats asimètriques de les dues formes de cristalls contenen només una molècula de complex, el qual està format per una cadena polipeptídica corresponent a la subunitat NeqN i una altra cadena, més petita, corresponent a la subunitat NeqC (Figura 4.50).

Taula 4.9. Col·lecció de dades i estadístics del refinament.

Col·lecció de dades ^a	Complex NeqN/NeqC	Complex NeqN/NeqC
Grup espacial	P42 ₁ 2	P6 ₅ 22
Dimensions de la cel·la		
a,b,c (Å)	67.43, 67.43, 50.89	45.67, 45.67, 254.41
A, B, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 120.00
Reflexions úniques	12, 621 (1803)	51, 160 (7, 240)
Resolució (Å)	50.89-1.74 (1.83-1.74)	42.40-1.20 (1.26-1.20)
Longitud d'ona	0.9792	0.9792
R _{sym} (%) ^b	0.061 (1.02)	0.094 (0.68)
<I/σI>	22.8 (2.7)	16.2 (1.0)
CC _{1/2} ^c	0.98 (0.61)	0.99 (0.65)
Completesa	100.0 (100.0)	99.90 (99.90)
Redundància	12.5 (12.9)	12.2 (12.3)
Estadístics de refinament		
Resolució	50.89-1.74	42.40-1.20
No. de reflexions	12, 000	48, 536
R _{cris} (%) ^d	20.44	13.73
R _{liure} (%) ^e	21.71	15.63
No. de residus	108	123
Molècules de solvent	201	49
Contingut de solvent (%)	41	55
Àtoms no H	1, 319	927
Mitjana factor B (Å ²)	36.40	16.09
Coordinada d'error (Å) ^f	0.084	0.018
RMSD bond (Å)	0.02	0.02
RMSD angles (°)	1.95	1.97

^a Els valors entre parèntesis corresponen a la Shell amb més resolució

^bR_{sym}= $\frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, on I_i(hkl) és la intensitat d'una observació i <I(hkl)> és el valor mitjà d'observacions per a una única reflexió.

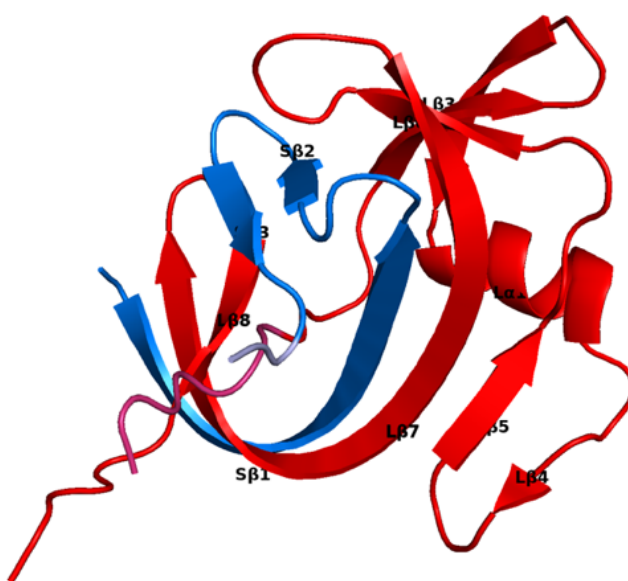
^cCC_{1/2} és el percentatge de la correlació entre intensitats de bases de dades aleatòries.

^dR_{cris}= $\frac{\sum_h |F_o(h) - F_c(h)|}{\sum_h |F_o(h)|}$, on F_o i F_c són les amplituds del factor estructura observades i calculades, respectivament.

^eR_{liure} va ser calculat amb el 5% de les dades, les quals van ser excloses del refinament.

^fes basa en la màxima probabilitat

Figura 4.50. Estructura cristal·logràfica obtinguda del complex NeqN/NeqC (5oxz). NeqN es troba representada en vermell i la NeqC en blau. Els 7 residus de l'exteina N salvatge es troben representats en rosa i els 3 residus corresponents a l'exteina C en gris.



La comparació de les estructures del complex NeqN/NeqC obtingudes per les dues formes dels cristalls donen un RMSD de tant sols 0.44 Å. Tot i això, en la sobreposició de les dues estructures (Figura 4.51), s'observa que existeixen diferències importants en les regions corresponents a les exteïnes N i C i en la cadena $\beta 4$ de la inteïna, la qual és més curta en l'estructura obtinguda en els cristalls hexagonals. D'altra banda, en el cristall hexagonal s'ha pogut resoldre l'estructura dels residus 1-83 de la inteïna N, 1-30 de la inteïna C, residus (-7)-(-1) de l'exteïna N i tres residus, TDS, de l'exteïna C. En canvi, en l'estructura obtinguda a partir dels cristalls tetragonals només es veuen dels residus 1-77 de la inteïna N, 1-30 de la inteïna C, (-2)-(-1) residus de l'exteïna N. En aquest cas, no es veu cap dels residus corresponents a l'exteïna C. En aquest treball s'ha utilitzat de manera general l'estructura obtinguda a partir del cristall hexagonal.

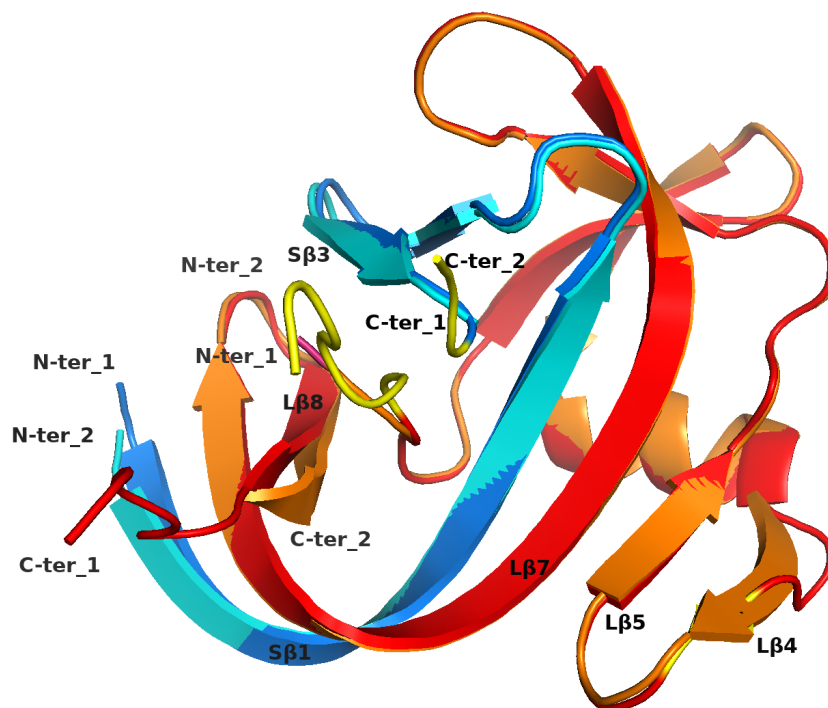


Figura 4.51. Sobreposició de les dues estructures del complex obtingudes dels cristalls hexagonals (5oxz) i dels cristalls tetragonals (5oxx). NeqN i NeqC del complex obtingut a partir dels cristalls hexagonals (5oxz) estan representades en vermell i blau, respectivament. NeqN i NeqC del complex obtingut dels cristalls tetragonals (5oxx) estan representades en taronja i turquesa, respectivament. En groc es representen les principals diferències entre les dues estructures. Els extrems N- i C-terminals del complex hexagonal (5oxz) es troben indicats amb _1, i els del complex tetragonal (5oxx) amb un _2. La lletra L davant dels elements d'estructura secundària indiquen que corresponen a la NeqN, i la lletra S a la NeqC.

Estructura secundària

L'estructura del complex NeqN/NeqC està formada principalment, per estructures secundàries regulars de tipus β , les quals s'ordenen adoptant un plegament HINT [184], característic de les inteïnes. L'anàlisi DSSP del complex [145] defineix tres cadenes β (S β 1-S β 3) per a la subunitat NeqC i set cadenes β (L β 2-L β 8) i una hèlix α (L α 1) per a la subunitat NeqN (Figura 4.52). Sempre que ha estat possible els elements d'estructura secundària s'han enumerat com en l'estructura de la NeqN aïllada. Els residus de l'exteïna N, Asp(-1) – Phe(-7), així com els residus de l'extrem N-terminal de la NeqN, Ser1-Thr5, es troben desordenats. El mateix passa amb els tres residus de l'exteïna C, Thr+1-Ser+3, i els dos residus de l'extrem C-terminal de la NeqC, His29 i Asn30. És a dir, la inteïna presenta un plegament globular estable, mentre que els residus que es troben en els llocs d'unió exteïna-inteïna es troben desordenats. Aquest tipus de plegaments s'han observat en altres inteïnes.

Seguint un ordre seqüencial la inteïna N plega NeqN (Figura 4.52). Començant amb una cadena β 2, formada pels residus 6-11, la qual està unida a través d'un gir β (residus 14-19) a la cadena β 3. Com a resultat s'obté la primera forquilla β antiparal·lela de la NeqN en el complex. A continuació, es troba una hèlix α amfipàtica composta pels residus 20-28, que es troba per un canto exposada a solvent i per l'altre empaquetada entre les cadenes laterals de les cadenes β 2 i β 3. La cadena β 4, la qual és molt curta en l'estructura 5oxz (residus 32-33), es troba unida per l'extrem N-terminal a l'hèlix α 1 a través d'una regió desplegada (residus 29-31) i per l'extrem C-terminal mitjançant un gir β a la cadena β 5, residus 36-39. D'aquesta manera la β 4 i β 5 formen la segona forquilla β antiparal·lela. Aquesta forquilla β només s'ha trobat en estructures tridimensionals obtingudes d'inteïnes hipertermofíliques, no s'observa en les seves homòlogues mesofíliques (Figura 4.59; Figura 4.60). Per aquesta raó s'ha proposat que aquesta estructura secundària, anomenada HTH, augmenta l'estabilitat d'aquestes inteïnes, permetent-les que catalitzin la reacció de *splicing* a temperatures molt elevades [19]. Seguidament, es troba la cadena β 6 formada pels residus 44-47, que es troba orientada aproximadament formant un angle recte respecte la cadena β 5. Aquesta cadena β 6 es troba unida a través d'un gir β a la cadena 7, l'estructura secundària més llarga de tota la inteïna (residus 52-77), donant forma al que seria una tercera forquilla β antiparal·lela. Aquesta mateixa cadena β 7, pel seu extrem C-terminal, es troba formant una forquilla β antiparal·lela amb l'última cadena β de la inteïna, la β 8 formada pels residus 73-77.

Pel que fa a la NeqC i també seguint un ordre seqüencial NeqN (Figura 4.52), aquesta comença plegant-se, sempre que es trobi acomplexada amb la NeqN, en una cadena β , β 1, composta pels residus 2-16 de l'extrem N-terminal. A continuació, a través d'un petit fragment desordenat

es troba la cadena β_2 , residus 20-23, la qual forma una forquilla β antiparal·lela amb la cadena β_3 , en l'extrem C-terminal de la inteïna C.

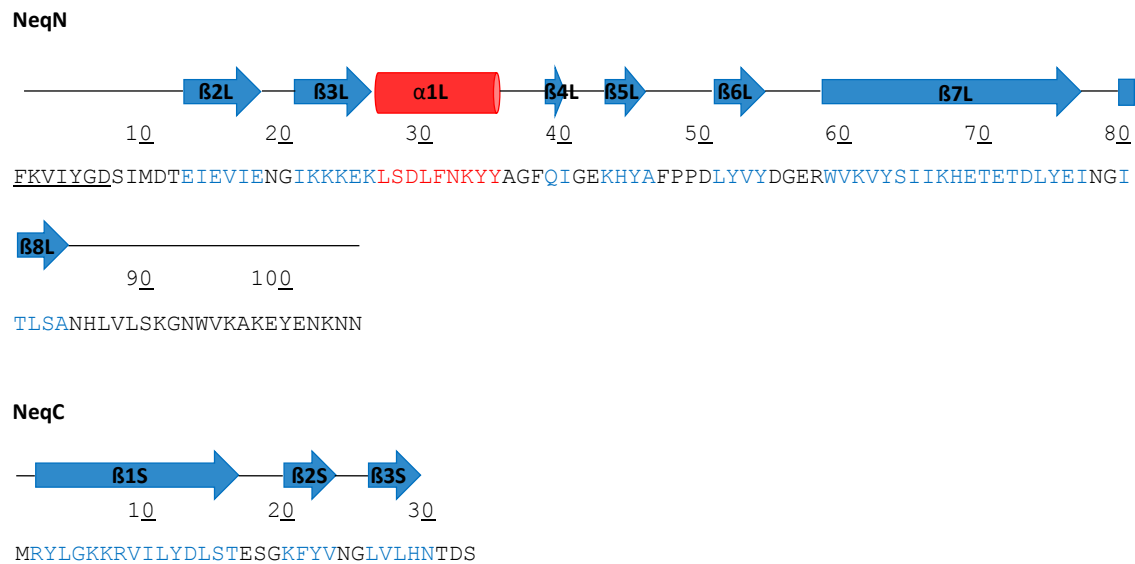


Figura 4.52. Diagrama de l'estructura secundària de la NeqN i de la NeqC en el complex NeqN/NeqC. En fletxes blaves es representen les cadenes β i en cilindres vermells les hèlixs α . La localització de cada element d'estructura secundària es senyalitza marcant la seqüència amb el mateix color.

Estructura quaternària

Com ja s'ha comentat anteriorment, el complex NeqN/NeqC està format per una subunitat corresponent a la NeqN i una altra més petita corresponent a la NeqC, les quals interaccionen entre si a través de dues cadenes β antiparal·leles, L β_7 -S β_1 .

Per tal d'estimar l'estabilitat de la interfície formada entre els dos fragments del complex NeqN/NeqC es va utilitzar el servidor PISA [185]. L'estabilitat estimada de la interfície dictamina si l'estructura macromolecular pot existir en solució o només és un artefacte del cristall. El servidor per calcular l'estabilitat de la interfície utilitza la variació d'energia lliure i el canvi d'entropia entre l'estat monomèric i el complex.[186]. Com es pot veure en la Taula 4.10, el complex format per les dues inteïnes és estable en solució.

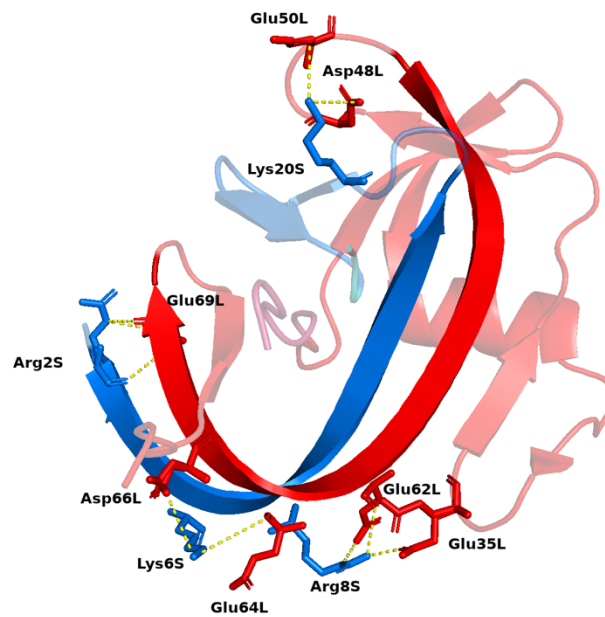
Taula 4.10. Paràmetres fisicoquímics de la interacció predits pel servidor PISA.

Paràmetre	Informació	
Numero monòmers	2	Complex format per una molècula de NeqN i una molècula de NeqC
Fórmula	AB	
Estable	Si	Estable en solució
Àrea de Superfície (Å ²)	7620	Àrea de la superfície exposada a solvent
Àrea Enfonsada (Å ²)	3770	Àrea de la superfície enfonsada
ΔG int (kcal/mol)	-22.1	Guany en energia lliure de solvatació en formar el complex
ΔG dis (kcal/mol)	28.2	Energia lliure de dissociació del complex. Valors positius indiquen que el complex és termodinàmicament estable.
Interfície		
N res B	32	Número de residus de la NeqC que formen la interfície.
Superfície (Å ²) B	3964	
N res A	51	Número de residus de la NeqN que formen la interfície.
Superfície (Å ²) A	7424	
X Àrea de la interfície (Å ²)	1885.9	Àrea de la superfície que forma la interfície d'associació
ΔG int (kcal/mol)	-22.1	Guany en energia lliure en formar el complex
ΔG P-valor	0.385	Indica el p-valor del guany d'energia lliure de solvatació observada. Valors <0.5 indiquen que la interfície té un caràcter hidrofòbic i per tant que la interacció observada és específica.
N _{PH}	35	Nombre de ponts d'hidrogen presents en la interfície.
N _{SH}	5	Nombre de ponts salins presents en la interfície.

L'associació dels dos fragments en el complex es produeix per l'establiment d'un gran nombre de ponts d'hidrogen entre les cadenes laterals i principals dels residus que conformen la interfície del complex (Annex K) i la formació de nombroses interaccions de caràcter hidrofòbic, d'acord amb PISA (p-valor<0.5) [185]. De les interaccions electroestàtiques trobades cal destacar quatre grups: (1) Glu35L, Glu62L i Arg8S; (2) Glu64L, Asp66L, i Lys6S; (3) Glu69L i Arg2S; (4) Asp48L, Glu50L, i Lys20S (Figura 4.53).

La comparació del complex NeqN/NeqC amb el complex Npu (4lx3) [73] indica que els grups trobats d'interaccions electroestàtiques 1, 2 i 3 es poden considerar equivalents als descrits per la Npu [75]. Com era d'esperar, tots els residus bàsics es troben en la subunitat NeqC, i tots els residus àcids, excepte el Glu35L, formen part de la cadena β7 de la subunitat NeqN. En l'estructura obtinguda de la NeqN aïllada (5oxw) es pot observar com aquests residus àcids, excepte el Glu69, es troben en regions desestructurades o girs, i tots ells exposats a solvent (Figura 4.54). La distribució de càrregues entre els dos fragments de la inteïna i la localització dels residus àcids en la NeqN indiquen que existeix una complementació electrostàtica entre els dos fragments, la qual probablement, igual que en la Npu DnaE [57], sigui essencial en el procés d'associació entre els dos fragments de la NeqN.

a



b

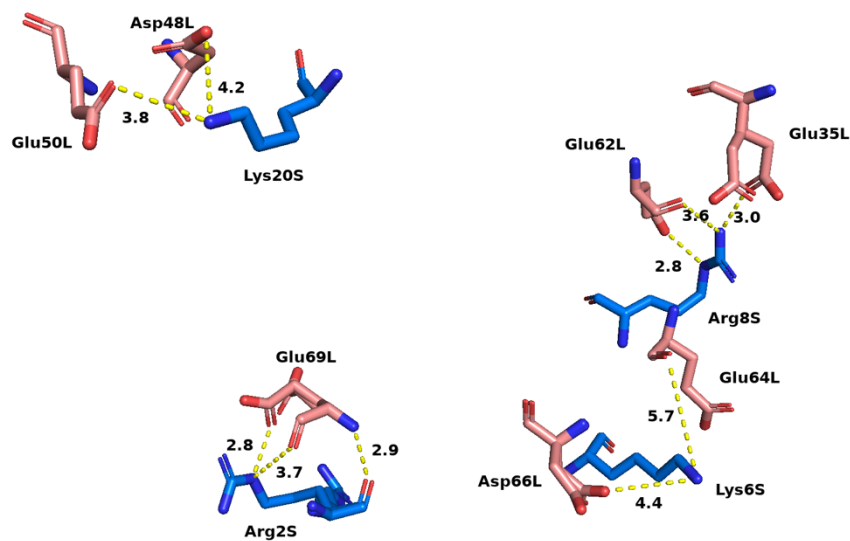


Figura 4.53. Complementació electroestàtica existent entre la NeqN i la NeqC. (a) Estructura del complex 5oxz on es senyalitzen els quatre grups d'interaccions iòniques, les quals probablement desenvolupen un paper clau en el procés d'associació entre la NeqN i la NeqC. Les cadenes polipeptídiques de la NeqN i NeqC es mostren en vermell i en blau, respectivament. (b) Detall dels quatre grups d'interaccions que es donen entre la NeqN i la NeqC. Totes les distàncies entre àtoms es troben en Ansgtoms (Å). Els ponts d'hidrogen es representen com a línies discontinües. Els residus que formen part de la NeqN estan indicats amb la lletra L i els que corresponent a la NeqC amb la lletra S.

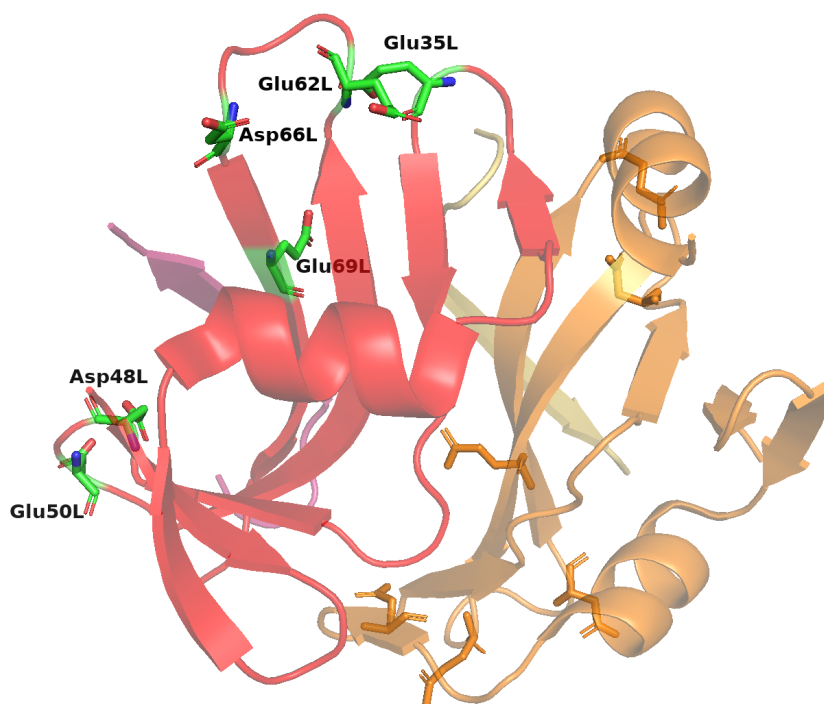


Figura 4.54. Estructura cristal·logràfica de la Neq N aïllada (5oxw), on es ressalten els residus àcids implicats en el procés d'associació. Dímer format per dues subunitats de NeqN (monòmer B i C). Les cadenes polipeptídiques del monòmer B i C es mostren en vermell i taronja, respectivament. L'exteïna N del monòmer B es mostra en rosa i l'exteïna N del monòmer C en groc. En groc es troba senyalitzada la interfície dimèrica. Com es pot observar els residus àcids que posteriorment participen en interaccions electroestàtiques amb la NeqC, es troben en regions desestructurades o girs, excepte el Glu69. D'altra banda, s'observa que tots els residus àcids es troben en la cara oposada de la interfície dimèrica, quedant d'aquesta manera exposats a solvent i facilitant la interacció entre les dues subunitats. Els residus que formen part de la NeqN estan indicats amb la lletra L i els que corresponen a la NeqC amb la lletra S.

Paper de les exteïnes

Com s'ha comentat anteriorment, els residus nadius que conformen l'exteïna N i l'exteïna C, en el complex NeqN/NeqC es troben desordenats. Aquest fet es pot atribuir a la petita mida dels segments de l'exteïna presents en aquest precursor, els quals no aconseguen plegar-se en absència de la resta dels residus nadius de l'exteïna, però també podria ser degut al fet que la inteïna necessiti certa flexibilitat en aquestes regions d'unió exteïna-inteïna per a dur a terme la reacció de *splicing* [46].

Tot i no presentar cap estructura ordenada, en l'estructura del complex NeqN/NeqC s'observen interaccions específiques entre l'exteïna N i l'exteïna C (interaccions 3-4 de la Taula 4.11; Figura 4.55), les quals poden tenir una funció indirecta en la reacció de *splicing* des d'un punt de vista estructural. Per altra banda, l'exteïna N podria desenvolupar un paper essencial en la reacció de

splicing, ja que l'Asp-1 interacciona a través de dos ponts d'hidrogen amb l'His29S (interaccions 1-2 de la Taula 4.11; Figura 4.55). De fet aquests resultats estructurals concorden amb els resultats obtinguts de *trans-splicing*.

Taula 4.11. Ponts d'hidrogen entre exteïna N i exteïna C, i exteïna N i inteïna C.

#	Nom	Atom	Nom	Atom	Distància (Å)
1	Asp-1L	[OD2]	His29S	[N]	2.86
2	Asp-1L	[N]	His29S	[O]	2.75
3	Gly-2L	[N]	Thr+1S	[OG1]	2.87
4	Ile-4L	[O]	Asp+2S	[N]	3.25
5	Ile-4L	[N]	Asp+2S	[O]	3.38

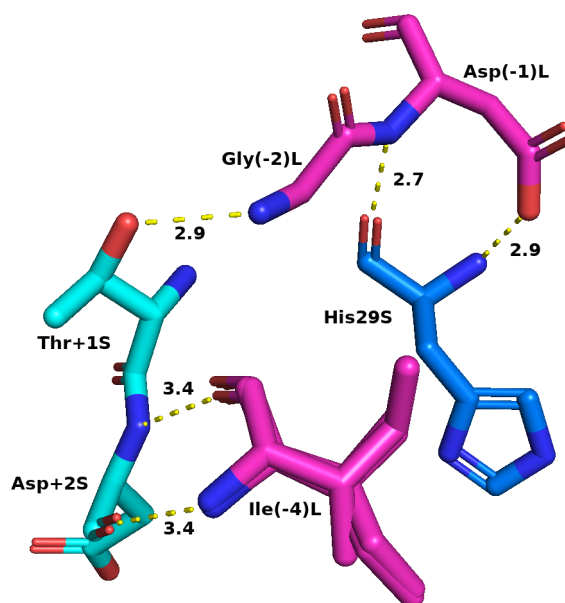


Figura 4.55. Distàncies entre els residus de l'exteïna N (rosa) i la NeqC (blau) i l'exteïna C (turquesa). Els ponts d'hidrogen es representen com a línies discontinües. Les distàncies es troben mesurades en Å. Els residus que formen part de la NeqN estan indicats amb la lletra L i els que corresponen a la NeqC amb la lletra S.

Canvis conformacional en la NeqN per la formació del complex NeqN/NeqC

La comparació de l'estructura de la NeqN aïllada amb la NeqN formant complex revela que la NeqN experimenta canvis conformacionals durant la formació del complex, els quals no impliquen una pèrdua o guany d'estructura global. Aquest resultat concorda amb els espectres obtinguts per CD (Figura 4.48). Els canvis observats afecten principalment a les regions N- i C-

terminal de la NeqN (Figura 4.56). A continuació, es descriuen els canvis conformacionals experimentats per la NeqN durant la formació del complex (Figura 4.56, Figura 4.57).

En el monòmer de la NeqN la cadena $\beta 8$ (residus 67-74), $\beta 7$ (residus 52-70) i $\beta 1$ (residus (-7)-(-1)) formen un full β antiparal·lel (Figura 4.56, Figura 4.57). Com s'ha vist anteriorment l'associació dels dos fragments de la inteïna es dona principalment gràcies a l'establiment de nombroses interaccions no covalents entre les cadenes $\beta 7$ de la NeqN i $\beta 1$ de la NeqC, com a conseqüència la cadena $\beta 1$ de la NeqN aïllada desapareix, la cadena $\beta 7$ s'esdevé més llarga, i la cadena $\beta 8$ es torna més curta i canvia de posició. A més a més, els residus de l'exteïna N perden la seva organització regular en la cadena $\beta 1$ i es desplacen cap al centre catalític.

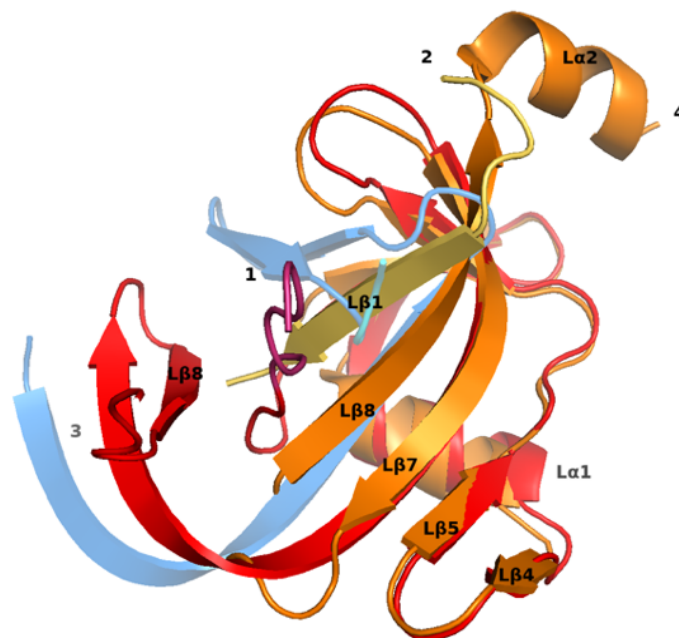


Figura 4.56. Sobreposició de la NeqN monomèrica (50xw) i formant complex amb la NeqC (50xz). En l'estructura de la NeqN monomèrica el fragment N-terminal de la inteïna es representa en taronja i els residus de l'exteïna N en groc. La NeqN, la NeqC, els residus de l'exteïna N i de l'exteïna C estan representats en vermell, blau, rosa i turquesa, respectivament. 1- N-terminal de l'exteïna N formant el complex; 2- N-terminal de l'exteïna N monomèrica; 3- C-terminal de la NeqN formant el complex; 4- C-terminal de la NeqN monomèrica.

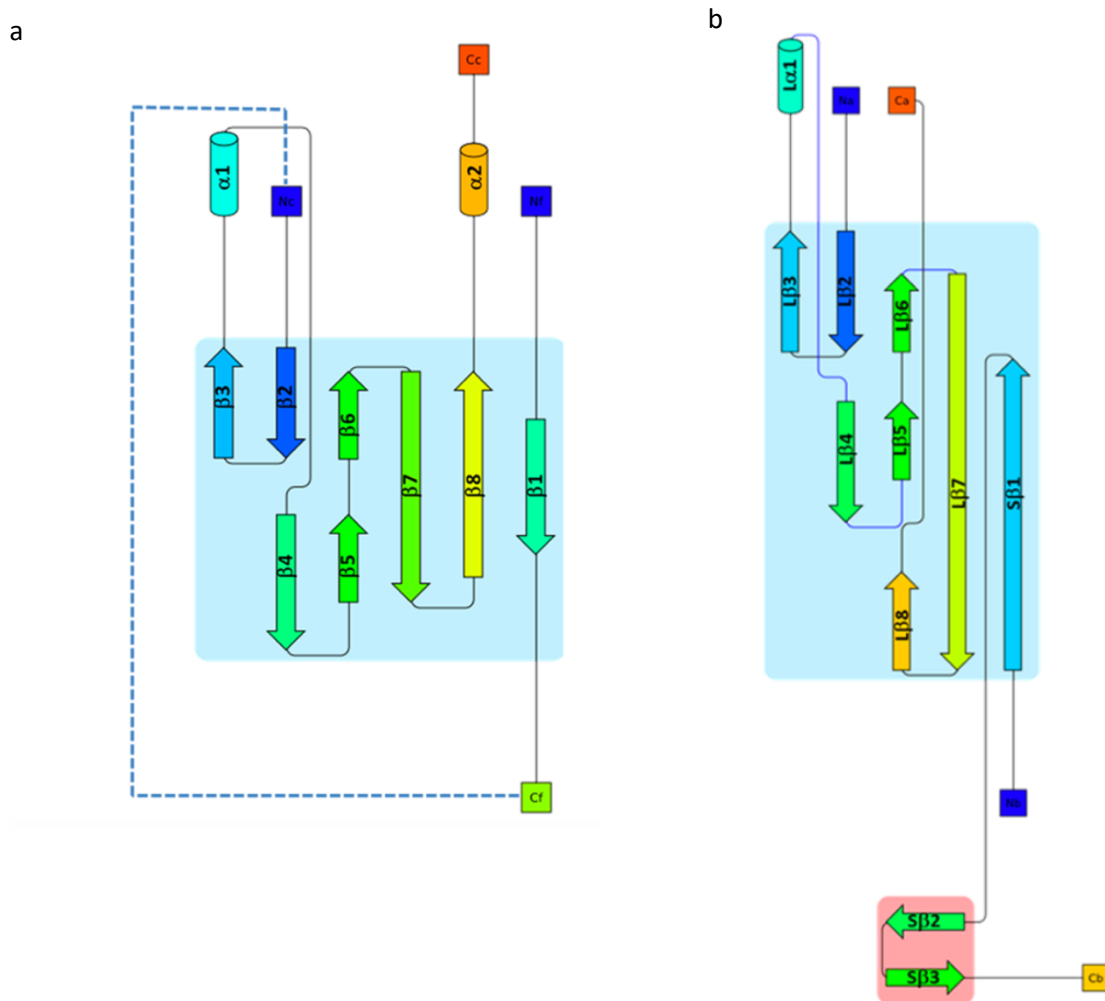


Figura 4.57. Esquema de l'estructura secundària de la NeqN aïllada i en el complex NeqN/NeqC. (a) Representació de la topologia de la NeqN. L'hèlix $\alpha 2$ només es troba en els monòmers C i D. Els residus de l'exteina N formen la cadena $\beta 1$, mentre que els cinc residus de l'extrem N-terminal de la NeqN no s'han pogut modelar (lineal discontinua). El full β que caracteritza la topologia de la NeqN es troba marcat en blau. (b) representació de l'estructura secundària del complex NeqN/NeqC. Els elements d'estructura secundària de la NeqN i de la NeqC estan indicats amb la lletra L i S, respectivament. La numeració dels elements d'estructura secundària del complex ha seguit sempre que ha estat possible l'enumeració de l'estructura de la NeqN aïllada. La cadena $\beta 1$ en el complex és absent on els residus de l'exteina N no adopten cap estructura secundària definida. Els dos fulls β es troben marcats en blau i en vermell. Les dues representacions s'han generat utilitzant DSSP [145] i pro-origami [174]

Cal destacar que en el monòmer de la NeqN el Trp en posició 52 es troba immobilitzat per l'establiment interaccions no covalents en un entorn hidrofòbic. En el cas dels monòmers A i B la interacció es dona entre el grup carbonil del Trp52, el qual forma un pont d'hidrogen amb l'enllaç amida que uneix la Thr74 i la Leu 75. En canvi en els monòmers C i D aquesta interacció es produeix entre els anells aromàtics del Trp52 i la Tyr47 (Figura 4.58). En el complex NeqN/NeqC aquest Trp queda lliure i la seva cadena lateral queda exposada al solvent. Aquest canvi de posició en la cadena lateral del Trp52 explica els espectres obtinguts per CD en la regió de l'UV proper.

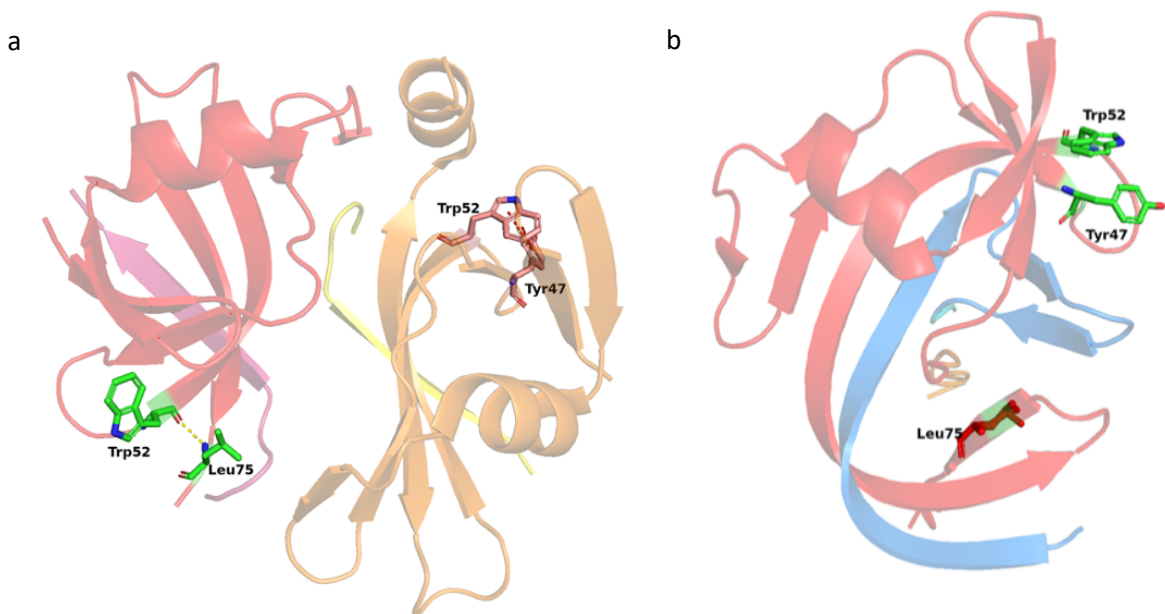


Figura 4.58. Efecte sobre el Trp52 del canvi conformacional de la NeqN provocat per l'associació amb la NeqC. (a) Dímer format pels monòmers B (extrem N-terminal de la inteïna indicat en vermell i els residus de l'exteïna N en rosa) i C (en taronja s'indica la NeqN i en groc els residus de l'exteïna N). El pont d'hidrogen establert entre el Trp52 i la Leu75 en el monòmer B es troba representat per una línia discontinua groga, i la interacció π entre el Trp 52 i la Tyr 47 del monòmer C per línia discontinua vermella. L'establiment d'aquestes interaccions manté el residu de Trp en un entorn ordenat causant dels espectres de CD en la regió UV-proper obtinguts; (c) Complex NeqN/NeqC on es pot apreciar la llibertat de moviment en la cadena lateral del Trp52. La NeqN i la NeqC del complex es troben representades en vermell i blau.

Relació estructural amb altres membres de la família

Les anàlisis comparatives de seqüències i estructures proteïques tenen un paper clau en la comprensió de les funcions de les proteïnes. Assumint que existeix una continuïtat evolutiva entre estructura i funció, descriure les relacions de semblança estructural entre proteïnes permet conèixer millor la proteïna d'estudi.

Amb la finalitat de trobar inteïnes homòlogues estructuralment a la NeqN es va utilitzar el servidor DALI. Aquest servidor on-line va permetre comparar l'estructura resolta de la NeqN formant complex, 5oxz, amb la resta d'estructures dipositades en el Protein Data Bank (PDB) [187]. L'homologia estructural en aquest servidor es mesura amb la puntuació Z, valors per sobre de 2 indiquen que existeix una homologia significativa entre les proteïnes, generalment correspon a plegaments similars. Fortes homologies es caracteritzen per presentar una identitat seqüencial per sobre del 20% o bé una puntuació Z per sobre del límit, el qual depèn de la mida de la proteïna d'estudi. El límit s'estableix empíricament com $n/10 - 4$, on n és el nombre de residus de l'estructura [188]. En el cas de la NeqN el valor Z límit és 5. Tenint-ho en compte s'han

trobat quatre inteïnes d'arqueobacteris hipertermòfils que presenten una forta homologia amb la Neq pol, , *P. furiosus* RIR1-1 (Pfu RIR1-1; Z-score = 7.0; PDB code 1dq3), *T. kodakaraensis* Pol-2 (Tko Pol-2; Z-score = 6.8; PDB code 2cw7), *P. horikosii* RadAmini (Pho RadAmini; Z-score = 6.7; PDB code 4e2u) i *P. abyssi* PolIII (Pab PolIII; Z-score = 5.5; PDB code 2lcj). Cap d'aquestes inteïnes és una inteïna partida, per lo que es va creure necessari incorporar en l'estudi comparatiu la inteïna partida Npu DnaE (PDB code 4lx3), la qual mostra una puntuació Z per sota de la límit, 4.3. Aquests resultats indiquen que en l'estructura tridimensional de la inteïna Neq pol predominen els trets característics de les inteïnes hipertermofíliques sobre els de les inteïnes partides. En la Figura 4.59 es mostra l'alineament seqüencial basat en l'estructura de la NeqN amb les seves homòlogues. En aquest alineament es mostra que la identitat seqüencial de la NeqN les altres inteïnes és baixa tot i tenir una estructura tridimensional similar, la qual cosa suggereix que probablement aquesta inteïna presenti un origen evolutiu diferent.

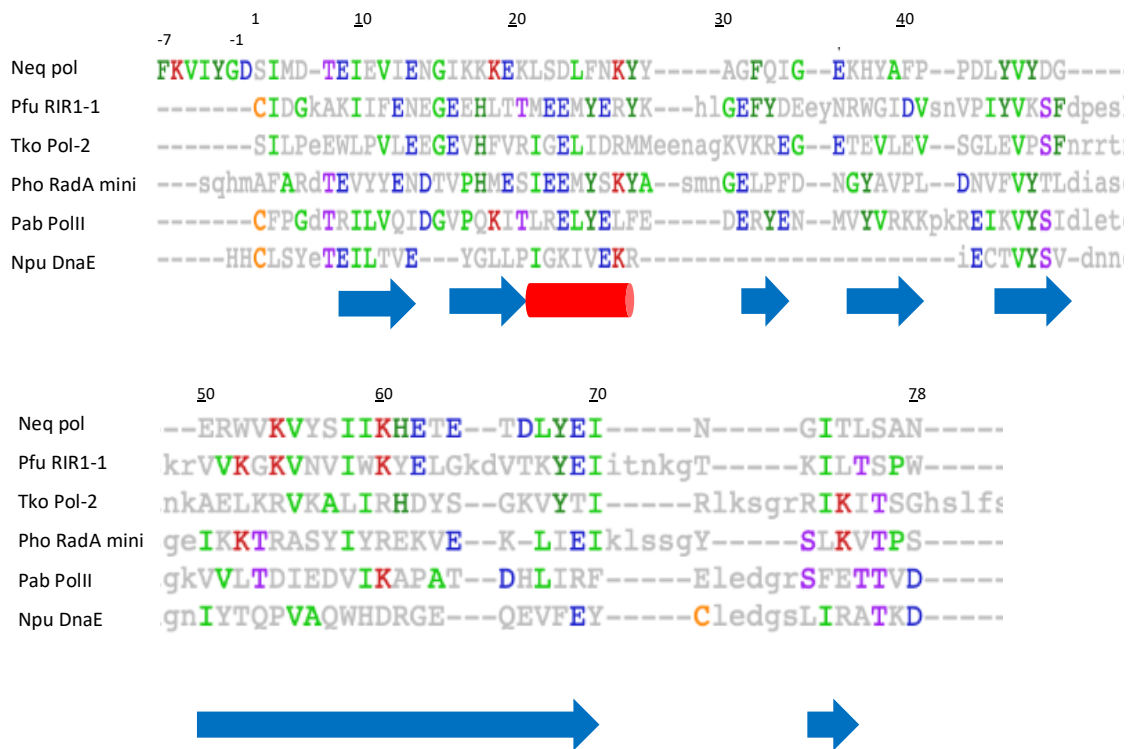


Figura 4.59. Alineament seqüencial basat en l'estructura de la NeqN formant complex amb altres inteïnes les quals presenten una elevada homologia estructural. L'enumeració dels residus de la NeqN es troben a sobre de l'alineació. A sota de l'alineació s'indiquen els elements d'estructura secundària que componen la NeqN. Les lletres en majúscula indiquen posicions estructuralment equivalents en la NeqN, i les lletres en minúscula insercions relatives. Els tipus d'aminoàcids més freqüents es troben acolorits en cada columna. Fletxes blaves: Cadenes β ; Cilindres vermells: hèlix α .

Com es pot observar en la Figura 4.60 la superposició de l'estructura tridimensional de la Neq pol amb les inteïnes Pfu RIR1-1, Tko Pol-2, Pho RadAmini, Pab Pol-2 i Npu DnaE és bona. Tanmateix, es pot observar que tant el domini endonucleasa present en les inteïnes Pfu RIR1-1 i Tko Pol-2, com el domini extra Stirrup que es troba en la inteïna Pfu RIR1-1, no interfereixen en el plegament del domini HINT (RMSD= 1.9 Å per l'alineació entre Neq pol i Pfu RIR1-1 (1dq3); RMSD= 1.5 Å per l'alineació entre Neq pol i Tko Pol-2 (2cw7)). Com ja s'ha comentat anteriorment el plegament del domini HINT es caracteritza per adoptar una estructura semblant a una ferradura, la qual es divideix en dos lòbuls. La superposició de la Neq pol amb les diferents inteïnes mostra que la Neq pol és la inteïna amb el segon lòbul més petit.

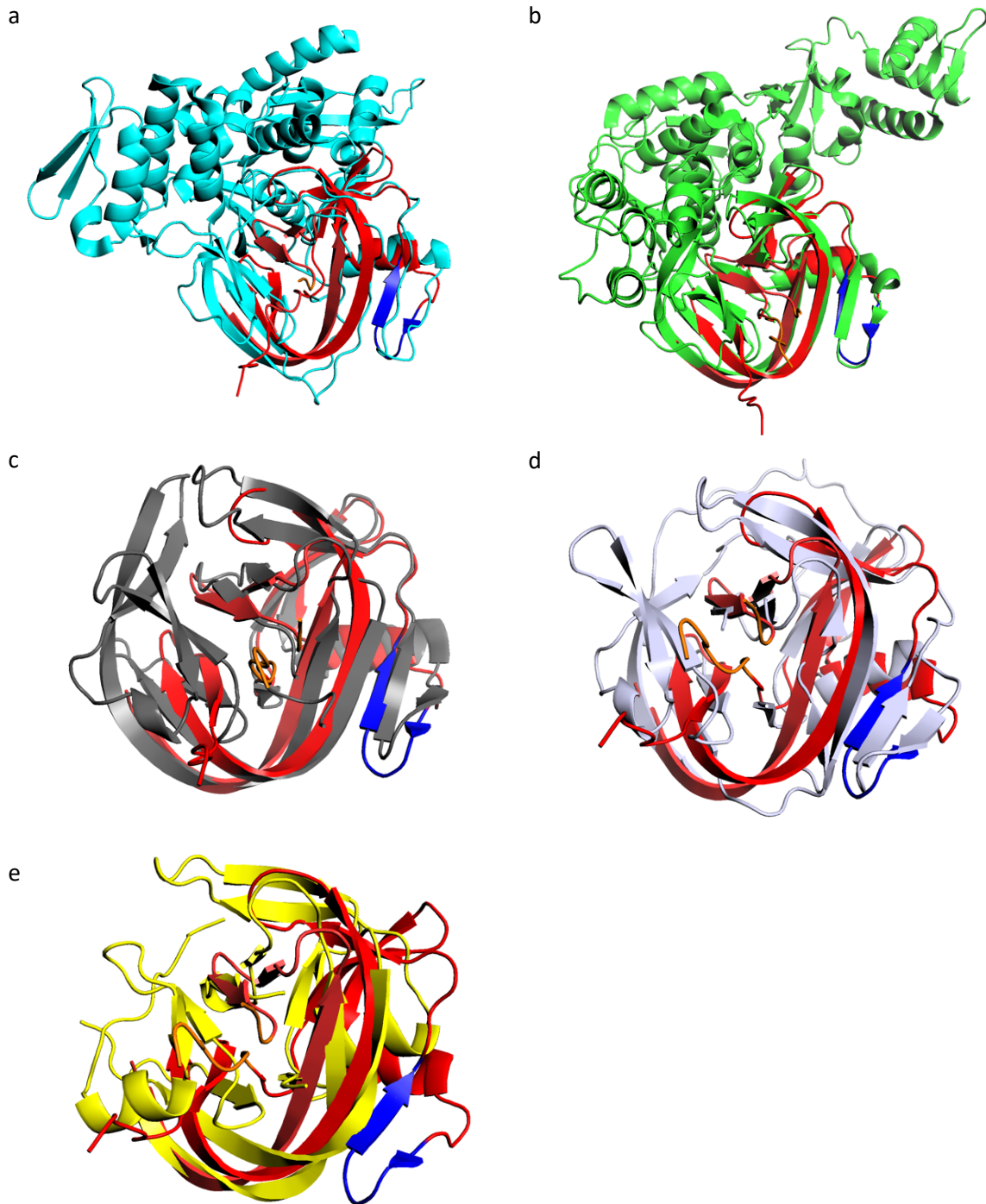


Figura 4.60. Superposició de les estructures tridimensionals de la *Neq pol* amb altres inteïnes homòlogues. Aquestes superposicions permeten veure del domini HINT i la forquilla hipertèrmofílica (HTH) (senyalitzada en blau). (a) Superposició amb la inteïna Pfu RIR1-1 (RMSD= 1.9 Å; PDB 1dq3); (b) superposició amb la inteïna Tko pol-2 (RMSD= 1.5 Å; PDB 2cw7); (c) superposició amb la inteïna Pho RadAmini (RMSD= 2.1 Å; PDB 4e2u); (d) superposició amb la inteïna Pab pol-II (RMSD= 3.6 Å; PDB 2lcj); (e) superposició amb la inteïna Npu DnaE (RMSD= 8.1 Å; PDB 4lx3). La inteïna *Neq pol* i les seves exteïnes s'assenyalen en vermell i taronja, respectivament.

4.3.4. Centre catalític de la Neq pol: Implicacions en el mecanisme de *splicing*

Com s'ha vist en la introducció d'aquesta tesi els quatre blocs que contenen els residus essencials per a la reacció de *splicing* comuns a totes les inteïnes estan dividits en les inteïnes partides, dos es troben en la inteïna N (A i B) i els altres dos en la inteïna C (F i G) [6]. Les estructures obtingudes en aquesta tesi permeten reconsiderar els blocs definits per a la Neq pol, proporcionant un marc estructural únic per aquesta reacció de *splicing* on no participa cap Cys, en la qual sembla que juguen un paper clau els residus de l'exteïna N i l'exteïna C.

Formació de l'intermediari lineal

Wang i col·laboradors han proposat la His10 del bloc B com a residu essencial per a la formació de l'intermediari lineal [30]. En el cas de la Neq pol la His79L del bloc B es troba massa lluny de la Ser1L per a desenvolupar el rol descrit per Wang i col·laboradors (14.8 Å entre el C α de la Ser1L i la His79L). En el cristall hexagonal, la Ser1L de la Neq pol adopta dues conformacions diferents, en una el grup hidroxil interacciona amb l'àtom NE de l'anell d'imidazol de la His61L, 3.5 Å (Figura 4.61). D'aquesta manera, la His61L podria actuar com a base i facilitar la desprotonació del grup hidroxil de la Ser1L iniciant el desplaçament N-O acil. No obstant això, la distància entre l'amida de la cadena principal de la Ser1L i l'àtom ND de l'His61L és força llarga (6.7 Å), especialment si es compara amb altres estructures d'inteïnes de classe 1 (~ 2.8 Å) [44], [189]–[191], perquè la His61L actuï com a àcid protonant el grup amida sortint. En el cas de la Neq pol aquesta funció la podria realitzar l'Asp-1, el qual podria ser l'encarregat de protonar el grup amida sortint de l'intermediari lineal. L'oxigen del carboxilat de la cadena lateral de l'Asp-1 es troba formant un pont d'hidrogen amb una molècula d'aigua, la qual a la vegada interacciona amb el grup amida de la Ser1L (Figura 4.61). Tanmateix, aquesta molècula d'aigua estableix simultàniament ponts d'hidrogen amb oxígens de la cadena principal de la Ser1L i Ser76L i amb el grup hidroxil de la Ser76L, el que recolza la seva funció en el desenvolupament de l'intermediari tetraèdric, rol que normalment desenvolupa la histidina del bloc B.

Un altre residu que es proposa com a important per l'activitat de les inteïnes és l'Asp del bloc F. En la Neq pol, l'Asp13S adopta diferents conformacions (Figura 4.61), en una d'elles el grup carboxilat de la cadena lateral interacciona amb el grup hidroxil de la Ser1L (3.3 Å). La cadena lateral del Asp13S protonada podria actuar com a donadora d'hidrògens, estabilitzant l'oxianió de la Ser1L i facilitant la formació de l'anell oxioxazolidina en l'intermediari tetraèdric, de manera similar al proposat per la inteïna Mtu RecA [52].

Altres estudis indiquen que una distorsió en el pla de l'enllaç peptídic -1 facilita l'atac nucleofílic de la histidina del bloc B [41], [45], [191], [192]. En les estructures del complex de la Neq pol no s'observen grans desviacions del pla de l'enllaç peptídic -1, la qual cosa concorda amb el fet que el grup ND de l'anell d'imidazol es trobi lluny del N amida de la Ser1L.

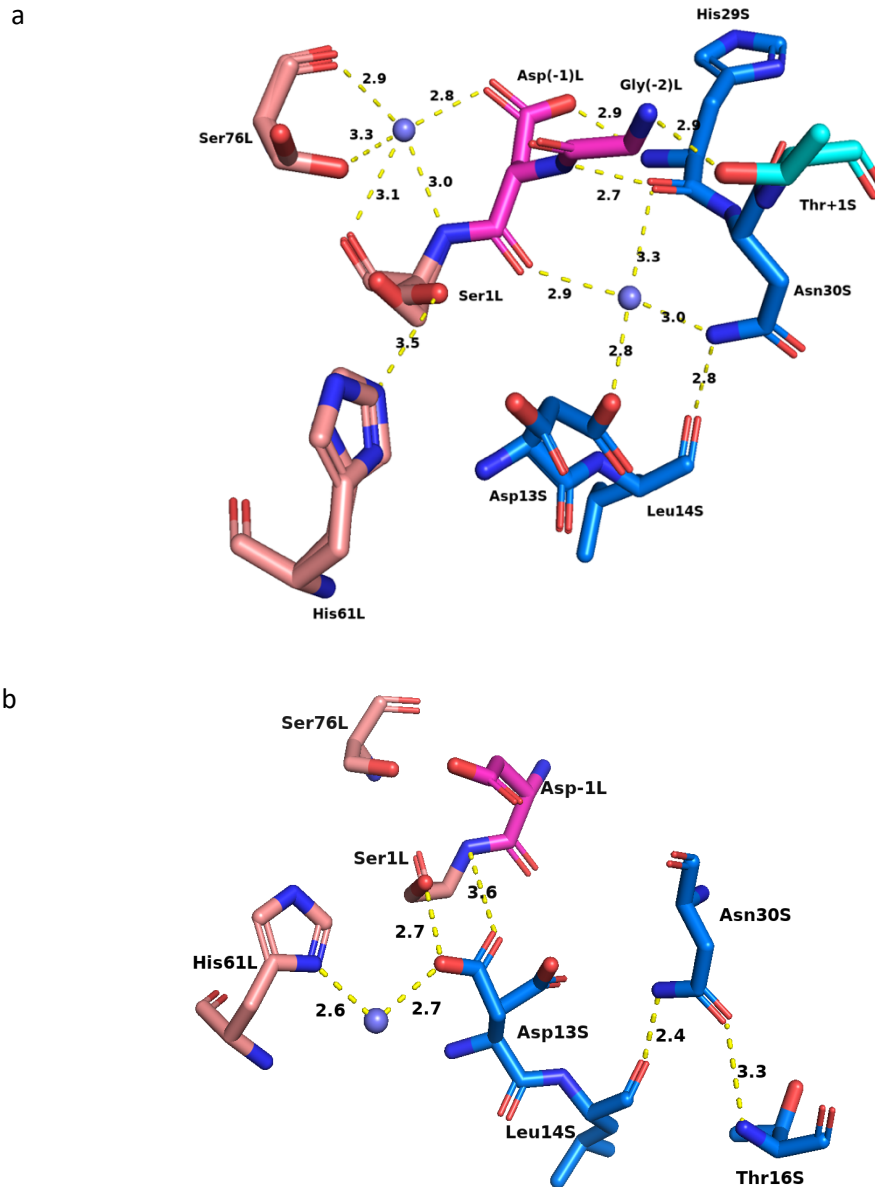


Figura 4.61. Centre actiu de la Neq pol. (a) Obtingut a partir dels cristalls hexagonals. En una de les conformacions, la His61 interacciona amb el grup hidroxil de la Ser1. Una molècula d'aigua interacciona amb l'enllaç amida de la Ser1L i l'oxigen del carboxilat de l'Asp-1. (b) En els cristalls tetragonals no s'observen els residus de l'exteïna C. En una de les conformacions, l'Asp13S interacciona amb el grup hidroxil de la Ser1 i amb l'amida de l'Asp-1. Els residus del fragment N- i C-terminal de la inteïna estan indicats en vermell i blau, respectivament. En rosa i turquesa es troben indicats els residus de les exteïna N i C. Les molècules d'aigua es troben representades com esferes blaves. Els residus de la NeqN i de la NeqC estan indicats amb la lletra L i S, respectivament. Les distàncies entre els residus s'indiquen amb línies discontinües grogues i estan mesurades amb Å.

Trans-Esterificació

En la *trans*-(tio)esterificació de l'exteïna N l'exteïna N passa d'estar unida del primer residu catalític de la inteïna a la cadena lateral del primer residu de l'exteïna C, mitjançant un enllaç (tio)-ester, formant-se d'aquesta manera l'intermediari ramificat. Aquest pas es troba poc caracteritzat, però estudis recents suggereixen que l'Asp del bloc F juga un paper clau. L'estructura del complex NeqN/NeqN aporta per primer cop la distància entre els dos nucleòfils (~ 7.8 Å entre els hidroxils de la Ser1L i la Thr+1S, i ~ 7.4 Å entre el grup carbonil de l'Asp(-1)L i de l'hidroxil de la Thr+1). En una de les conformacions observades per la Ser1L i l'Asp13S, l'Asp es troba interaccionant amb la Ser1L i l'His61L, fent possible la seva participació en el primer pas de la reacció de *splicing* com s'ha descrit anteriorment (Figura 4.61-b). En una altra conformació, el grup carboxílic de la cadena lateral de l'Asp13S es dirigeix cap a l'enllaç peptídic entre l'Asn30S i la Thr+1, tot i que la distància és molt gran (6.8 Å entre el grup OD2 de l'Asp13S i el grup OG de la Thr+1) per a l'activació del grup hidroxil de la Thr+1, tret que es produeixi un canvi conformacional que els acosti i/o que participi una molècula d'aigua (Figura 4.61-a). D'altra banda, en aquesta conformació l'Asp13S interacciona a través d'una molècula d'aigua amb la cadena lateral de l'Asn30S (Figura 4.61-a), la qual catalitza la hidròlisi de l'intermediari ramificat després de la *trans*-esterificació. La interacció entre l'Asp13S i l'Asn30S col·loca la cadena lateral de l'Asn lluny de l'enllaç peptídic a hidrolitzar, prevenint d'aquesta manera la ciclació de l'Asn. Aquesta interacció implica que és necessari que es produeixi un canvi conformacional en l'estructura per a què la cadena lateral de l'Asn cicli. Aquest tipus d'interacció s'ha trobat en la inteïna RadA de *Pyrococcus horikoshii* [193].

Ciclació de l'Asn i alliberació de la inteïna

En l'últim pas de la reacció de *splicing* juga un paper essencial l'Asn del bloc G, el nitrogen de la cadena lateral, el qual ataca nucleofilicament el carboni carbonil de l'enllaç peptídic +1, formant-se succinimida. En l'estructura del complex NeqN/NeqC la cadena lateral de l'Asn30S interacciona establint ponts d'hidrogen amb els àtoms O i N de l'esquelet polipeptídic de la Leu14S i la Thr16S, respectivament (Figura 4.61-b). Així com ja s'ha comentat anteriorment també amb el grup OD2 de l'Asp13S a través d'una molècula d'aigua. Les interaccions directes i indirectes amb els residus Leu14S i Asp13S, respectivament, fan que la cadena lateral de l'Asn estableixi una conformació desfavorable per a la seva ciclació. Com ja s'ha comentat anteriorment, aquests resultats indiquen que perquè es produeixi aquest últim pas de la reacció és necessari un canvi conformacional. Això podria ser una estratègia per a minimitzar les reaccions d'hidròlisi de l'extrem C-terminal en un context natiu.

5. Conclusions

- 1- Els residus flanquejants de les regions d'unió exteïna-inteïna són essencials perquè es produeixi la reacció de *trans-splicing*. En aquest treball s'ha demostrat que la inteïna Neq pol tan sol catalitza la reacció de *trans-splicing* de manera coordinada quan en el precursor N es conserva la seqüència peptídica de set residus de l'exteïna salvatge i en el precursor C es mantenen com a mínim tres residus salvatges. En l'estructura cristal·logràfica obtinguda per difracció de raigs X del complex NeqN/NeqC (oxz) s'observa com els residus Ile-4, Gly-2 i Asp-1 de l'exteïna N interaccionen específicament amb l'His29 de la NeqC i els residus Thr+1 i Asp+2 de l'exteïna C. Probablement aquestes interaccions orienten correctament el centre catalític de la inteïna.
- 2- La reacció de *trans-splicing* catalitzada per la inteïna partida Neq pol depèn de la temperatura atès que no s'observa la reacció a temperatures per sota de 50°C.
- 3- En aquelles reaccions on el precursor N no conserva la seqüència salvatge de l'exteïna N, la inteïna Neq pol no és capaç de coordinar les diferents etapes que constitueixen la reacció de *trans-splicing*, produint-se les reaccions d'hidròlisi de l'extrem N- i C-terminal. S'ha comprovat també que la ciclació de la cadena lateral de l'Asn es dona de manera prematura, i comporta la hidròlisi de l'extrem C-terminal. Aquesta reacció és independent de la temperatura. Al contrari, sembla que la hidròlisi de l'extrem N-terminal es produeix només a temperatures superiors als 50°C, indicant que la temperatura modula la primera etapa de la reacció de *trans-splicing* de la Neq pol.
- 4- La substitució de la Ser en posició 1 per Cys provoca que l'eficiència de la reacció de *trans-splicing* sigui més elevada, ja que s'augmenta tant la velocitat de la reacció com el rendiment d'aquesta.
- 5- La NeqN és resistent a la desnaturalització per SDS, per urea i a la digestió per proteases el que indica que es tracta d'una proteïna cinèticament estable, la qual presenta una inusual rigidesa en la seva estructura nativa, propietat característica de les proteïnes hipertermofíliques. Presenta una T_m de 105 °C i el seu desplegament és irreversible.
- 6- L'estructura cristal·logràfica obtinguda de la NeqN lliure es va resoldre amb una resolució de 2.6 Å a partir de cristalls monocíclics. Cada cristall està format per quatre monòmers per unitat asimètrica que estan organitzats en dos dímers. D'acord amb els resultats obtinguts per SEC i ultracentrifugació, el dímer és l'estat predominant de la NeqN en solució a concentracions elevades. Probablement, la dimerització sigui un dels mecanismes adoptats per la inteïna per augmentar la seva termoestabilitat, ja que d'aquesta manera es redueix l'àrea de superfície hidrofòbica exposada a solvent.

- 7- El monòmer NeqN es troba estabilitzat per nombrosos ponts d'hidrogen intramoleculars i intermoleculars entre la inteïna i l'exteïna i entre els monòmers que conformen el dímer. També s'han observat ponts salins i interaccions de tipus π . Aquestes interaccions no s'han descrit en altres inteïnes partides, la qual cosa suggereix que poden jugar un paper important tant en el manteniment de l'estructura nativa com en la termoestabilitat de la proteïna.
- 8- La velocitat d'associació entre els fragments NeqN/NeqC augmenta gradualment amb la temperatura. El seu procés d'associació presenta una elevada energia d'activació (89.5 KJ/mol), característica de proteïnes molt estables, les quals necessiten trencar enllaços per associar-se. Probablement, aquesta elevada energia d'activació sigui conseqüència de què per poder formar el complex NeqN/NeqC s'han de trencar interaccions que estabilitzen el dímer NeqN. Aquests dos fragments s'associen amb una elevada afinitat, presentant un valor de K_d de, 18.6 nM a 25°C. El procés és espontani amb un valor de ΔG (25°C)=-1.73 kcal·mol⁻¹ essent desfavorable des d'un punt de vista entròpic, ja que presenta un valor de $\Delta H^\circ = 2.36 \pm 1.5$ kcal·mol⁻¹.
- 9- D'acord amb els resultats obtinguts es proposa el següent mecanisme d'associació entre els dos fragments de la inteïna partida Neq pol:
- La NeqN lliure es troba en un estat dimèric i la NeqC desplegada. En un primer contacte governat per interaccions electroestàtiques entre la NeqC i la NeqN dimèrica, la NeqC promou la desestabilització de la forma dimèrica de la NeqN. Aquest procés, com qualsevol altre, es veu afavorit per la temperatura, però no està controlat per aquesta. Després es produeix una reorganització estructural formant-se el complex NeqN/NeqC, on la NeqC assoleix la seva estructura terciària.
- 10- La T_m del complex NeqN/NeqC és la mateixa que per la NeqN lliure, 105°C, la qual cosa indica que l'associació de la NeqC no estabilitza la NeqN.
- 11- El complex NeqN/NeqC adopta l'estructura en ferradura ordenada característica dels dominis HINT. En canvi, els residus que es troben en els llocs d'unió exteïna-inteïna es troben desordenats, la qual cosa suggereix que la inteïna necessiti certa flexibilitat en aquestes regions per dur a terme la reacció de *splicing*.
- 12- La interfície del complex NeqN/NeqC està estabilitzada per nombrosos ponts d'hidrogen i interaccions hidrofòbiques entre els residus que la conformen i per complementarietat electroestàtica la qual segurament és molt important en el procés d'associació entre els fragments de la Neq pol.

- 13- En l'estructura tridimensional de la NeqN formant complex amb la NeqC predominen els trets característics de les inteïnes hipertermofíliques sobre els de les inteïnes partides mesòfiles. En l'estructura de la NeqN es troba la forquilla β HTH, formada per les cadenes β 4 i 5, la qual no es troba en cap de les estructures resoltes d'inteïnes mesòfiles partides.
- 14- Amb relació al mecanisme de *splicing*, el paper àcid/base descrit per la His del bloc B en la formació de l'intermediari lineal, en la inteïna Neq pol el duria a terme la His61 i l'Asp-1.
- 15- L'Asp13S pot tenir un doble paper en la reacció de *trans-splicing* i intervenir en la formació de l'intermediari lineal i ramificat. En el primer pas de la reacció pot actuar com a donador de protons estabilitzant l'oxianió en l'intermediari tetraèdric. En la formació de l'intermediari ramificat l'Asp13S podria comportar-se com una base i captar un protó del grup hidroxil de la Thr, encara que la distància entre el grup carboxilat de l'Asp13S i el grup hidroxil de la cadena lateral de la Thr és molt gran perquè es produeixi, el que indica que s'ha de donar un canvi conformacional que ho permeti.
- 16- L'Asn30S interacciona amb els residus Leu14S, Asp13S i Thr16, provocant que aquest residu adopti una conformació desfavorable per la ciclació de la seva cadena lateral. Aquest potser un mecanisme per a minimitzar les reaccions d'hidròlisi de l'extrem C-terminal.

6. Bibliografia

- [1] C. W. Lennon and M. Belfort, "Inteins," *Curr. Biol.*, vol. 27, no. 6, pp. 204–206, 2017.
- [2] R. Hirata, Y. Ohsumi, A. Nakano, H. Kawasaki, K. Suzuki, and Y. Anraku, "Molecular structure of a gene, VMA1, encoding the catalytic subunit of H⁺-translocating adenosine triphosphatase from vacuolar membranes of *Saccharomyces cerevisiae*," *J. Biol. Chem.*, vol. 265, no. 12, pp. 6726–6733, 1990.
- [3] P. M. Kane, C. T. Yamashiro, D. F. Wolczyk, N. Neff, M. Goebel, and T. H. Stevens, "Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H⁺-adenosine triphosphatase," *Science*, vol. 250, no. 4981, pp. 651–657, 1990.
- [4] T. M. Tanaka Hall, J. A. Porter, K. E. Young, E. V. Koonin, P. A. Beachy, and D. J. Leahy, "Crystal structure of a Hedgehog autoprocessing domain: Homology between Hedgehog and self-splicing proteins," *Cell*, vol. 91, no. 1, pp. 85-97, 1997.
- [5] T. Pavankumar, "Inteins: Localized Distribution, Gene Regulation, and Protein Engineering for Biological Applications," *Microorganisms*, vol. 6, no. 1, p. 19, 2018.
- [6] F. B. Perler, G. J. Olsen, and E. Adam, "Compilation and analysis of intein sequences," *Nucleic Acids Res.*, vol 25, no. 6, pp. 1087-1093, 1997.
- [7] F. B. Perler, "InBase: InteIn Database," *Nucleic Acids Res.*, vol. 30, no. 1, pp. 383-384, 2002.
- [8] M. L. Pedulla, M. E. Ford, J. M. Houtz, T. Karthikeyan, C. Wadsworth, J. A. Lewis, D. Jacobs-Sera, J. Falbo, J. Gross, N. R. Pannunzio, W. Brucker, V. Kumar, J. Kandasamy, L. Keenan, S. Bardarov, J. Kriakov, J. G. Lawrence, W. R. Jacobs Jr, R. W. Hendrix, and G. F. Hatfull, "Origins of highly mosaic mycobacteriophage genomes," *Cell*, vol. 113, no. 2, pp. 171-182, 2003.
- [9] R. T. M. Poulter, T. J. D. Goodwin, and M. I. Butler, "The nuclear-encoded inteins of fungi," *Fungal Genetics and Biology*, vol. 44, no. 3, pp. 153-179, 2007.
- [10] K. S. Swithers, S. M. Soucy, E. Lasek-Nesselquist, P. Lapierre, and J. P. Gogarten, "Distribution and evolution of the mobile vma-1b intein," *Mol. Biol. Evol.*, vol. 30, no. 12, pp. 2676-2687, 2013.
- [11] O. Novikova, N. Topilina, and M. Belfort, "Enigmatic distribution, evolution, and function of inteins.," *J. Biol. Chem.*, vol. 289, no. 21, pp. 14490–14497, 2014.
- [12] O. Novikova, P. Jayachandran, D. S. Kelley, Z. Morton, S. Merwin, N. Topilina, and Marlene Belfort, "Intein clustering suggests functional importance in different domains of life," *Mol. Biol. Evol.*, vol. 33, no. 3, pp. 783–799, 2016.
- [13] S. Pietrokovski, "Intein spread and extinction in evolution," *Trends in Genetics, Trends Genet*, vol. 17, no. 8, pp. 465–472, 2001.

- [14] K. S. Swithers, A. G. Senejani, G. P. Fournier, and J. P. Gogarten, "Conservation of intron and intein insertion sites: Implications for life histories of parasitic genetic elements," *BMC Evol. Biol.*, vol. 31, no. 9, p. 303, 2009.
- [15] J. P. Gogarten, A. G. Senejani, O. Zhaxybayeva, L. Olendzenski, and E. Hilario, "Inteins: Structure, function, and evolution," *Annual Review of Microbiology.*, vol. 56, pp. 263-287, 2002.
- [16] F. S. Gimble and J. Thorner, "Homing of a DNA endonuclease gene by meiotic gene conversion in *Saccharomyces cerevisiae*," *Nature*, vol. 357, no. 6376, pp. 301-306, 1992.
- [17] A. Burt and V. Koufopanou, "Homing endonuclease genes: The rise and fall and rise again of a selfish element," *Current Opinion in Genetics and Development.*, vol. 14, no.6, pp.609-615, 2004.
- [18] B. S. Chevalier and B. L. Stoddard, "Homing endonucleases: Structural and functional insight into the catalysts of intron/intein mobility," *Nucleic Acids Res.*, vol. 29, no. 18, pp. 3757-3774, 2001.
- [19] Z. Du, J. Liu, C. D. Albracht, A. Hsu, W. Chen, M D. Marieni, K. M. Colelli, J. E. Williams, J. N. Reitter, K. V. Mills, and C. Wang, "Structural and mutational studies of a hyperthermophilic intein from DNA polymerase II of *Pyrococcus abyssi*," *J. Biol. Chem.*, vol. 286, no. 44, pp. 38638–38648, 2011.
- [20] W. Chen, L. Li, Z. Du, J. Liu, J. N. Reitter, K. V. Mills, R. J. Linhardt, and C. Wang, "Intramolecular disulfide bond between catalytic cysteines in an intein precursor," *J. Am. Chem. Soc.*, vol. 134, no. 5, pp. 2500-3, 2012.
- [21] M. C. Nicastrì, K. Xega, L. Li, J. Xie, C. Wang, R. J. Linhardt, J. N. Reitter, and K. V. Mills, "Internal disulfide bond acts as a switch for intein activity," *Biochemistry*, vol. 52, no. 34, pp. 5920–5927, 2013.
- [22] N. I. Topilina, O. Novikova, M. Stanger, N. K. Banavali, and M. Belfort, "Post-translational environmental switch of RadA activity by extein-intein interactions in protein splicing," *Nucleic Acids Res.*, vol. 43, no. 13, pp. 6631-6648, 2015.
- [23] N. I. Topilina, C. M. Green, P. Jayachandran, D. S. Kelley, M. J. Stanger, C. L. Piazza, S. Nayak, M. Belfort, "SufB intein of *Mycobacterium tuberculosis* as a sensor for oxidative and nitrosative stresses," *Proc. Natl. Acad. Sci.*, vol. 112, no. 33, pp. 10348–10353, 2015.
- [24] J. N. Reitter, C. E. Cousin, M. C. Nicastrì, M. V. Jaramillo, and K. V. Mills, "Salt-Dependent Conditional Protein Splicing of an Intein from *Halobacterium salinarum*," *Biochemistry*, vol. 55, no. 9, pp. 1279–1282, 2016.

- [25] C. W. Lennon, M. Stanger, and M. Belfort, "Protein splicing of a recombinase intein induced by ssDNA and DNA damage," *Genes Dev.*, vol. 30, no. 24, pp. 2663–2668, 2016.
- [26] C. W. Lennon, M. Stanger, N. K. Banavali, and M. Belfort, "Conditional Protein Splicing Switch in Hyperthermophiles through an Intein-Extein Partnership," *M. Bio*, vol. 9, no. 1, pp. 2304-2317, 2018.
- [27] T. C. Evans and M. Q. Xu, "Mechanistic and kinetic considerations of protein splicing," *Chem. Rev.*, vol. 102, no. 12, pp. 4869-4884, 2002.
- [28] L. Saleh and F. B. Perler, "Protein splicing In Cis and In Trans," *Chem. Rec.*, vol. 6, no. 4, pp. 183–193, 2006.
- [29] A. Telenti, M. Southworth, F. Alcaide, S. Daugelat, W. R. Jacobs, and F. B. Perler, "The Mycobacterium xenopi GyrA protein splicing element: Characterization of a minimal intein," *J. Bacteriol.*, vol. 179, no. 20, pp. 6378-6382, 1997.
- [30] Z. Du, P. T. Shemella, Y. Liu, S. A. McCallum, B. Pereira, S. K. Nayak, G. Belfort, M. Belfort, and C. Wang, "Highly Conserved Histidine Plays a Dual Catalytic Role in Protein Splicing: A pKa Shift Mechanism," *J. Am. Chem. Soc.*, vol. 131, no. 32, pp. 11581–11589, 2009.
- [31] P. Van Roey, B. Pereira, Z. Li, K. Hiraga, M. Belfort, and V. Derbyshire, "Crystallographic and mutational studies of Mycobacterium tuberculosis recA mini-inteins suggest a pivotal role for a highly conserved aspartate residue," *J. Mol. Biol.*, vol. 367, no. 1, pp. 162–173, 2007.
- [32] S. Chong, K. S. Williams, C. Wotkowicz, and M. Q. Xu, "Modulation of protein splicing of the Saccharomyces cerevisiae vacuolar membrane ATPase intein.," *J. Biol. Chem.*, vol. 273, no. 17, pp. 10567–10577, 1998.
- [33] H. Iwai, S. Züger, J. Jin, and P.-H. Tam, "Highly efficient protein trans-splicing by a naturally split DnaE intein from Nostoc punctiforme.," *FEBS Lett.*, vol. 580, no. 7, pp. 1853–1858, 2006.
- [34] G. Amitai, B. P. Callahan, M. J. Stanger, G. Belfort, and M. Belfort, "Modulation of intein activity by its neighboring extein substrates," *Proc. Natl. Acad. Sci.*, vol. 106, no. 27, pp. 11005–11010, 2009.
- [35] N. H. Shah, E. Eryilmaz, D. Cowburn, and T. W. Muir, "Extein residues play an intimate role in the rate-limiting step of protein trans-splicing.," *J. Am. Chem. Soc.*, vol. 135, no. 15, pp. 5839-5847, 2013.

- [36] E. J. Pearl, A. A. M. Bokor, M. I. Butler, R. T. M. Poulter, and S. M. Wilbanks, "Preceding hydrophobic and β -branched amino acids attenuate splicing by the CnePRP8 intein," *Biochim. Biophys. Acta - Proteins Proteomics*, vol. 1774, no. 8, pp. 995–1001, 2007.
- [37] A. Wasmuth, C. Ludwig, and H. D. Mootz, "Structure-activity studies on the upstream splice junction of a semisynthetic intein.," *Bioorg. Med. Chem.*, vol. 21, no. 12, pp. 3495–3503, 2013.
- [38] J. N. Reitter, C. E. Cousin, M. C. Nicastrì, M. V. Jaramillo, and K. V. Mills, "Salt-Dependent Conditional Protein Splicing of an Intein from *Halobacterium salinarum*," *Biochemistry*, vol. 55, no. 9, pp. 1279–1282, 2016.
- [39] H. Paulus, "Protein splicing and related forms of protein autoprocessing," *Annu. Rev. Biochem.*, no. 69, pp. 447–496, 2000.
- [40] P. L. Starokadomskyy, "Protein splicing," *Mol. Biol.*, vol. 41, no. 2, pp. 278–293, 2007.
- [41] B. W. Poland, M. Q. Xu, and F. A. Quijochó, "Structural insights into the protein splicing mechanism of PI-Scel," *J. Biol. Chem.*, vol. 275, no. 22, pp. 16408–16413, 2000.
- [42] Y. Anraku, R. Mizutani, and Y. Satow, "Protein splicing: Its discovery and structural insight into novel chemical mechanisms," *IUBMB Life*, vol. 57, no. 8, pp. 563–574, 2005.
- [43] Y. Ding, M. Xu, I. Ghosh, X. Chen, S. Ferrandon, G. Lesage, and Z. Rao, "Crystal structure of a mini-intein reveals a conserved catalytic module involved in side chain cyclization of asparagine during protein splicing," *J. Biol. Chem.*, vol. 278, no. 40, pp. 39133–39142, 2003.
- [44] T. Klabunde, S. Sharma, A. Telenti, W. R. Jacobs, and J. C. Sacchettini, "Crystal structure of GyrA intein from *Mycobacterium xenopi* reveals structural basis of protein splicing," *Nat. Struct. Biol.*, vol. 5, no. 1, pp. 31–36, 1998.
- [45] A. Romanelli, A. Shekhtman, D. Cowburn, and T. W. Muir, "Semisynthesis of a segmental isotopically labeled protein splicing precursor: NMR evidence for an unusual peptide bond at the N-extein-intein junction," *PNAS*, vol. 101, no. 17, pp. 6397–6402, 2004.
- [46] M. A. Johnson, M. W. Southworth, F. B. Perler, and K. Wüthrich, "NMR assignment of a Klba intein precursor from *Methanococcus jannaschii*," *Biomol. NMR Assign.*, vol. 1, no. 1, pp. 19–21, 2007.
- [47] P. Sun, S. Ye, S. Ferrandon, T. C. Evans, M.-Q. Xu, and Z. Rao, "Crystal Structures of an Intein from the Split dnaE Gene of *Synechocystis* sp. PCC6803 Reveal the Catalytic Model Without the Penultimate Histidine and the Mechanism of Zinc Ion Inhibition of Protein Splicing," *J. Mol. Biol.*, vol. 353, pp. 1093–1105, 2005.

- [48] E. Eryilmaz, N. H. Shah, T. W. Muir, and D. Cowburn, "Structural and dynamical features of inteins and implications on protein splicing," *J. Biol. Chem.*, vol. 289, no. 21, pp. 14506–14511, 2014.
- [49] G. Volkmann and Mootz, "Recent progress in intein research: from mechanism to directed evolution and applications," *Cell. Mol. Life Sci.*, vol. 70, no. 7, pp. 1185–1206, 2013.
- [50] S. Frutos, M. Goger, B. Giovani, D. Cowburn, and T. W. Muir, "Branched intermediate formation stimulates peptide bond cleavage in protein splicing," *Nat. Chem. Biol.*, vol. 6, no. 7, pp. 527–533, 2010.
- [51] B. Pereira, P. T. Shemella, G. Amitai, G. Belfort, S. K. Nayak, and M. Belfort, "Spontaneous proton transfer to a conserved intein residue determines on-pathway protein splicing," *J. Mol. Biol.*, vol. 406, no. 3, pp. 430–442, 2011.
- [52] C. Z. Du, Y. Zheng, M. Patterson, Y. Liu and C. Wang, "pKa Coupling at the Intein Active Site: Implications for the Coordination Mechanism of Protein Splicing with a Conserved Aspartate," *J Am Chem Soc.*, vol. 133, no. 26, pp. 10275–10282, 2011.
- [53] Z. Liu, S. Frutos, M. J. Bick, M. Vila-Perelló, G. T. Debelouchina, S. A. Darst, and T. W. Muir, "Structure of the branched intermediate in protein splicing," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 111, no. 23, pp. 8422–8427, 2014.
- [54] L. Chen, J. Benner, and F. B. Perler, "Protein splicing in the absence of an intein penultimate histidine," *J. Biol. Chem.*, vol. 275, no. 27, pp. 20431–20435, 2000.
- [55] S. Ellilä, J. M. Jurvansuu, and H. Iwai, "Evaluation and comparison of protein splicing by exogenous inteins with foreign exteins in *Escherichia coli*," *FEBS Lett.*, vol. 585, no. 21, pp. 3471–3477, 2011.
- [56] N. M. Nichols and T. C. Evans, "Mutational analysis of protein splicing, cleavage, and self-association reactions mediated by the naturally split *Ssp DnaE* intein," *Biochemistry*, vol. 43, no. 31, pp. 10265–10276, 2004.
- [57] N. H. Shah, E. Eryilmaz, D. Cowburn, and T. W. Muir, "Naturally split inteins assemble through a 'capture and collapse' mechanism," *J. Am. Chem. Soc.*, vol. 135, no. 49, pp. 18673–18681, 2013.
- [58] P. T. Shemella, N. I. Topilina, I. Soga, B. Pereira, G. Belfort, M. Belfort, S. K. Nayak, "Electronic structure of neighboring extein residue modulates intein C-terminal cleavage activity," *Biophys. J.*, vol. 100, no. 9, pp. 2217–2225, 2011.
- [59] E. J. Pearl, J. D. A. Tyndall, R. T. M. Poulter, and S. M. Wilbanks, "Sequence requirements for splicing by the *Cne PRP8* intein," *FEBS Lett.*, vol. 581, no. 16, pp. 3000–3004, 2007.

- [60] S. Chong, Y. Shao, H. Paulus, J. Benner, F. B. Perler, and M. Q. Xu, "Protein splicing involving the *Saccharomyces cerevisiae* VMA intein: The steps in the splicing pathway, side reactions leading to protein cleavage, and establishment of an in vitro splicing system," *J. Biol. Chem.*, vol. 271, no. 36, pp. 22159–22168, 1996.
- [61] M. W. Southworth, J. Benner, and F. B. Perler, "An alternative protein splicing mechanism for inteins lacking an N-terminal nucleophile," *EMBO J.*, vol. 19, no. 18, pp. 5019–5026, 2000.
- [62] K. Tori, B. Dassa, M. A. Johnson, M. W. Southworth, L. E. Brace, Y. Ishino, S. Pietrokovski, and F. B. Perler, "Splicing of the mycobacteriophage bethlehem DnaB intein: Identification of a new mechanistic class of inteins that contain an obligate block of nucleophile," *J. Biol. Chem.*, vol. 285, no. 4, pp. 2515–2526, 2010.
- [63] X. Qi, J. Wang, Q. Meng, and X. Q. Liu, "Alternative nucleophilic residues in intein catalysis of protein splicing," *Protein Pept. Lett.*, vol. 18, no. 12, pp. 1226–32, Dec. 2011.
- [64] M. Cheriyan, S.-H. Chan, and F. Perler, "Traceless splicing enabled by substrate-induced activation of the *Nostoc punctiforme* Npu DnaE intein after mutation of a catalytic cysteine to serine," *J. Mol. Biol.*, vol. 426, no. 24, pp. 4018–29, Dec. 2014.
- [65] O. D. Ekici, M. Paetzel, and R. E. Dalbey, "Unconventional serine proteases: Variations on the catalytic Ser/His/Asp triad configuration," *Protein Sci.*, vol. 17, no. 12, pp. 2023–2037, 2008.
- [66] B. Dassa, N. London, B. L. Stoddard, O. Schueler-Furman, and S. Pietrokovski, "Fractured genes: A novel genomic arrangement involving new split inteins and a new homing endonuclease family," *Nucleic Acids Res.*, vol. 37, no. 8, pp. 2560–2573, 2009.
- [67] J. Caspi, G. Amitai, O. Belenkiy, and S. Pietrokovski, "Distribution of split DnaE inteins in cyanobacteria," *Mol. Microbiol.*, vol. 50, no. 5, pp. 1569–1577, 2003.
- [68] J. J. Choi, K. H. Nam, B. Min, S.-J. Kim, D. Söll, and S. T. Kwon, "Protein Trans-splicing and Characterization of a Split Family B-type DNA Polymerase from the Hyperthermophilic Archaeal Parasite *Nanoarchaeum equitans*," *J. Mol. Biol.*, vol. 356, no. 5, pp. 1093–1106, 2006.
- [69] B. P. Callahan, N. I. Topilina, M. J. Stanger, P. Van Roey, and M. Belfort, "Structure of catalytically competent intein caught in a redox trap with functional and evolutionary implications," *Nat. Struct. Mol. Biol.*, vol. 18, no. 5, pp. 630–633, 2011.
- [70] A. K. Dearden, B. Callahan, P. V. Roey, Z. Li, U. Kumar, M. Belfort, and S. K. Nayak, "A conserved threonine spring-loads precursor for intein splicing," *Protein Sci.*, vol. 22, no. 5, pp. 557–63, 2013.

- [71] A. S. Aranko, J. S. Oeemig, T. Kajander, and H. Iwai, "Intermolecular domain swapping induces intein-mediated protein alternative splicing," *Nat. Chem. Biol.*, vol. 9, no.10, pp. 616-622, 2013.
- [72] J. S. Oeemig, A. S. Aranko, J. Djupsjöbacka, K. Heinämäki, and H. Iwai, "Solution structure of DnaE intein from *Nostoc punctiforme*: Structural basis for the design of a new split intein suitable for site-specific chemical modification," *FEBS Lett.*, vol. 583, no. 9, pp. 1451-1456, 2009.
- [73] Q. Wu, Z. Gao, Y. Wei, G. Ma, Y. Zheng, Y. Dong, and Y. Liu, "Conserved residues that modulate protein trans-splicing of Npu DnaE split intein," *Biochem. J.*, vol. 461, no. 2, pp. 247-255, 2014.
- [74] B. Dassa, G. Amitai, J. Caspi, O. Schueler-Furman, and S. Pietrokovski, "Trans protein splicing of cyanobacterial split inteins in endogenous and exogenous combinations," *Biochemistry*, vol. 46, no. 1, pp. 322-330, 2007.
- [75] N. H. Shah, M. Vila-Perelló, and T. W. Muir, "Kinetic control of one-pot trans-splicing reactions by using a wild-type and designed split intein," *Angew. Chemie - Int. Ed.*, vol. 50, no. 29, pp. 6511-6515, 2011.
- [76] J. Shi and T. W. Muir, "Development of a tandem protein trans-splicing system based on native and engineered split inteins," *J. Am. Chem. Soc.*, vol. 127, no. 17, pp. 6198-6206, 2005.
- [77] Y. Zheng, Q. Wu, C. Wang, M.-Q. Xu, and Y. Liu, "Mutual synergistic protein folding in split intein," *Biosci. Rep.*, vol. 32, no. 5, pp. 433-42, 2012.
- [78] H. Yu, "Extending the size limit of protein nuclear magnetic resonance," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 96, no. 2, pp. 332-334, 1999.
- [79] T. Yamazaki, T. Otomo, N. Oda, Y. Kyogoku, K. Uegaki, N. Ito, Y. Ishino, and H. Nakamura, "Segmental isotope labeling for protein NMR using peptide splicing," *J. Am. Chem. Soc.*, vol. 120, no. 22, pp. 5591-5592, 1998.
- [80] R. Xu, B. Ayers, D. Cowburn, and T. W. Muir, "Chemical ligation of folded recombinant proteins: Segmental isotopic labeling of domains for NMR studies," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 96, no. 2, pp. 388-393, 1999.
- [81] T. Otomo, N. Ito, Y. Kyogoku, and T. Yamazaki, "NMR observation of selected segments in a larger protein: Central- segment isotope labeling through intein-mediated ligation," *Biochemistry*, vol. 38, no. 49, pp. 16040-16044, 1999.

- [82] S. Züger and H. Iwai, "Intein-based biosynthetic incorporation of unlabeled protein tags into isotopically labeled proteins for NMR studies," *Nat. Biotechnol.*, vol. 23, no. 6, pp. 736–740, 2005.
- [83] T. W. Muir, D. Sondhi, and P. A. Cole, "Expressed protein ligation: A general method for protein engineering," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 95, no. 12, pp. 6705-6710, 1998.
- [84] T. C. Evans, J. Benner, and M. Q. Xu, "Semisynthesis of cytotoxic proteins using a modified protein splicing element," *Protein Sci.*, vol. 95, no. 12, pp. 6705-6710, 1998.
- [84b] M. Vila-Perelló, and T. W. Muir, "Biological applications of protein splicing," *Cell*, vol.143, no. 2, pp. 191-200, 2010.
- [85] C. P. Scott, E. Abel-Santos, M. Wall, D. C. Wahnou, and S. J. Benkovic, "Production of cyclic peptides and proteins in vivo," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 96, no. 24, pp. 13638-13643, 1999.
- [86] J. A. Kritzer, S. Hamamichi, J. M. McCaffery, S. Santagata, T. A. Naumann, K. A. Caldwell, G. A. Caldwell, and S. Lindquist, "Rapid selection of cyclic peptides that reduce α -synuclein toxicity in yeast and animal models," *Nat. Chem. Biol.*, vol. 5, no. 9, pp. 655-663, 2009.
- [87] L. Berrade, Y. Kwon, and J. A. Camarero, "Photomodulation of protein trans-splicing through backbone photocaging of the DnaE split intein," *ChemBioChem*, vol. 11, no. 10, pp. 1368-1372, 2010.
- [88] A. R. Buskirk, Y. C. Ong, Z. J. Gartner, and D. R. Liu, "Directed evolution of ligand dependence: Small-molecule-activated protein splicing," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 29, pp. 10505-10510, 2004.
- [89] H. Jeon, E. Lee, D. Kim, M. Lee, J. Ryu, C. Kang, S. Kim, and Y Kwon, "Cell-Based Biosensors Based on Intein-Mediated Protein Engineering for Detection of Biologically Active Signaling Molecules," *Anal. Chem.*, vol. 90, no. 16, pp. 9779-9786, 2018.
- [90] L. J. Rothschild and R. L. Mancinelli, "Life in extreme environments," *Nature*, vol. 409, no. 6823, pp. 1092-1101, 2001.
- [91] C. Vieille, D. S. Burdette, and J. G. Zeikus, "Thermozymes," *Biotechnol. Annu. Rev.*, vol. 2, pp. 1-83, 1996.
- [92] K. O. Stetter, "Hyperthermophiles in the history of life," *Philos. Trans. R. Soc. Lon. B. Biol. Sci.*, vol. 361, no. 1474, pp. 1837-1842, 2006.
- [93] D. Stiefler-Jensen, "Stability of hyperthermophilic proteins-Post-translational modifications and other factors," PhD thesis, Danish Archaea Centre, 2016.

- [94] S. Fukuchi and K. Nishikawa, "Protein surface amino acid compositions distinctively differ between thermophilic and mesophilic bacteria," *J. Mol. Biol.*, vol. 309, no. 4, pp. 835–843, 2001.
- [95] C. Cambillau and J. M. Claverie, "Structural and genomic correlates of hyperthermostability," *J. Biol. Chem.*, vol. 275, no. 42, pp. 32383–32386, 2000.
- [96] A. Szilágyi and P. Závodszy, "Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: Results of a comprehensive survey," *Structure*, vol. 8, no. 5, pp. 493–504, 2000.
- [97] S. Hait, S. Mallik, S. Basu, and S. Kundu, "Finding the generalized molecular principles of protein thermal stability," *Proteins Struct. Funct. Bioinforma.*, vol. 88, no. 6, pp. 788–808, 2020.
- [98] G. Sælensminde, Ø. Halskau, and I. Jonassen, "Amino acid contacts in proteins adapted to different temperatures: Hydrophobic interactions and surface charges play a key role," *Extremophiles*, vol. 13, no. 1, pp. 11–20, 2009.
- [99] M. Robinson-Rechavi, A. Alibés, and A. Godzik, "Contribution of electrostatic interactions, compactness and quaternary structure to protein thermostability: Lessons from structural genomics of *Thermotoga maritima*," *J. Mol. Biol.*, vol. 356, no. 2, pp. 547–557, 2006.
- [100] Y. Li, C. R. Middaugh, and J. Fang, "A novel scoring function for discriminating hyperthermophilic and mesophilic proteins with application to predicting relative thermostability of protein mutants," *BMC Bioinformatics*, vol. 62, pp. 1–11, 2010.
- [101] V. V. Loladze, B. Ibarra-Molero, J. M. Sanchez-Ruiz, and G. I. Makhatadze, "Engineering a thermostable protein via optimization of charge-charge interactions on the protein surface," *Biochemistry*, vol. 38, no. 50, pp. 16419–16423, 1999.
- [102] F. Sterpone and S. Melchionna, "Thermophilic proteins: Insight and perspective from in silico experiments," *Chem. Soc. Rev.*, vol. 41, no. 5, pp. 1665–1676, 2012.
- [103] F. Sterpone, C. Bertonati, G. Briganti, and S. Melchionna, "Key role of proximal water in regulating thermostable proteins," *J. Phys. Chem. B*, vol. 113, no. 1, pp. 131–137, 2009.
- [104] F. Sterpone, C. Bertonati, G. Briganti, and S. Melchionna, "Water around thermophilic proteins: The role of charged and apolar atoms," *J. Phys. Condens. Matter*, vol. 22, no. 28, p. 284113, 2010.
- [105] H. R. Bosshard, D. N. Marti, and I. Jelesarov, "Protein stabilization by salt bridges: Concepts, experimental approaches and clarification of some misunderstandings," *J. Mol. Recognit.*, vol. 17, no. 1, pp. 1–16, 2004.

- [106] A. H. Elcock, "The stability of salt bridges at high temperatures: Implications for hyperthermophilic proteins," *J. Mol. Biol.*, vol. 284, no. 2, pp. 489-502, 1998.
- [107] M. M. Gromiha, M. C. Pathak, K. Saraboji, E. A. Ortlund, and E. A. Gaucher, "Hydrophobic environment is a key factor for the stability of thermophilic proteins," *Proteins Struct. Funct. Bioinforma.*, vol. 81, no. 4, pp. 715–721, 2013.
- [108] E. van Dijk, A. Hoogeveen, and S. Abeln, "The Hydrophobic Temperature Dependence of Amino Acids Directly Calculated from Protein Structures," *PLoS Comput. Biol.*, vol. 11, no. 5, pp. 1-17, 2015.
- [109] D. R. Tompa, M. M. Gromiha, and K. Saraboji, "Contribution of main chain and side chain atoms and their locations to the stability of thermophilic proteins," *J. Mol. Graph. Model.*, vol. 64, pp. 85-93, 2016.
- [110] D. W. Sammond, N. Kastelowitz, M. E. Himmel, H. Yin, M. F. Crowley, and Y. J. Bomble, "Comparing residue clusters from thermophilic and mesophilic enzymes reveals adaptive mechanisms," *PLoS One*, vol. 11, no. 1, pp. 1-18, 2016.
- [111] R. Goldstein, "Thermophiles , and hyperthermophiles : Insights from the quasi-chemical approximation," *Protein Sci.*, vol. 16, no. 9, pp. 1887-1895, 2007.
- [112] J. Chen, Z. Lu, J. Sakon, and W. E. Stites, "Increasing the thermostability of staphylococcal nuclease: Implications for the origin of protein thermostability," *J. Mol. Biol.*, vol. 303, no. 2, pp. 125–130, 2000.
- [113] K. M. Oda K, "Physicochemical origin of high correlation between thermal stability of a protein and its packing efficiency: a theoretical study for staphylococcal nuclease mutants.," *Biophys Physicobiol*, vol. 12, pp. 1–12, 2015.
- [114] K. M. Tych *et al.*, "Differential Effects of Hydrophobic Core Packing Residues for Thermodynamic and Mechanical Stability of a Hyperthermophilic Protein," *Langmuir*, vol. 32, no. 29, pp. 7392-7402, 2016.
- [115] R. Scandurra, V. Consalvi, R. Chiaraluce, L. Politi, and P. C. Engel, "Protein thermostability in extremophiles," *Biochimie*, vol. 80, no. 11, pp. 933-941, 1998.
- [116] P. Závodszy, J. Kardos, Á. Svingor, and G. A. Petsko, "Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 95, no. 13, pp. 7406-7411, 1998.
- [117] R. Jaenicke and G. Böhm, "The stability of proteins in extreme environments," *Curr. Opin. Struct. Biol.*, vol. 8, no. 6, pp. 738-748, 1998.

- [118] H. Bönisch, J. Backmann, T. Kath, D. Naumann, and G. Schäfer, "Adenylate kinase from *Sulfolobus acidocaldarius*: Expression in *Escherichia coli* and characterization by Fourier transform infrared spectroscopy," *Arch. Biochem. Biophys.*, vol. 333, no. 1, pp. 75-84, 1996.
- [119] A. Gershenson, J. A. Schauerte, L. Giver, and F. H. Arnold, "Tryptophan phosphorescence study of enzyme flexibility and unfolding in laboratory-evolved thermostable esterases," *Biochemistry*, vol. 39, no. 16, pp. 4658-4665, 2000.
- [120] C. Vieille and G. J. Zeikus, "Hyperthermophilic Enzymes: Sources, Uses, and Molecular Mechanisms for Thermostability," *Microbiol. Mol. Biol. Rev.*, vol. 65, no. 1, pp. 1-43, 2001.
- [121] T. J. Taylor, and I. I. Vaisman, "Discrimination of thermophilic and mesophilic proteins," *BMC Struct. Biol.*, vol. 10, no. 1, 2010.
- [122] B. Folch, Y. Dehouck, and M. Rooman, "Thermo- and mesostabilizing protein interactions identified by temperature-dependent statistical potentials," *Biophys. J.*, vol. 98, no. 4, pp. 667-677, 2010.
- [123] K. M. Makwana and R. Mahalakshmi, "Implications of aromatic-aromatic interactions: From protein structures to peptide models," *Protein Sci.*, vol. 24, no. 12, pp. 1920-1933, 2015.
- [124] R. S. Prajapati, M. Sirajuddin, V. Durani, S. Sreeramulu, and R. Varadarajan, "Contribution of cation- π interactions to protein stability," *Biochemistry*, vol. 45, no. 50, pp. 15000-15010, 2006.
- [125] M. Matsumura, G. Signor, and B. W. Matthews, "Substantial increase of protein stability by multiple disulphide bonds," *Nature*, vol. 342, no. 6247, pp. 291-293, 1989.
- [126] G. Bertani, "Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*," *J. Bacteriol.*, vol. 62, no. 3, pp. 293-300, 1951.
- [127] F. Sanger, S. Nicklen, and A. R. Coulson, "DNA sequencing with chain-terminating inhibitors," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 74, no. 12, pp. 5463-5467, 1977.
- [128] H. D. Mootz and T. W. Muir, "Protein splicing triggered by a small molecule," *J. Am. Chem. Soc.*, vol. 124, no. 31, pp. 9044-9045, 2002.
- [129] M. R. Wilkins, E. Gasteiger, A. Bairoch, J. C. Sanchez, K. L. Williams, R. D. Appel, and D. F. Hochstrasser, "Protein Identification and Analysis Tools on the ExpASY Server," *Methods Mol. Biol.*, vol. 112, pp. 531-552, 2005.
- [130] C. Notredame, D. G. Higgins, and J. Heringa, "T-coffee: A novel method for fast and accurate multiple sequence alignment," *J. Mol. Biol.*, vol. 302, no. 1, pp. 205-217, 2000.

- [131] V. Vacic, V. N. Uversky, A. K. Dunker, and S. Lonardi, "Composition Profiler: A tool for discovery and visualization of amino acid composition differences," *BMC Bioinformatics*, vol 8, p. 211, 2007.
- [132] M. Sickmeier, J. A. Hamilton, T. LeGall, V. Vacic, M. S. Cortese, A. Tantos, B. Szabo, P. Tompa, J. Chen, V. N. Uversky, Z. Obradovic, and A. K. Dunker., "DisProt: The database of disordered proteins," *Nucleic Acids Res.*, vol. 35, pp. 86-93, 2007.
- [133] P. Romero, Z. Obradovic, X. Li, E. C. Garner, C. J. Brown, and A. K. Dunker, "Sequence complexity of disordered protein," *Proteins*, vol. 42, no. 1, pp. 38-48, 2001.
- [134] V. N. Uversky, J. R. Gillespie, and A. L. Fink, "Why Are 'Natively Unfolded' Proteins Unstructured Under Physiologic Conditions?," *Proteins*, vol. 41, no. 3, pp. 415-427, 2000.
- [135] U. K. Laemmli, "Cleavage of structural proteins during the assembly of the head of bacteriophage T4," *Nature*, vol. 227, no. 5259, pp. 680-685, 1970.
- [136] W. Ebeling, N. Hennrich, M. Klockow, H. Metz, H. D. Orth, and H. Lang, "Proteinase K from *Tritirachium album Limber*," *Eur. J. Biochem.*, vol. 47, no. 1, pp. 91-97, 1974.
- [137] K. Morihara and H. Tsuzuki, "Specificity of Proteinase K from *Tritirachium album Limber* for Synthetic Peptides," *Agr. Biol. Chem.*, vol 39, no. 7, pp. 1489-1492, 1975.
- [138] M. Manning and W. Colón, "Structural Basis of Protein Kinetic Stability: Resistance to Sodium Dodecyl Sulfate Suggests a Central Role for Rigidity and a Bias Toward β -Sheet Structure [†]," *Biochemistry*, vol. 43, no. 35, pp. 11248–11254, 2004.
- [139] J. Pous, A. Canals, S. S. Terzyan, A. Guasch, A. Benito, M. Ribó, M. Vilanova, and M. Coll, "Three-dimensional structure of a human pancreatic ribonuclease variant, a step forward in the design of cytotoxic ribonucleases," *J. Mol. Biol.*, vol. 303, no. 1, pp. 49-60, 2000.
- [140] C. Park and S. Marqusee, "Pulse proteolysis: A simple method for quantitative determination of protein stability and ligand binding," *Nat. Methods*, vol. 2, no. 3, pp. 207-212, 2005.
- [141] J. Font, A. Benito, R. Lange, M. Ribó, and M. Vilanova, "The contribution of the residues from the main hydrophobic core of ribonuclease A to its pressure-folding transition state," *Protein Sci.*, vol. 15, no. 5, pp. 1000-1009, 2006.
- [142] S. Ruiz Martínez, "Structural and functional characterization of Apoptin and design of a monomeric variant with antitumor activity," PhD thesis, Universitat de Girona, 2016.
- [143] S. Ruiz-Martínez, D. Pantoja-Uceda, J. Castro, M. Vilanova, M. Ribó, M. Bruix, A. Benito, and D. V. Laurents., "Insights into the mechanism of Apoptin's exquisitely selective anti-tumor action from atomic level characterization of its conformation and dynamics," *Arch. Biochem. Biophys.*, vol. 614, pp. 53-64, 2017.

- [144] S. Ruiz-Martínez, J. Castro, M. Vilanova, M. Bruix, D. V. Laurents, M. Ribó, and A. Benito, "A truncated apoptin protein variant selectively kills cancer cells," *Invest. New Drugs*, vol. 35, no. 3, pp. 260-268, 2017.
- [145] W. Kabsch and C. Sander, "Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features," *Biopolymers*, vol. 22, no. 12, pp. 2577-2637, 1983.
- [146] M. Sammito, C. Millán, D. D. Rodríguez, I. M. de Ilarduya, K. Meindl, I. De Marino, G. Petrillo, R. M. Buey, J. M. de Pereda, K. Zeth, G. M. Sheldrick, and I. Usón, "Exploiting tertiary structure through local folds for crystallographic phasing," *Nat. Methods*, vol. 10, no. 11, pp. 1099-1101, 2013.
- [147] R. J. Read and A. J. McCoy, "A log-likelihood-gain intensity target for crystallographic phasing that accounts for experimental error," *Acta Crystallogr. Sect. D Struct. Biol., Crystallogr.*, vol. 72, no. 3, pp. 375-387, 2016.
- [148] I. Usón and G. M. Sheldrick, "An introduction to experimental phasing of macromolecules illustrated by SHELX; New autotracing features," *Acta Crystallogr. Sect. D. Struct. Biol., Crystallogr.*, vol. 74, no. 2, pp. 106-116, 2018.
- [149] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read, "Phaser crystallographic software," *J. Appl. Crystallogr.*, vol. 40, no. 4, pp. 658-674, 2007.
- [150] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan, "Features and development of Coot," *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 66, no. 4, pp. 486-501, 2010.
- [151] G. N. Murshudov, A. A. Vagin, and E. J. Dodson, "Refinement of macromolecular structures by the maximum-likelihood method," *Acta Crystallog. Sect. D. Biol. Crystallogr.*, vol. 53, no. 3, pp. 240-255, 1997.
- [152] V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, and D. C. Richardson, "MolProbity: All-atom structure validation for macromolecular crystallography," *Acta Crystallogr. D. Biol. Crystallogr.*, vol. 66, no. 1, pp. 12-21, 2010.
- [153] V. Novokhatny and K. Ingham, "Thermodynamics of maltose binding protein unfolding," *Protein Sci.*, vol. 6, no. 1, pp. 141-146, 1997.
- [154] R. B. Kapust and D. S. Waugh, "Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused.," *Protein Sci.*, vol. 8, no. 8, pp. 1668-74, 1999.

- [155] P. Alexander, S. Fahnstock, T. Lee, J. Orban, and P. Bryan, "Thermodynamic analysis of the folding of the streptococcal protein G IgG-binding domains B1 and B2: why small proteins tend to have high denaturation temperatures.," *Biochemistry*, vol. 31, no. 14, pp. 3597–603, 1992.
- [156] B. Nilsson, T. Moks, B. Jansson, L. Abrahmsén, A. Elmlad, E. Holmgren, C. Henrichson, T. A. Jones, and M. Uhlén, "A synthetic IgG-binding domain based on staphylococcal protein A," *Protein Eng. Des. Sel.*, vol. 1, no. 2, pp. 107-113, 1987.
- [157] J. Feldwisch, V. Tolmachev, C. Lendel, N. Herne, A. Sjöberg, B. Larsson, D. Rosik, E. Lindqvist, G. Fant, I. Höidén-Guthenberg, J. Galli, P. Jonasson, and L. Abrahmsén., "Design of an Optimized Scaffold for Affibody Molecules," *J. Mol. Biol.*, vol. 398, no. 2, pp. 232–247, 2010.
- [158] M. Q. Xu, M. W. Southworth, F. B. Mersha, L. J. Hornstra, and F. B. Perler, "In vitro protein splicing of purified precursor and the identification of a branched intermediate," *Cell*, vol. 75, no. 7, pp. 1371–1377, 1993.
- [159] D. D. Martin, M.-Q. Xu, and T. C. Evans, "Characterization of a Naturally Occurring Trans-Splicing Intein from *Synechocystis* sp. PCC6803," *Biochemistry*, vol. 40, no. 5, pp. 1393–1402, 2001.
- [160] A. S. Aranko, S. Züger, E. Buchinger, and H. Iwai, "In vivo and in vitro protein ligation by naturally occurring and engineered split DnaE inteins," *PLoS One*, vol. 4, no. 4, 2009.
- [161] K. Iwai and T. Ando, "[²⁹N] → O acyl rearrangement," *Methods Enzymol.*, vol. 11, no. 1967, pp. 263-282, 1967.
- [162] "PROTOCOLS & APPLICATIONS GUIDE-Buffers for Biochemical Reactions.," www.Promega.com
- [163] L. Greenough, J. F. Menin, N. S. Desai, Z. Kelman, and A. F. Gardner, "Characterization of Family D DNA polymerase from *Thermococcus* sp. 9°N," *Extremophiles*, vol. 18, no. 4, pp. 653-664, 2014.
- [164] A. Ghasemi, A. H. Salmanian, N. Sadeghifard, A. A. Salarian, and M. K. Gholi, "Cloning, expression and purification of pwo polymerase from *pyrococcus woesei*," *Iran. J. Microbiol.*, vol. 3, no. 3, pp. 118–122, 2011.
- [165] F-X. Theillet, L. Kalmar, P. Tompa, K. Han, P. Selenko, A. K. Dunker, G. W. Daughdrill, and V. N. Uversky, "The alphabet of intrinsic disorder: I. Act like a Pro: On the abundance and roles of proline residues in intrinsically disordered proteins," *Intrinsically Disord. Proteins*, vol. 1, no. 1, p. e24360, 2013.

- [166] J. Winkelmann, G. Calloni, S. Campioni, B. Mannini, N. Taddei, and F. Chiti, "Low-Level Expression of a Folding-Incompetent Protein in *Escherichia coli*: Search for the Molecular Determinants of Protein Aggregation In Vivo," *J. Mol. Biol.*, vol. 398, no. 4, pp. 600-613, 2010.
- [167] A. Espargaró, V. Castillo, N. S. de Groot, and S. Ventura, "The in Vivo and in Vitro Aggregation Properties of Globular Proteins Correlate With Their Conformational Stability: The SH3 Case," *J. Mol. Biol.*, vol. 378, no. 5, pp. 1116-1131, 2008.
- [168] V. Castillo, A. Espargaró, V. Gordo, J. Vendrell, and S. Ventura, "Deciphering the role of the thermodynamic and kinetic stabilities of SH3 domains on their aggregation inside bacteria," *Proteomics*, vol. 10, no. 23, pp. 4172-4185, 2010.
- [169] S. Kelly and N. Price, "The Use of Circular Dichroism in the Investigation of Protein Structure and Function," *Curr. Protein Pept. Sci.*, vol. 1, no. 4, pp. 349-384, 2005.
- [170] J. J. Merelo, M. A. Andrade, A. Prieto, and F. Morán, "Proteinotopic feature maps," *Neurocomputing*, vol. 6, no. 4, pp. 383-390, 1994.
- [171] M. Yahiya Khan, G. Villanueva, and S. A. Newman, "On the origin of the positive band in the far-ultraviolet circular dichroic spectrum of fibronectin," *J. Biol. Chem.*, vol. 264, no. 4, pp. 2139-2142, 1989.
- [172] G. V. Semisotnov, N. A. Rodionova, O. I. Razgulyaev, V. N. Uversky, A. F. Gripas, and R. I. Gilmanshin, "Study of the 'molten globule' intermediate state in protein folding by a hydrophobic fluorescent probe," *Biopolymers*, vol. 31, no. 1, pp. 119-128, 1991.
- [173] M. Cardamone and N. K. Puri, "Spectrofluorimetric assessment of the surface hydrophobicity of proteins," *Biochem. J.*, vol. 282, no. 2, pp. 589-593, 1992.
- [174] A. Stivala, M. Wybrow, A. Wirth, J. C. Whisstock, and P. J. Stuckey, "Automatic generation of protein structure cartoons with pro-origami," *Bioinformatics*, vol. 27, no. 23, pp. 3315-3316, 2011.
- [175] P. Carvajal-Vallejos, R. Pallissé, H. D. Mootz, and S. R. Schmidt, "Unprecedented rates and efficiencies revealed for new natural split inteins from metagenomic sources.," *J. Biol. Chem.*, vol. 287, no. 34, pp. 28686-96, 2012.
- [176] M. Cheriyan, C. S. Pedomallu, K. Tori, and F. Perler, "Faster protein splicing with the *Nostoc punctiforme* DnaE intein using non-native extein residues.," *J. Biol. Chem.*, vol. 288, no. 9, pp. 6202-11, 2013.
- [177] S. M. Singh and A. K. Panda, "Solubilization and refolding of bacterial inclusion body proteins," *J. Biosci. Bioeng.*, vol. 99, no. 4, pp. 303-310, 2005.

- [178] C. T. Calderone and D. H. Williams, "An enthalpic component in cooperativity: The relationship between enthalpy, entropy, and noncovalent structure in weak associations," *J. Am. Chem. Soc.*, vol. 123, no. 26, pp. 6262–6267, 2001.
- [179] X. L. He, D. C. Chow, M. M. Martick, and K. C. Garcia, "Allosteric activation of a spring-loaded natriuretic peptide receptor dimer by hormone," *Science*, vol. 293, no. 5535, pp. 1657-1662, 2001.
- [180] T. Abraham, R. N. A. H. Lewis, R. S. Hodges, and R. N. McElhaney, "Isothermal titration calorimetry studies of the binding of a rationally designed analogue of the antimicrobial peptide gramicidin S to phospholipid bilayer membranes," *Biochemistry*, vol. 44, no. 6, pp. 2103–2112, 2005.
- [181] K. Luke, D. Apiyo, and P. Wittung-Stafshede, "Dissecting homo-heptamer thermodynamics by isothermal titration calorimetry: Entropy-driven assembly of co-chaperonin protein 10," *Biophys. J.*, vol. 89, no. 5, pp. 3332–3336, 2005.
- [182] C. Tanford, "The hydrophobic effect and the organization of living matter," *Science*, vol. 200, no. 4345, pp. 1012-1018, 1978.
- [183] C. Millán, M. Sammito, and I. Usón, "Macromolecular ab initio phasing enforcing secondary and tertiary structure," *IUCrJ*, vol. 2, no. 1, pp. 95-105, 2015.
- [184] A. G. Murzin, S. E. Brenner, T. Hubbard, and C. Chothia, "SCOP: A structural classification of proteins database for the investigation of sequences and structures," *J. Mol. Biol.*, vol. 247, no. 4, pp. 536-540, 1995.
- [185] E. Krissinel and K. Henrick, "Inference of Macromolecular Assemblies from Crystalline State," *J. Mol. Biol.*, vol. 372, no. 3, pp. 774-797, 2007.
- [186] G. Capitani, J. M. Duarte, K. Baskaran, S. Bliven, and J. C. Somody, "Understanding the fabric of protein crystals: Computational classification of biological interfaces and crystal contacts," *Bioinformatics*, vol. 32, no. 4, pp. 481-489, 2016.
- [187] L. Holm and C. Sander, "Protein structure comparison by alignment of distance matrices," *Journal of Molecular Biology.*, vol. 233, no. 1, pp. 123-138, 1993.
- [188] L. Holm, S. Kääriäinen, P. Rosenström, and A. Schenkel, "Searching protein structure databases with DaliLite v.3," *Bioinformatics*, vol. 24, no. 23, pp. 2780-2781, 2008.
- [189] K. Hiraga, I. Soga, J. T. Dansereau, B. Pereira, V. Derbyshire, Z. Du, C. Wang, P. V. Roey, G. Belfort, and M. Belfort, "Selection and structure of hyperactive inteins: peripheral changes relayed to the catalytic center.," *J. Mol. Biol.*, vol. 393, no. 5, pp. 1106–1117, 2009.

- [190] Q. Wu, Z. Gao, Y. Wei, G. Ma, Y. Zheng, Y. Dong, and Y. Liu, "Conserved residues that modulate protein trans-splicing of Npu DnaE split intein," *Biochem. J.*, vol. 461, no. 2, pp. 247-255, 2014.
- [191] R. Mizutani, S. Nogami, M. Kawasaki, Y. Ohya, Y. Anraku, and Y. Satow, "Protein-splicing reaction via a thiazolidine intermediate: Crystal structure of the VMA1-derived endonuclease bearing the N and C-terminal propeptides," *J. Mol. Biol.*, vol. 316, no. 4, pp. 919-929, 2002.
- [192] Q. Xu, D. Buckley, C. Guan, and H. C. Guo, "Structural insights into the mechanism of intramolecular proteolysis," *Cell*, vol. 98, no. 5, pp. 651-661, 1999.
- [193] H. Jesper S. Oeemig , D. Zhou , T. Kajander , A. Wlodawer and H. Iwai, "NMR and crystalstructures of the *Pyrococcus horikoshii* RadA intein guide a strategy for engineering a highly efficient and promiscuous intein," *J. Mol. Biol.*, vol. 421, no. 1, pp. 85-99, 2012.

7. Annex

Annex A. Purificacions per cromatografia d'afinitat Ni-NTA a pressió atmosfèrica dels precursors N i C i del complex NeqN/NeqC per a posteriors reaccions de *splicing*.

Les cèl·lules obtingudes de la producció del precursor corresponent a 37°C es van resuspendre amb tampó de lisi (20 mM Tampó fosfat a pH 7.4, 300 mM NaCl, 10 mM Imidazol, 1mM PMSF), després de la lisi cel·lular la fracció soluble es va carregar en la resina Ni-NTA prèviament equilibrada amb el tampó de lisi. Després de recollir el través, la columna es va rentar amb 5 VC de solució de rentat 1 (tampó de lisi), 5 VC de tampó de rentat 2 (tampó de lisi + 25 mM) i 3 VC de tampó de rentat 3 (tampó de lisi + 50 mM Imidazol) i finalment, es va passar 5 CV de solució E (tampó de lisi + 250 mM Imidazol). Les fraccions que presentaven la proteïna es van ajuntar i dialitzar enfront del tampó corresponent.

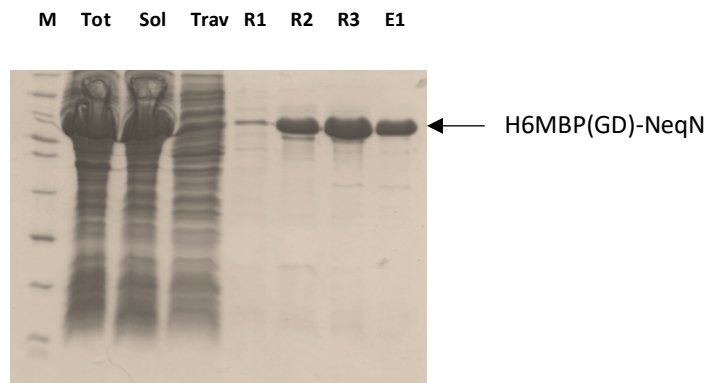


Figura 1. Anàlisi per SDS-PAGE de l'expressió i de la purificació per cromatografia d'afinitat Ni-NTA a pressió atmosfèrica del precursor N: H6MBP(GD)-NeqN (56492.6 Da). Tot: Fracció total; Sol: Fracció soluble; Trav: Través; R1: Rentat 1; R2: Rentat 2; R3: Rentat 3; E1: Elució.

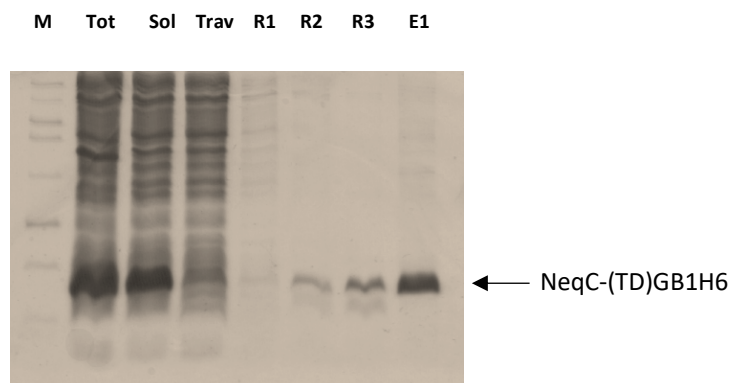


Figura 2. Anàlisi per SDS-PAGE de l'expressió i de la purificació per cromatografia d'afinitat Ni-NTA a pressió atmosfèrica del precursor C: NeqC-(TD)GB1H6 (12146.5 Da). Tot: Fracció total; Sol: Fracció soluble; Trav: Través; R1: Rentat 1; R2: Rentat 2; R3: Rentat 3; E1: Elució.

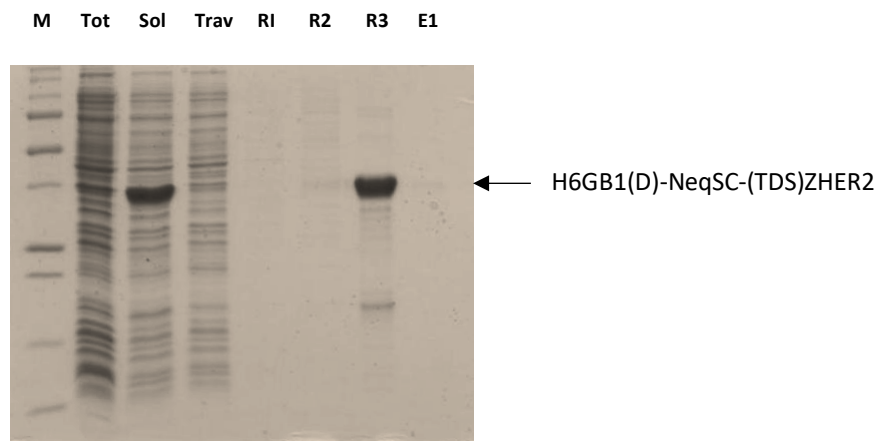


Figura 3. Anàlisi per SDS-PAGE de l'expressió i de la purificació per cromatografia d'afinitat Ni-NTA a pressió atmosfèrica del precursor en cis: H6GB1(D)-NeqSC-(TDS)ZHER2 (31997.7 Da). Tot: Fracció total; Sol: Fracció soluble; Trav: Través; R1: Rentat 1; R2: Rentat 2; R3: Rentat 3; E1: Elució.

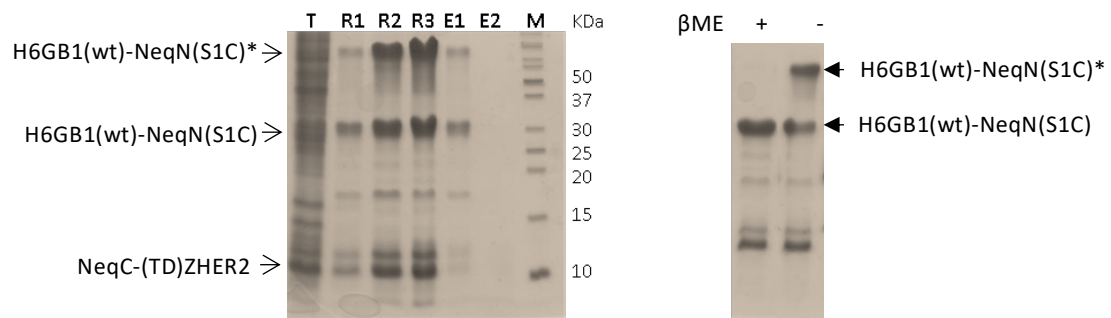


Figura 4. Anàlisi per SDS-PAGE de la copurificació del complex H6GB1(wt)-NeqN(S1C)/NeqC-(TD)ZHER2. (a) Anàlisi de la copurificació per cromatografia d'afinitat Ni-NTA del complex format pels precursors: H6GB1(wt)-NeqN(S1C) i NeqC-(TD)ZHER2. (b) Comprovació de l'origen de les estructures multimèriques formades per molècules de precursor N. (*)Estructures multimèriques; T: Través; R1: rentat 1; R2: rentat 2; R3: rentat 3; E1: elució 1; E2: elució 2; M: Marcadors de massa molecular.

Annex B. Purificació per cromatografia d'afinitat Ni-NTA a pressió atmosfèrica sota condicions desnaturalitzants de la NeqC utilitzat per a la caracterització del procés d'associació.

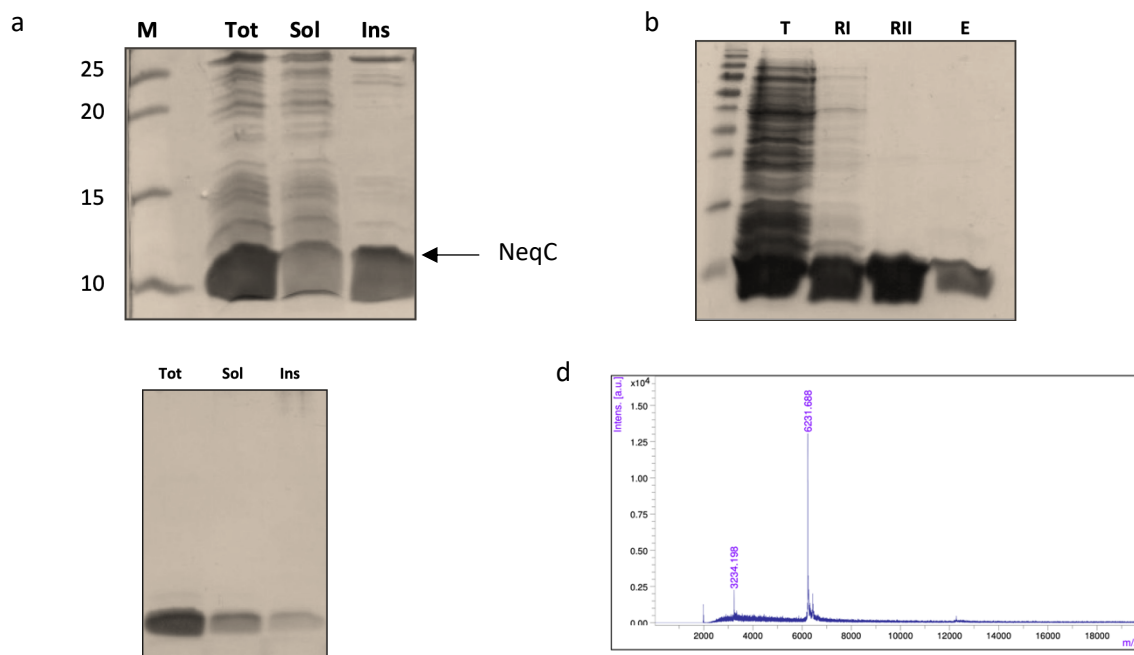


Figura 1. Anàlisi per SDS-PAGE del procés de purificació del precursor C NeqC (6139.8 Da) per cromatografia d'afinitat a pressió atmosfèrica sota condicions desnaturalitzats. (a) Les cèl·lules obtingudes de la producció de 100 ml de cultiu es van resuspendre amb 10 ml de tampó de lisi (20 mM fosfat a pH 7.2, 0.5 M NaCl, 1 mM EDTA). L'extracte cel·lular es va lisar i es va separar la fracció soluble de la insoluble per centrifugació (20 minuts, 14000 rpm a 4°C). M: Marcador de massa molecular; Tot: Fracció total; Sol: Fracció soluble; Ins: Fracció insoluble. (b) Purificació per cromatografia Ni-NTA sota condicions desnaturalitzants. Les cèl·lules obtingudes de la producció a 37°C es va resuspendre amb tampó de lisi (20 mM tampó fosfat a pH 7.2, 0.5 M NaCl, 20 mM Imidazol, 8 M urea), la fracció soluble es va carregar en la resina Ni-NTA prèviament equilibrada amb el tampó de lisi. Després d'eluir el través, la columna es va rentar amb 5 VC de solució RI (tampó de lisi), 5 VC de solució RII (tampó de lisi + 50 mM Imidazol) i finalment, es va passar 5 CV de solució E (tampó de lisi + 250 mM Imidazol). Les fraccions que presentaven la proteïna es van sotmetre a diferents passos de diàlisi per a retirar la urea, finalment es va dialitzar en front H₂O MilliQ. T: Través; RI: rentat 1; RII: rentat 2; E: Elució. (c) Anàlisi de solubilitat del precursor C després de la diàlisi en front H₂O Milli-Q després de la purificació. Tot: Total; Sol: Soluble; Ins: Insoluble. (d) MALDI-TOF.

Annex C. Identificació per Western blot de bandes obtingudes per SDS-PAGE de diferents reaccions de *trans-splicing*.

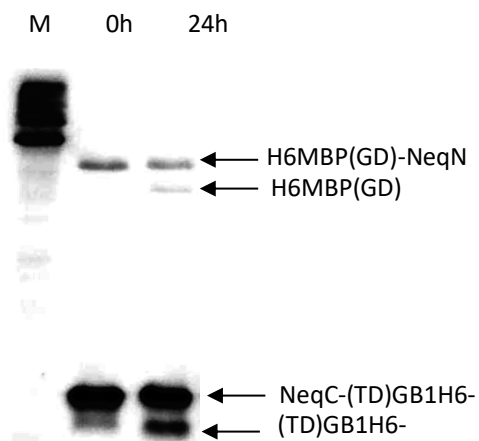


Figura 1. Western blot (Anti-His tag) de la reacció de *trans-splicing* catalitzada pel precursor N, H6MBP(GD)-NeqN, i el precursor C, NeqC-(TD)GB1H6 a 50°C.

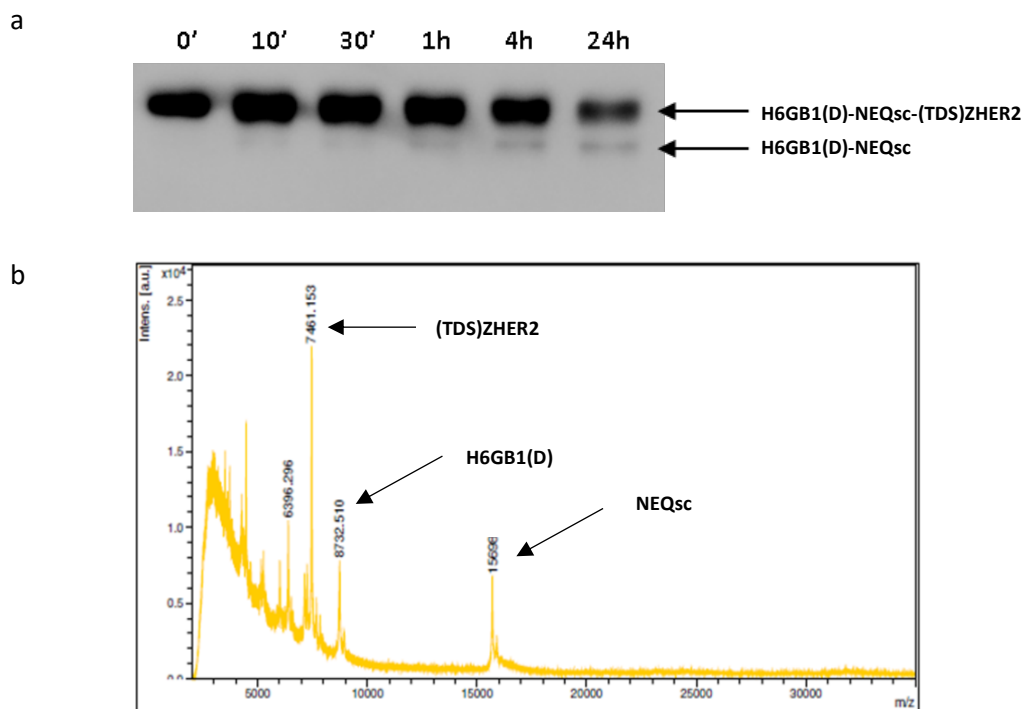


Figura 2. Identificació de les bandes obtingudes per 12%SDS-PAGE de la reacció *splicing* utilitzant el precursor en cis H6GB1(D)-NeqSC-(TDS)ZHER2 a pH 8.0 a 70°C. (a) Western blot (Anti-His tag) (b) MALDI-TOF.

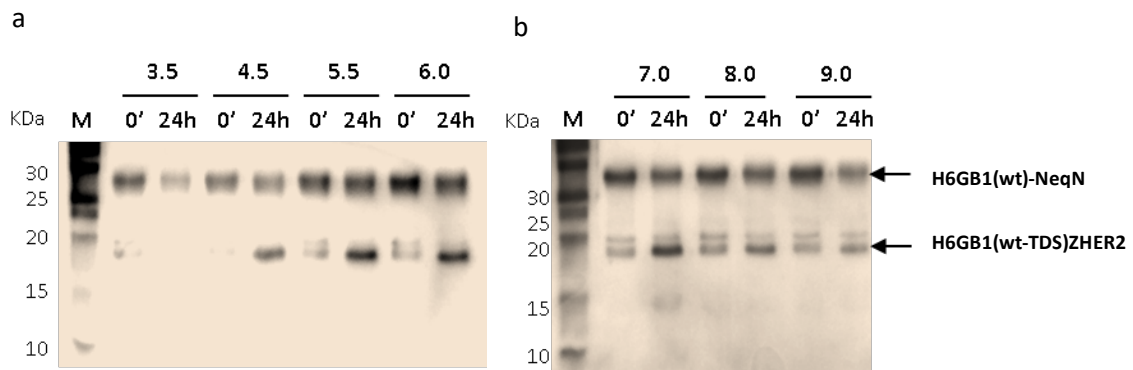


Figura 3. Western-Blot (Anti-His tag) per l'anàlisi de la reacció de *trans-splicing* catalitzada pels precursors: H6GB1(wt)-NeqN i NeqC-(TDS)ZHER2 en un rang de pH entre 3.5-9.0. Els precursors s'han coexpressat i copurificat per cromatografia d'afinitat Ni-NTA. Per cada pH s'ha preparat una solució dels precursors a una concentració final de 20 μ M. a) Rang de pH: 3.5-6.0 (b) Rang de pH: 7.0-10.8

Annex D. Anàlisi d'expressió i solubilització del precursor C, NeqC-(TD)ZHER2

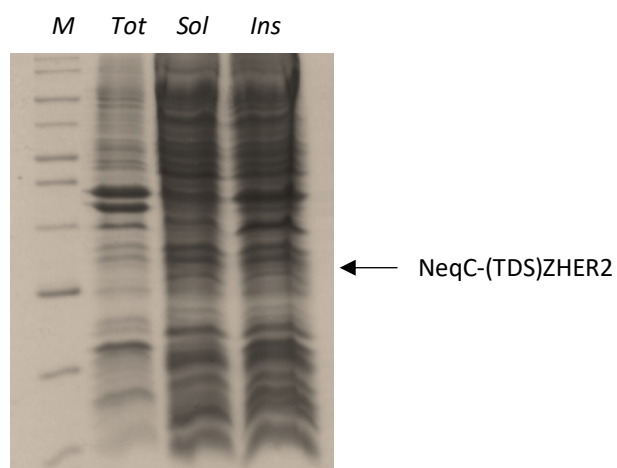


Figura 1. Anàlisi d'expressió i de solubilitat del precursor C, NeqC-(TD)ZHER2 (pskBAD2) a 18°C. M: Marcador de massa molecular; Tot: fracció total; Sol: Fracció soluble; Ins: Fracció insoluble.

Annex E. Purificació de la NeqN

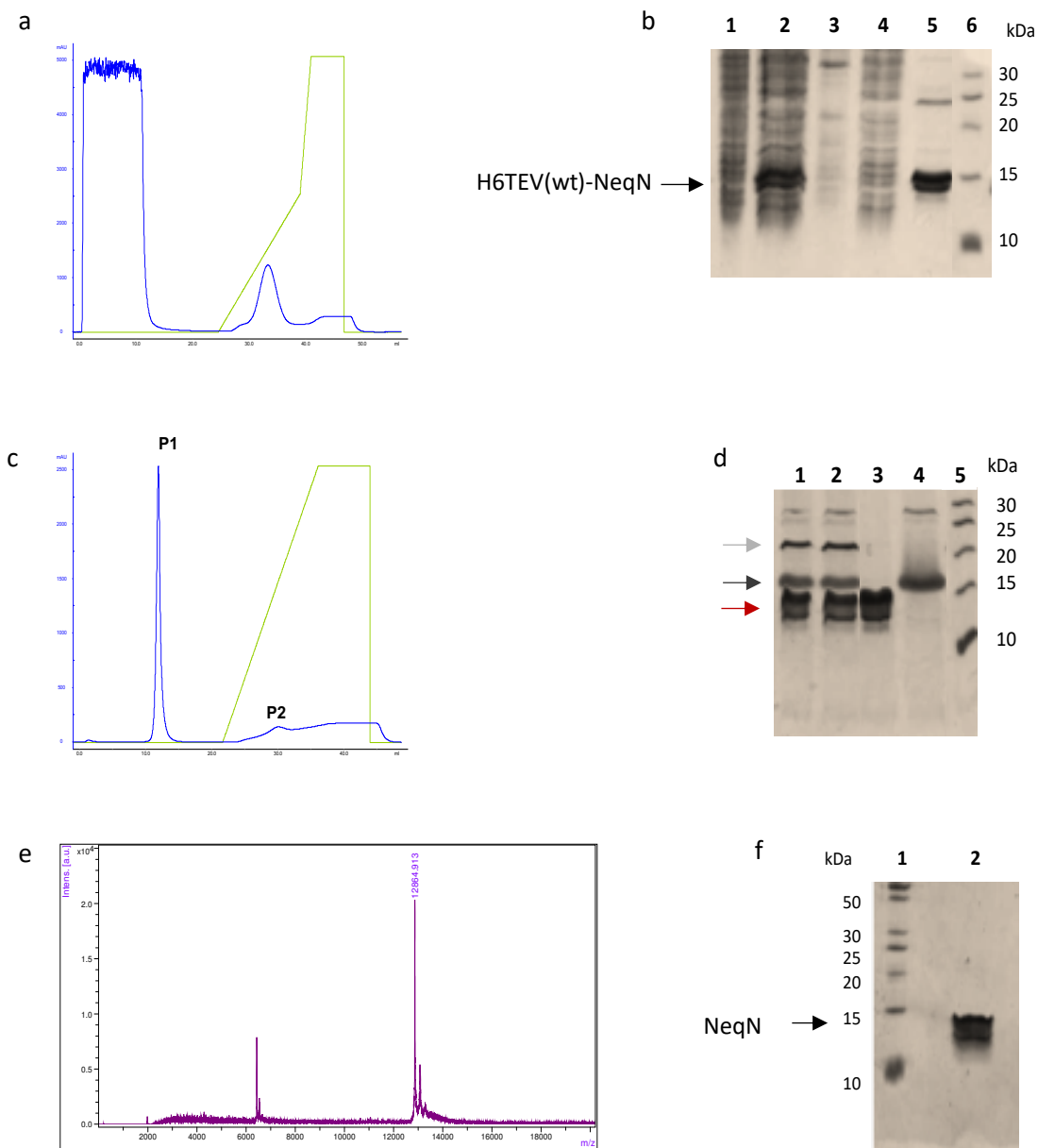
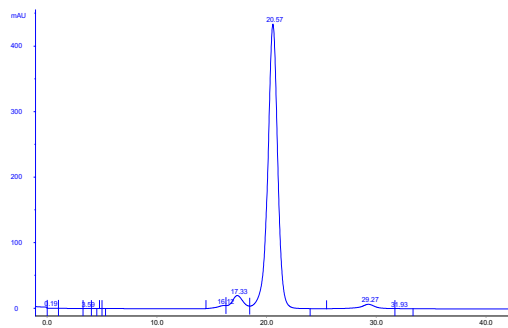
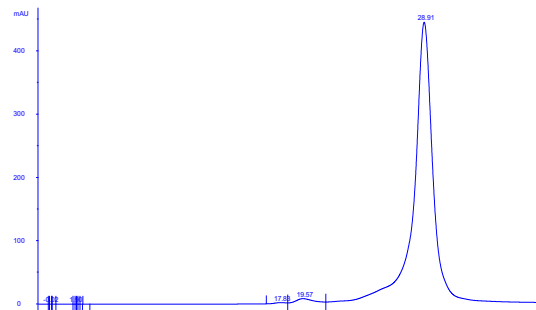


Figura 1. Seguiment del procés de purificació de la NeqN. Purificació del precursor H6TEV(wt)-NeqN (14841.6 Da) (a) Perfil d'absorbància a 280 nm obtingut de la cromatografia Ni-NTA. (b) Anàlisi per SDS-PAGE de les fraccions obtingudes de la cromatografia. (1) Extracte cel·lular abans d'afegir l'IPTG; (2) Fracció soluble i (3) insoluble després de la lisi cel·lular; (4) Través; (5) Eluïció de la cromatografia d'afinitat Ni-NTA. (6) Marcadors de massa molecular. Purificació de la NeqN (12861.5 Da). (c) Perfil d'absorbància a 280 nm obtingut de la cromatografia Ni-NTA realitzada després de la digestió amb la proteasa TEV. (d) Anàlisi per SDS-PAGE de la digestió per la proteasa TEV i de les fraccions obtingudes de la cromatografia. (1) Mostra després de la digestió; (2) Mostra després de la diàlisi; Fraccions obtingudes per cromatografia Ni-NTA: (3) Través (4) Eluïció; (5) Marcadors de massa molecular. Fletxa taronja: H6-TEV (27495.2 Da) Fletxa verda: H6TEV (1998.6 Da). Fletxa vermella: NeqN. Anàlisi de la puresa obtinguda al final del procés de purificació per: (e) MALDI-TOF (f) i per SDS-PAGE. (1) Marcadors de massa molecular; (2) NeqN.

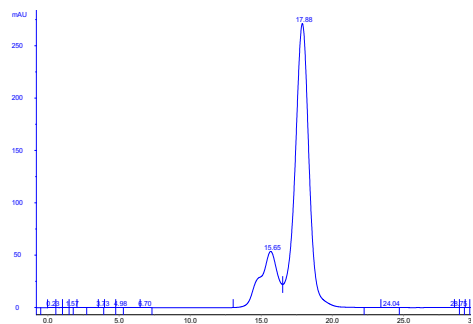
Annex F. Identificació per SEC de la NeqN, NeqC i complex NeqN/NeqC



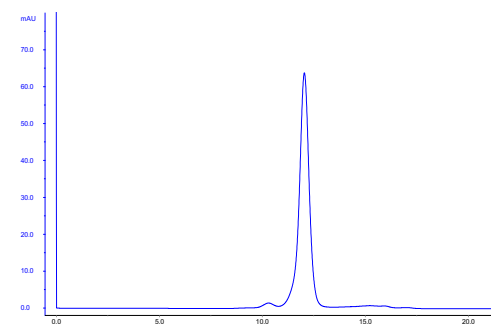
Anhidrasa Carbónica (29000 Da)



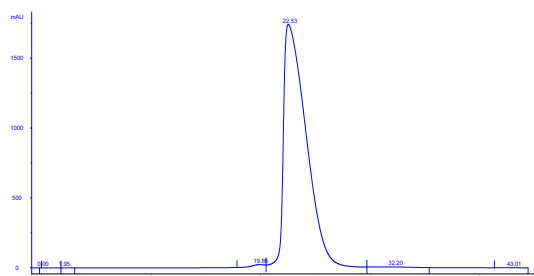
Aprotinina (6500 Da)



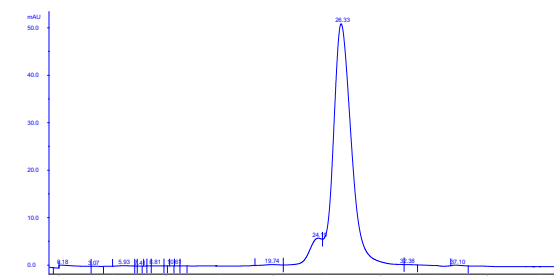
Ovoalbúmina (44000 Da)



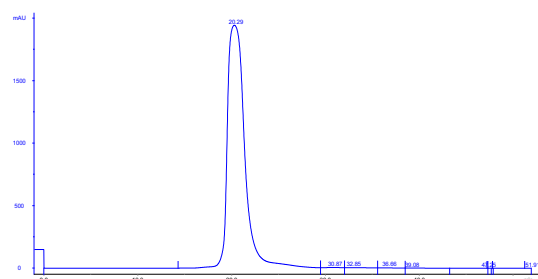
Ribonucleasa A (13700 Da)



NEQn + NEQc (19001,3 Da)



NEQc (6139,8 Da)



NEQn (12861,5 Da)

Figura 1. Perfils d'absorbància a 280 nm obtinguts per SEC.

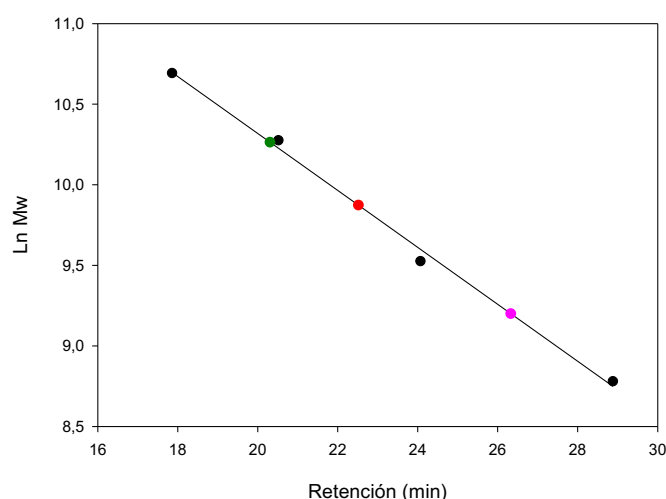


Figura 2. Recta patró per a la determinació de la massa molecular de la NeqN, NeqC i el complex per SEC. La recta patró es va obtenir a partir dels temps de retenció obtinguts de l'anàlisi per cromatografia d'exclusió molecular de diferents proteïnes globular de massa molecular coneguts. (a) Perfils d'absorbància obtinguts de l'anàlisi de les diferents proteïnes amb massa molecular conegut i la NeqN, NeqC i el complex NeqN/NeqC per cromatografia d'exclusió molecular. (b) Recta patró obtinguda a partir dels temps de retenció obtinguts per SEC i el ln de la massa molecular ($R^2= 0.996$) Els punts negres corresponent a les diferents proteïnes de massa molecular conegut, el punt verd correspon a la NeqN, el punt rosa a la NeqC i el punt vermell al complex NeqN/NeqC.

Taula 1. Masses moleculars experimentals obtinguts per SEC de la NeqN, la NeqC i el complex NeqN/NeqC.

	$[\mu M]$	Retenció (min)	Mitjana	Desv. Stand.	Ln Mw	Mw (Da) Experimental	Mw (Da) Teòrica
Complex	356	22,53	22,52	0,01	9,872716	19,3939	19001,3
		22,52					
		22,51					
NEQn	325	20,29	20,31	0,096	10,263223	28,659	12861,5
		20,22					
		20,41					
NEQc	50	26,33			9,199489	9892,07	6139,8

Annex G. Estudi per DSC de l'efecte del tampó sobre la NeqN

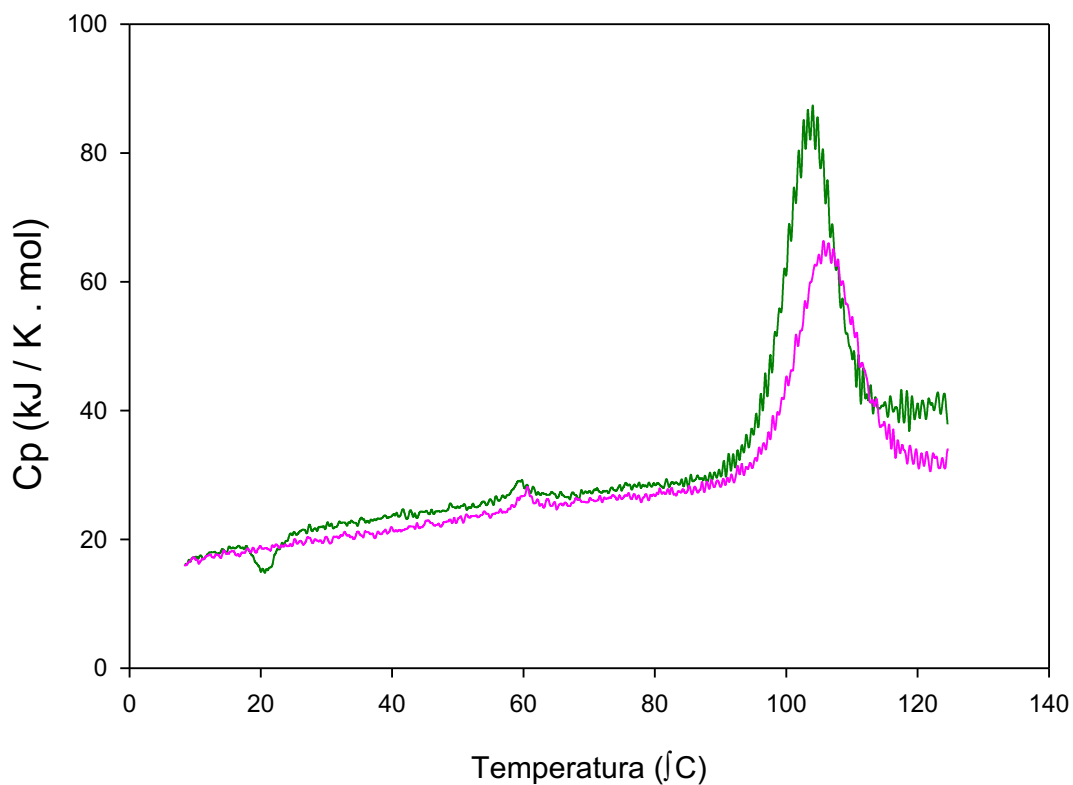


Figura 1. Termograma obtingut per DSC del procés de desnaturalització de la inteïna NeqN en tampó fosfat i tampó Tris-HCl. Experiments realitzats a una concentració de $26 \mu\text{M}$ de proteïna amb 20 mM tampó fosfat/Tris-HCl a pH 8.0, 150 mM NaCl i 1 mM EDTA a una velocitat d'escombrat de $90\text{K}/\text{min}$. Línia rosa: traça de desplegament de la NeqN amb Tris ; Línia verda: traça de desplegament de la NeqN amb Tris

Annex H. Interaccions entre l'exteïna N i la inteïna N, així com les interaccions que estableixen la interfície dimèrica

Taula 1. Ponts d'hidrogen entre l'exteïna N i la inteïnaN.

#	Nom	Atom	Nom	Atom	Distància (Å)
1	Ser-10	[OG]	Ser76	[OG]	3.38
2	Gly-9	[N]	Thr74	[O]	2.83
3	Gly-9	[N]	Thr74	[OG1]	3.51
4	Ser-8	[N]	Thr74	[OG1]	3.40
5	Ser-8	[OG]	Thr74	[OG1]	3.07
6	Lys-6	[N]	Gly72	[O]	2.72
7	Lys-6	[NZ]	Asp48	[OD1]	3.39
8	Lys-6	[NZ]	Asp48	[OD2]	3.47
9	Ile-4	[N]	Ile70	[O]	3.02
10	Tyr-3	[OH]	Glu69	[OE1]	3.45
11	Gly-2	[N]	Tyr68	[O]	3.07
12	Ala-11	[O]	Arg51	[NH1]	2.64
13	Ala-11	[O]	Thr74	[OG1]	3.52
14	Ser-10	[OG]	Arg51	[NH]	3.21
15	Ser-10	[OG]	Ser76	[N]	3.87
16	Ser-8	[O]	Thr74	[N]	3.10
17	Ser-8	[O]	Thr74	[OG1]	3.34
18	Lys-6	[O]	Gly72	[N]	3.01
19	Ile-4	[O]	Ile70	[N]	2.66

7-8: ponts salins

Taula 2. Ponts d'hidrogen i entre cadenes laterals de residus aromàtics presents en la interfície dimèrica.

#	Nom	Atom	Nom	Atom	Distància (Å)
1	Ser83	[OG]	Gln32	[O]	3.04
2	Ser57	[OG]	Tyr56	[OH]	2.76
3	Tyr56	[OH]	Ser57	[OG]	2.66
4	Tyr56	[OH]	Asn71	[OD1]	3.29
5	Asn71	[OD1]	Tyr56	[OH]	3.45
6	Ser83	[OG]	Gln32	[N]	2.93
7	Phe40	-	Lys54	-	4.7
8	Phe-7	-	Phe-7	-	-3.9

7-8: interaccions π

1 i 6: interaccions asimètriques.

Annex I. Espectres d'emissió de fluorescència i corbes d'unió obtingudes del procés d'associació entre els dos fragments de la inteïna Neq pol a diferents temperatures per a estudiar la cinètica de la reacció

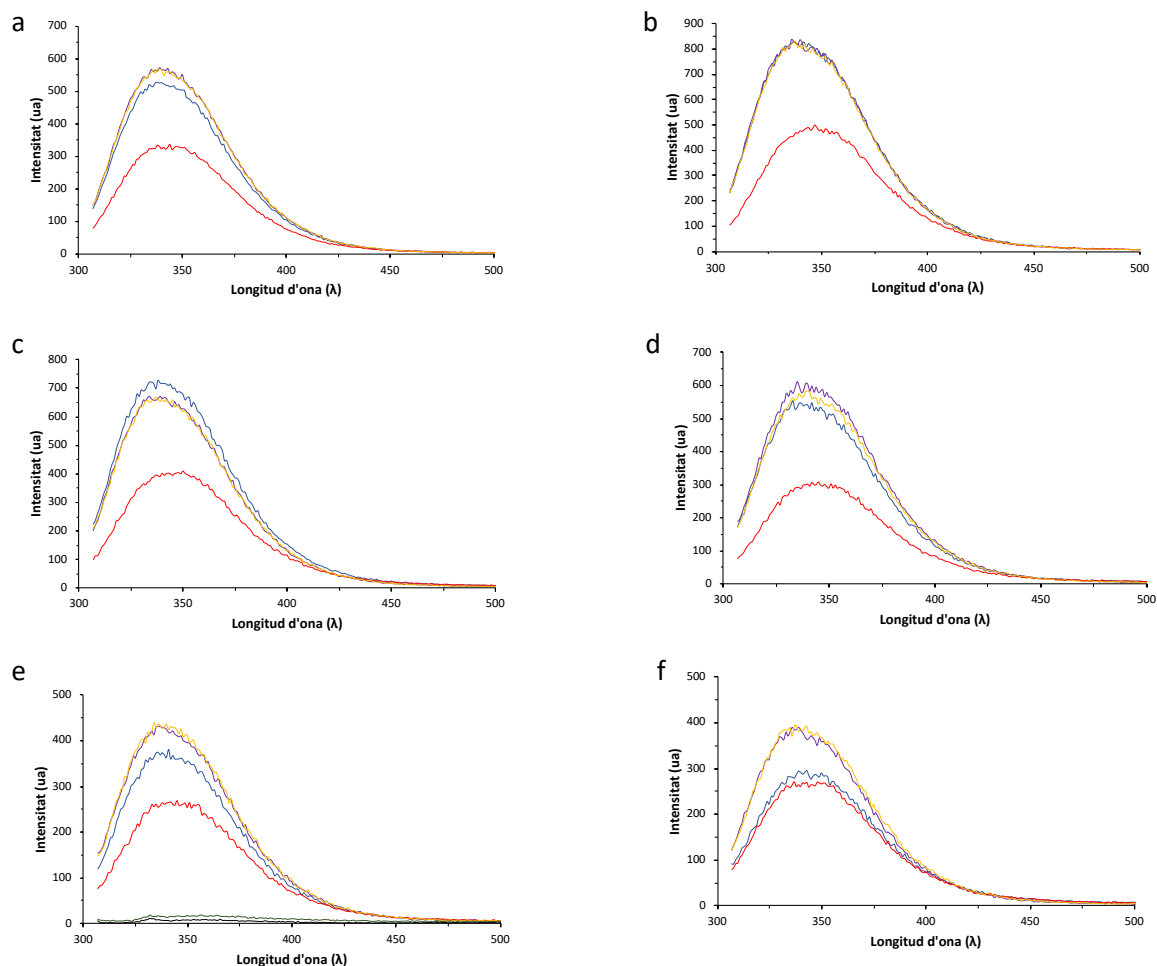


Figura 1. Espectres d'emissió de fluorescència ($\lambda = 298 \text{ nm}$) del procés d'associació entre NeqN i NeqC-obtinguts a diferents temperatures. (a) 25°C; (b) 36°C; (c) 46°C; (d) 55°C; (e) 64°C; (f) 74°C. Les línies de color morat i groc corresponen als espectres obtinguts de la NeqN sola al inici i al final de la reacció, respectivament. Les línies de color blau i vermell corresponen als espectres obtinguts de la NeqN al inici i al final del procés d'associació, respectivament. En l'espectre a 64°C en verd es troba l'espectre obtingut de la NeqC i en negre del tampó.

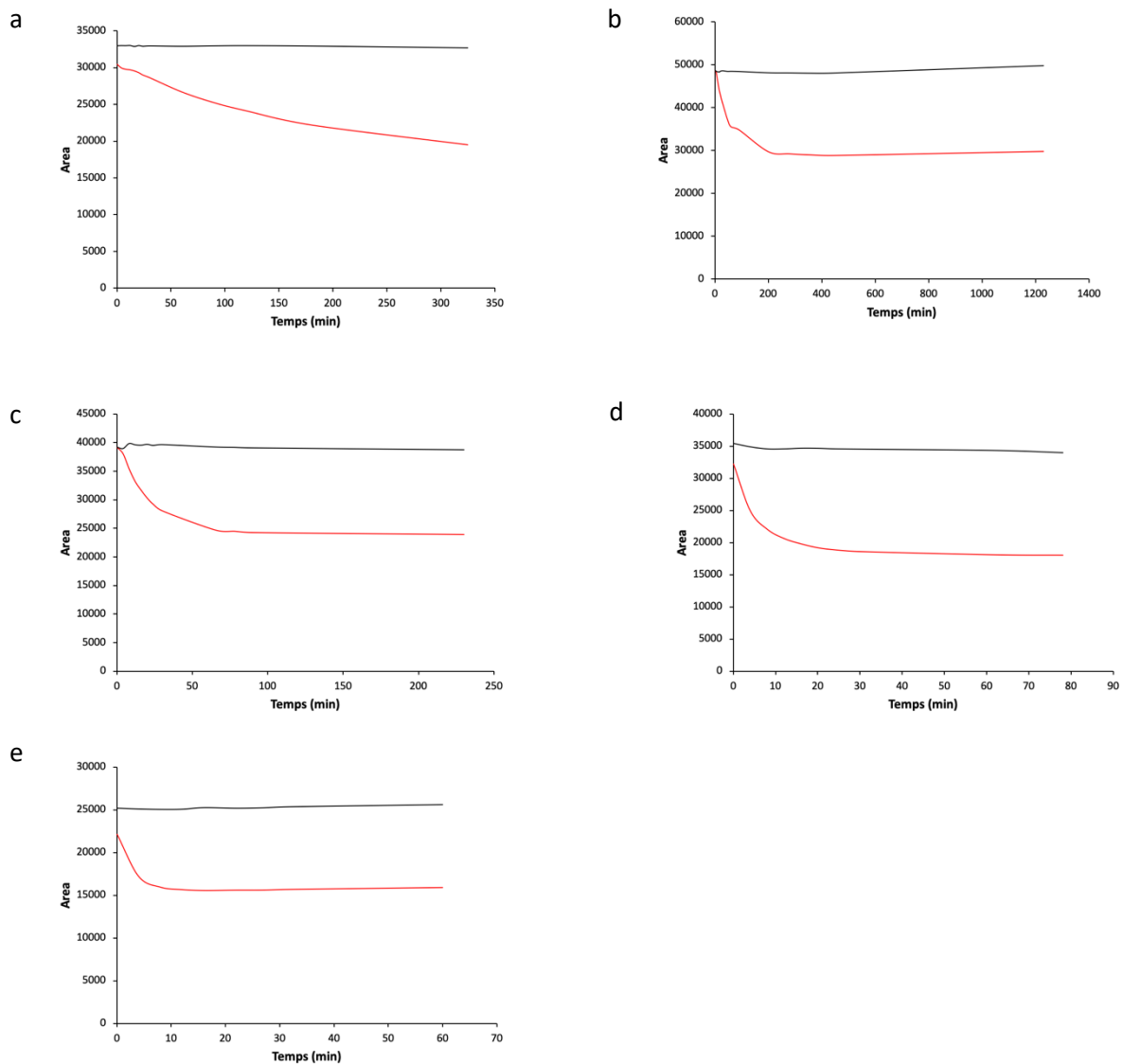


Figura 2. Corbes d'unió a diferents temperatures. (a) 25°C; (b) 36°C; (c) 46°C; (d) 55°C; (e) 64°C. La línia de color vermell representa la disminució de la fluorescència al llarg del temps de la NeqN acomplexada, i la línia de color negra correspon al seguiment de la fluorescència de la NeqN lliure.

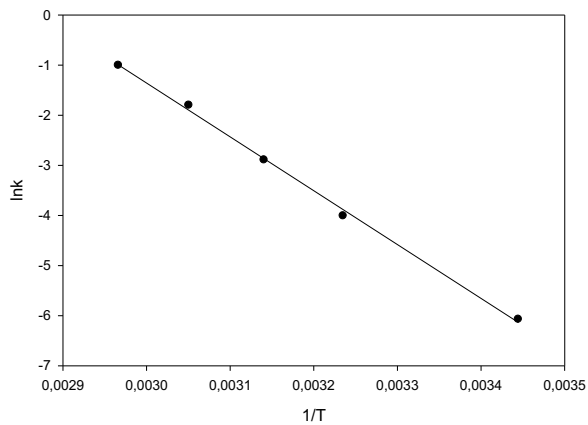


Figura 3. Gràfic d'Arrhenius.

Annex J. Espectres d'emissió de fluorescència i corbes d'unió obtingudes del procés d'associació entre els dos fragments de la inteïna Neq pol a concentracions creixents de NeqC per a l'estudi dels paràmetres termodinàmics del procés.

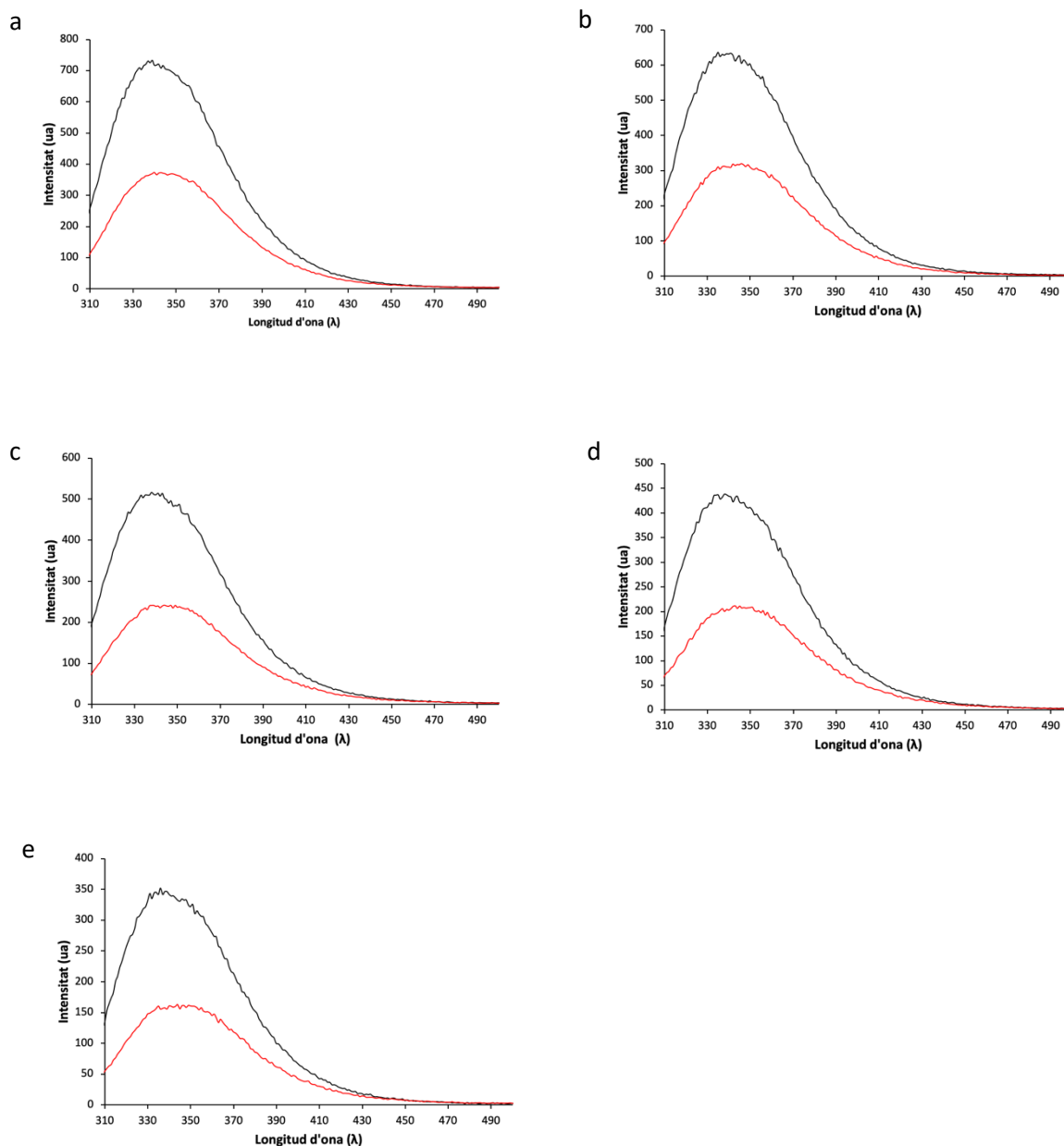


Figura 1. Espectres d'emissió de fluorescència ($\lambda = 298$ nm) de la (NeqN a una concentració de $24 \mu\text{M}$ en absència i en presència de concentracions creixents de NeqC a diferents temperatures. (a) 25°C ; (b) 30°C ; (c) 40°C ; (d) 50°C ; (e) 60°C . Els espectres en color negre corresponen als obtinguts per la NeqN sola, i els espectres de color vermell corresponen als adquirits de la combinació de $24 \mu\text{M}$ (NeqN + $61 \mu\text{M}$ de NeqC).

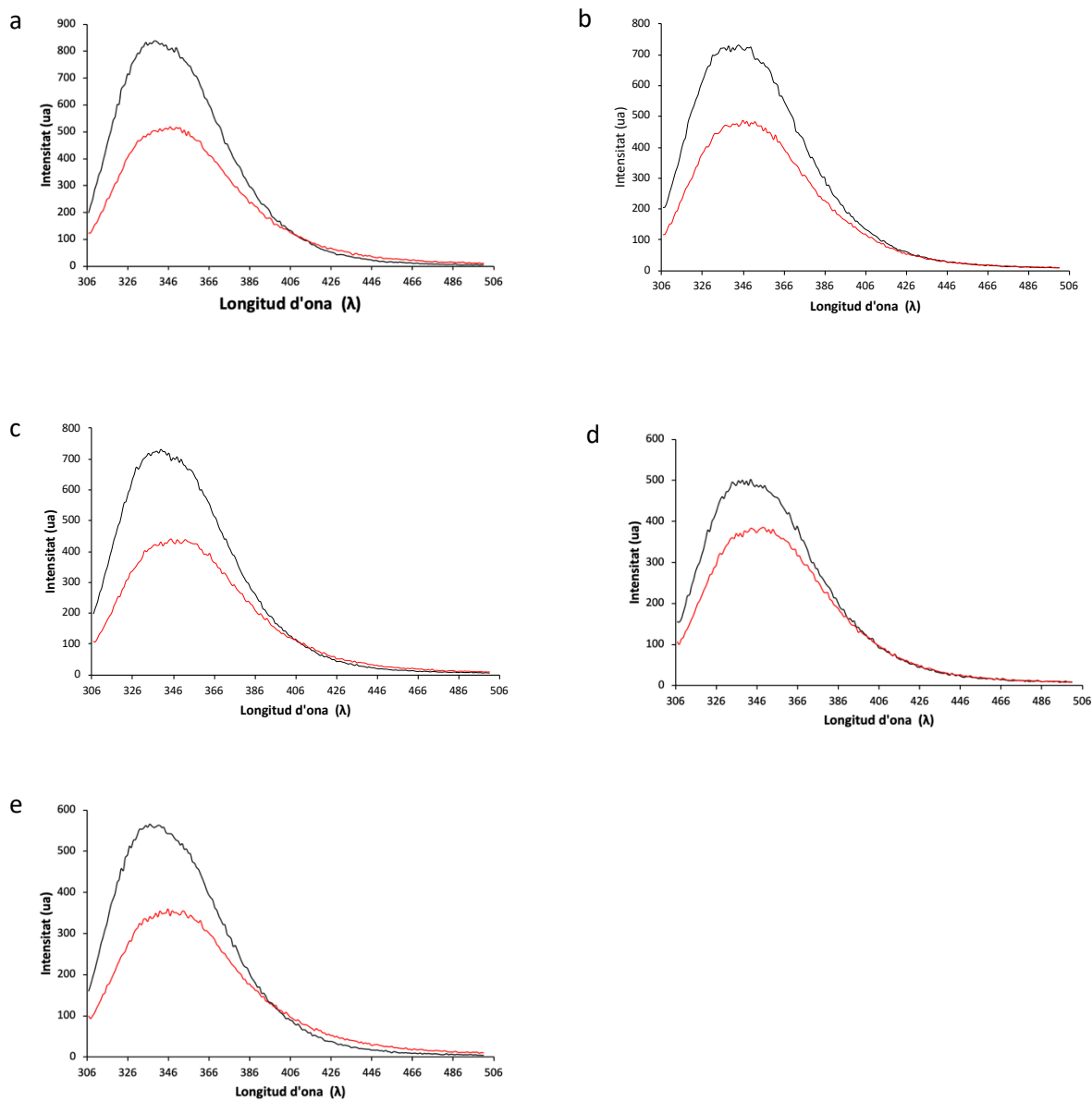


Figura 2. Espectres d'emissió de fluorescència ($\lambda=298$ nm) de la NeqN a una concentració de $14,4 \mu\text{M}$ en absència i en presència de concentracions creixents de NeqC a diferents temperatures (a) 20°C ; (b) 25°C ; (c) 30°C ; (d) 35°C ; (e) 40°C . Els espectres en color negre corresponen als obtinguts per la NeqN sola, i els espectres de color vermell corresponen als adquirits de la combinació de $14,4 \mu\text{M}$ NeqN + $46 \mu\text{M}$ de NeqC.

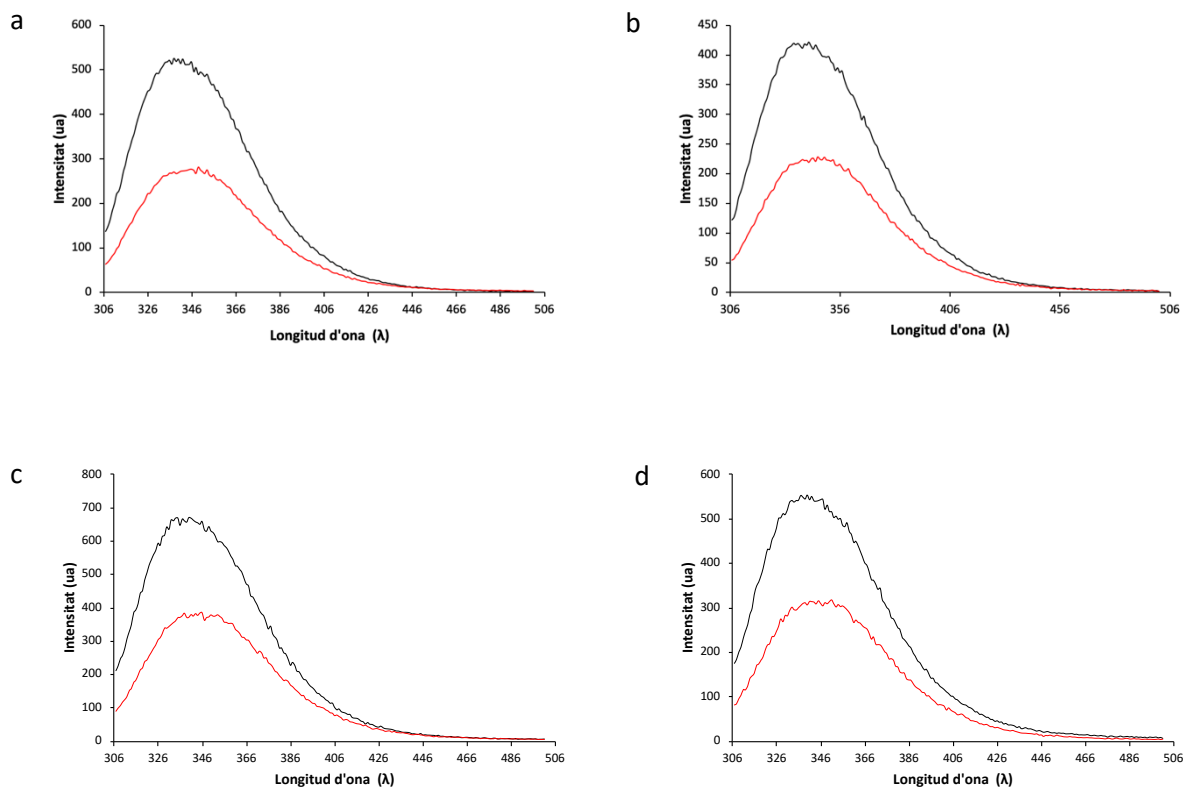


Figura 3. Espectres d'emissió de fluorescència ($\lambda = 298$ nm) de la NeqN a una concentració de $20 \mu\text{M}$ en absència i en presència de concentracions creixents de NeqC a diferents temperatures (a) 25°C ; (b) 35°C ; (c) 45°C ; (d) 55°C ; . Els espectres en color negre corresponen als obtinguts per la NeqN sola, i els espectres de color vermell corresponen als adquirits de la combinació de $14,4 \mu\text{M}$ NeqN + $46 \mu\text{M}$ de NeqC.

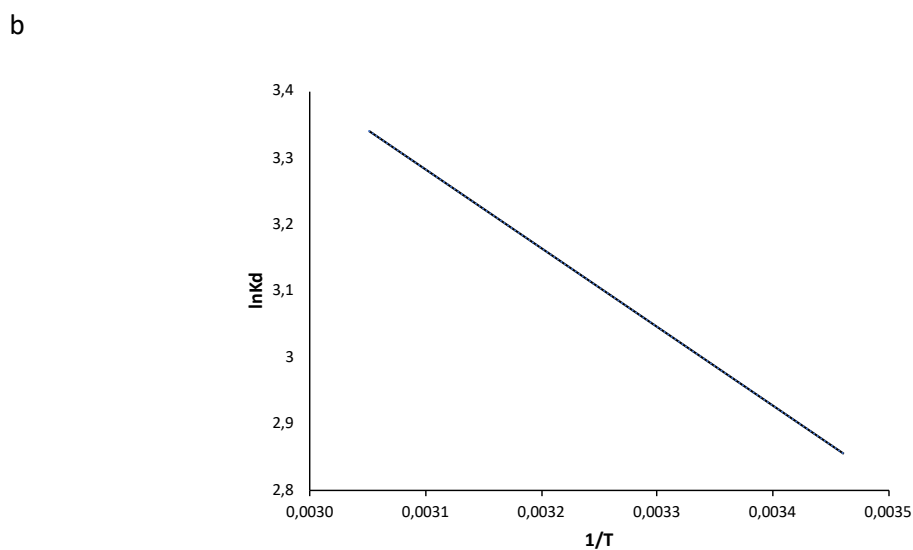
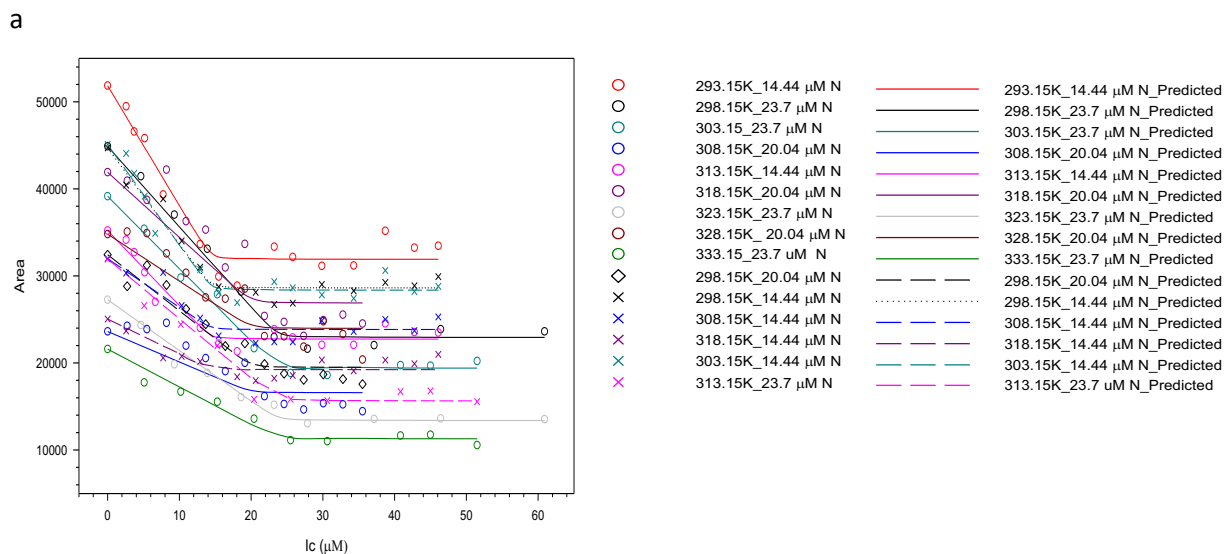


Figura 4. (a) Corbes d'unió (b) regressió lineal.

Annex K. Interaccions que estableixen la interfície del complex NeqN/NeqC

Taula 1. Ponts d'hidrogen presents en la interfície d'associació entre els dos fragments de la inteïna Neq.

#	NeqC		NeqN		Distància (Å)
	Nom	Atom	Nom	Atom	
1	Arg2	[N]	Glu69	[O]	3.24
2	Arg2	[NE]	Glu69	[OE1]	2.8
3	Arg2	[NH1]	Glu69	[O]	3.49
4	Tyr3	[OH]	Asp66	[OD1]	2.55
5	Leu4	[N]	Leu67	[O]	2.87
6	Lys6	[NZ]	Glu64	[O]	5.70
7	Lys6	[NZ]	Asp66	[OD2]	4.40
8	Lys7	[N]	Thr65	[O]	3.14
9	Arg8	[NE]	Glu62	[OE2]	2.80
10	Arg8	[NH2]	Glu35	[OE1]	2.97
11	Arg8	[NH2]	Glu62	[O]	3.64
12	Val9	[N]	Thr63	[O]	3.26
13	Leu11	[N]	His61	[O]	2.85
14	Tyr12	[N]	Ile2	[O]	2.77
15	Asp13	[N]	Ile59	[O]	2.93
16	Ser15	[N]	Ser57	[O]	2.87
17	Glu17	[N]	Lys54	[O]	2.97
18	Ser18	[OG]	Asp48	[OD2]	2.69
19	Lys20	[NZ]	Asp48	[OD2]	4.2
20	Lys20	[NZ]	Glu50	[OE1]	3.81
21	Tyr22	[N]	Tyr47	[O]	2.91
22	Asn24	[ND2]	Asn71	[O]	2.71
23	His29	[N]	Asp-1	[OD2]	2.86
24	Asp+2	[N]	Ile-4	[O]	3.25
25	Arg2	[O]	Glu69	[N]	2.89
26	Gly5	[O]	Leu67	[N]	3.01
27	Lys7	[O]	Thr65	[N]	2.98
28	Val9	[O]	Thr63	[N]	3.27
29	Leu11	[O]	His61	[N]	2.89
30	Tyr12	[O]	Ile2	[N]	2.93
31	Asp13	[O]	Ile59	[N]	2.97
32	Ser15	[O]	Tyr56	[N]	2.82
33	Ser15	[O]	Ser57	[N]	3.41
34	Glu17	[OE1]	Tyr56	[OH]	2.80
35	Asn24	[O]	Asn71	[ND2]	2.92
36	His29	[O]	Asp-1	[N]	2.75
37	Thr+1	[O]	Gly-2	[N]	2.87
38	Asp+2	[O]	Ile-4	[N]	3.38

Les interaccions entre els residus s'han obtingut utilitzant el servidor PISA[185].

El grup 1 d'interaccions polars està marcat en color verd, el grup 2 en color lila, el grup 3 en color blau i el grup 4 en color vermell