

II-Introducció general

II.A Els inhibidors de proteases

Els enllaços peptídics de les proteïnes són proteolitzats per les peptidases, també anomenades enzims proteolítics o proteases, que desenvolupen un paper essencial en un gran ventall de processos que tenen lloc en els organismes vius, des dels virus i bacteris fins als mamífers. Els inhibidors de proteases són els principals responsables de la regulació de l'activitat d'aquests enzims, convertint-se en un element clau en el control dels processos proteolítics que tenen lloc en els éssers vius. En alguns microorganismes s'han descrit inhibidors no proteics que tenen com a diana proteases de l'hoste (Bode i Huber, 2000)¹, però en mamífers els inhibidors endògens descrits són de natura proteica (Bode i Huber, 1992). El seu espectre d'inhibició és més o menys ampli en funció de cada cas. Tot i que s'han descrit alguns inhibidors que actuen simultàniament sobre 2 proteases de diferents classes, com és el cas de l'equistatina que inhibeix cisteïn-proteases i aspàrtic-proteases (Bode i Huber, 1992), normalment inhibeixen un o més enzims que pertanyen a la mateixa família de peptidases. Això ha fet que tradicionalment els inhibidors de proteases s'hagin classificat en quatre grans grups en funció del tipus de proteases que inhibeixen:

- Inhibidors de serín-proteases
- Inhibidors de cisteïn-proteases
- Inhibidors d'aspàrtic-proteases
- Inhibidors de metal.lo-proteases

Dins de cada grup els inhibidors es classifiquen en diferents famílies en funció d'homologies seqüencials o estructurals, o bé segons el mecanisme d'unió a la proteasa.

II.A.1 Inhibidors de serín-proteases

La majoria d'inhibidors que pertanyen a aquest grup són proteïnes de baix pes molecular (de 28 a 200 residus) i presenten un mecanisme comú d'unió amb l'enzim tipus substrat (*substrate-like*) (Bode i Huber, 1992). Les serpines, que són unes molècules de tamany més gran, representen una excepció dins aquest grup. El seu mecanisme d'interacció amb l'enzim, tot i ser també de tipus substrat, provoca una sèrie de canvis conformacionals irreversibles a l'inhibidor i això condueix a la formació de complexos molt més estables (Wright, 1996). Dins els inhibidors de serín-proteases més petits trobem els de la família *squash* que pel seu plegament pertanyen a la superfamília de les proteïnes T -knot, en la qual també trobem el PCI, les conotoxines de verins de serp i les proteïnes de la família EGF (Lin i Nussinov, 1995). Totes aquestes proteïnes comparteixen una característica comuna que és la presència de 3 ponts disulfur formant un nus de cisteïnes (T -knot), en el qual dos ponts disulfur i la cadena polipeptídica formen un anell a través del qual passa el tercer pont.

¹ Les referències d'aquest capítol es troben recollides a l'apartat **IX-Bibliografia**.

II.A.2 Inhibidors de cisteín-proteases

Dins aquest grup trobem els inhibidors de les calpaïnes, de les cisteín proteases tipus papaïna (*papain-like*) i de les fitocistatines.

Els inhibidors de les calpaïnes, les cisteín -proteases dependents de calci, estan involucrades en diversos processos patològics, com ara la inflamació, la distròfia muscular o l'Alzheimer (Suzuki i Sorimachi, 1998). No obstant, el grup més nombrós i més estudiat correspon al dels inhibidors de cisteín-proteases tipus papaïna, que es classifiquen en 3 famílies: estefines, cistatines i quininògens (Turk i Bode, 1991). La interacció entre les cistatines i les cisteín-proteases tipus papaïna representen el mecanisme d'unió tipus no-substrat (*non-substrate*). Les cistatines presenten una fulla beta amb 5 làmines beta antiparal·leles que embolcallen una hèlix- α central (Bode i col., 1988). Dos dels llaços que connecten les làmines β entre si són responsables de la interacció principal amb l'enzim, formant una zona hidrofòbica altament complementària a una zona de la superfície adjacent al seti actiu, impedit l'accés al substrat. La cua N-terminal interacciona amb el residu de cisteïna del seti actiu aconseguint que la unió sigui més forta i selectiva. Les estefines són molècules d'uns 11 kDa que no presenten ponts disulfur ni carbohidrats i normalment es troben al compartiment intracel·lular. Les cistatines tenen un pes molecular d'uns 13 kDa, contenen 2 o 3 ponts disulfur i són sintetitzades amb un pèptid senyal que les condueix probablement a l'espai extracel·lular (Sotiropoulou i col., 1997). Dins els quininògens es diferencien 3 tipus d'inhibidors en funció del pes molecular, però tots contenen 3 dominis tipus cistatina i dos d'ells presenten activitat inhibidora. A més d'inhibir les cisteín-proteases tipus papaïna, també estan involucrats en altres processos com la inflamació o la cascada de coagulació sanguínia. Les cistatines de plantes o fitocistatines presenten homologia seqüencial amb les cistatines però comparteixen amb les estefines l'absència de ponts disulfur o residus de cisteïna en la molècula i actualment es classifiquen com a una nova família dins la superfamília de les cistatines (Margis i col., 1998).

II.A.3 Inhibidors de metal·lo-proteases

Dins d'aquest grup trobem els inhibidors tissulars de les metal·loproteases de matriu (TIMPs) i els inhibidors de les metal·locarboxipeptidases. Les metal·loproteases de matriu (MMPs) es troben a la matriu extracel·lular i juntament amb els seus inhibidors són responsables del recanvi de la matriu i també estan involucrats en el procés de desenvolupament embrionari, en la remodelació tissular i en diverses malalties com l'artritis o l'arteriosclerosi (Nagase, 1996). S'han descrit només 4 inhibidors exògens de metal·locarboxipeptidases: l'inhibidor de patata (PCI) de 39 residus, l'inhibidor de tomata (TCI) de 37 residus, l'inhibidor del nemàtode *Ascaris suum* de 65 residus i el de sangonera (LCI) de 66 residus (Hass i col., 1976; Hass i Hermudson, 1981; Homandberg i col., 1989 i Reverter i col., 1998). També s'ha descrit un inhibidor endogen en mamífers que es troba en el cervell i inhibeix la CPA1, CPA2, CPA de mastòcits i CPB (Normant i col., 1995).

La interacció entre aquests inhibidors i els corresponents enzims, tant en el cas dels TIMPs com dels inhibidors de metal·locarboxipeptidases, és de tipus substrat; tan sols la interacció entre les metal·locarboxipeptidases i el seu segment pro és de tipus no-substrat.

II.A.4 Inhibidors d'aspàrtic-proteases

Actualment s'han descrit només dos inhibidors que pertanyen a aquest grup, l'inhibidor de pepsina que fou trobat en el nemàtode *Ascaris suum* (Khan i James, 1998 i Davies, 1990) i l'equistatina, que és un inhibidor bifuncional de cisteïn-proteases tipus papaïna i d'aspàrtic -proteases, que fou aïllat d'una anèmona marina (Lenarcic i Turk, 1999).

II.A.5 Els inhibidors de proteases emprats en aquest treball

II.A.5.1 L'inhibidor de carboxipeptidasa de patata (PCI)

El PCI és un dels pocs inhibidors proteics de metal·locarboxipeptidases que es coneixen. Es presenta a la natura en 5 variants que van de 37 a 39 residus i que responen a diferències isomòrfiques i de processament. Les diferències seqüencials entre elles són mínimes i totes presenten una activitat inhibidora similar enfront la CPA (Figura 1) (Hass i Derr, 1979a).

PCI-Ia:	EQHADPI C NKP C KTHDDCSGAWFCQACWNAAGTCGPYVG
PCI-Ib:	EHADPI C NKP C KTHDDCSGAWFCQACWNAAGTCGPYVG
PCI-IIa:	EQHADPI C NKP C KTHDDCSGAWFCQACWNSARTCGPYVG
PCI-IIb:	EHADPI C NKP C KTHDDCSGAWFCQACWNAAGTCGPYVG
PCI-III:	HADPI C NKP C KTHDDCSGAWFCQACWNAAGTCGPYVG

Figura 1. Seqüència aminoacídica de les isoformes del PCI (Hass i Derr, 1979a).

L'estructura del PCI complexat amb la CPA bovina fou resolta per difracció de raigs X (Rees i Lipscomb, 1982) i lliure en solució per RMN (Clare i col., 1987). La proteïna madura (isoforma IIa), de 39 residus, està organitzada en un nucli central globular de 27 residus, del qual protuberen una cua N-terminal de 7 residus i una cua C-terminal de 5 residus. El nucli, estabilitzat per 3 ponts disulfur, forma una estructura anomenada nus de cisteïnes (T-knot) i conté una volta d'hèlix 3_{10} (residus 14 a 18) que constitueix l'única estructura secundària regular de la proteïna (Figura 2). Els 4 últims residus de la cua C-terminal formen el lloc d'unió primari amb la CPA i estableixen interaccions en el solc on es troba el seti actiu de la proteasa. La glicina C-terminal del PCI és tallada en el complex, però roman ancorada a la butxaca del seti actiu de la CPA. La zona d'interacció secundària del PCI amb la peptidasa està formada pels residus 27 a 30 i la cadena lateral del residu 15 (His).

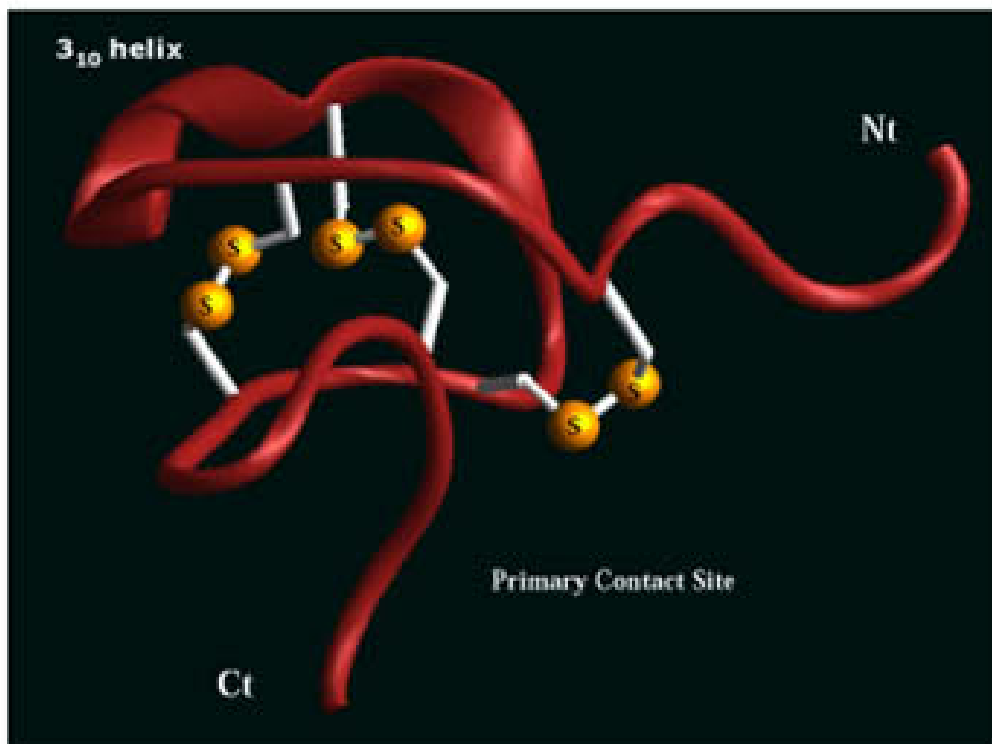


Figura 2. Estructura tridimensional del PCI

La major concentració de PCI en la planta es troba en el tubercle, tot i que també està present a les fulles, als brots i a la tija (Hass i Derr, 1979b). Mitjançant el triatge (*screening*) d'una genoteca de cDNA construïda a partir de mRNA de fulles de patata tractades amb àcid abscísic i usant com a sonda el cDNA del TCI, es va obtenir la seqüència completa de la regió codificant. Aquesta està formada per un pèptid senyal N-terminal de 29 residus, un pro-pèptid N-terminal de 27 residus, la proteïna madura de 39 residus i un pro -pèptid C-terminal de 7 residus (Villanueva i col., 1998). Estudis de *Northern blot* van mostrar que la transcripció del gen del PCI s'indueix per ferida i per acció de les fitohormones àcid abscísic i àcid jasmònic, que formen part de la ruta de resposta a ferida. La localització subcel.lular del PCI es va determinar per estudis immunocitoquímics i per microscopia electrònica i es va observar que la màxima acumulació tenia lloc en el vacúols (Hollander-Czytko i col., 1985 i Villanueva i col., 1998).

El caràcter hidrofòbic del propèptid C-terminal del PCI i les similituds seqüencials amb altres pèptids senyal de transport als vacúols suggereixen que la seva funció és el direccionament de la proteïna vers aquest compartiment cel.lular (Wilkins i Raikhel, 1989). Paral.lelament, també pot tenir un paper en la modulació de l'activitat de l'inhibidor, la qual cosa es desprèn d'estudis realitzats amb una proteïna recombinant que contenia la proteïna madura i el pro -pèptid C-terminal i on s'observà que aquesta proteïna, tot i estar ben plegada, no presentava activitat inhibidora enfront la CPA (Villanueva i col., 1998).

Fins ara no s'ha pogut determinar la funció del pro-pèptid N-terminal i tampoc existeix cap seqüència a la base de dades que presenti una homologia suficient com per deduir-ne una funcionalitat. No obstant, s'han descrit un gran nombre de proteïnes que se sintetitzen com a precursors i que necessiten el pro-pèptid per al seu plegament (Shinde i Inouye, 2000). En aquests casos el pro-pèptid rep el nom de chaperona intramolecular, ja que la seva funció és equivalent a la de les chaperones moleculars, que actuen unint-se a les proteïnes naixents dels ribosomes i catalitzen el seu plegament. Aquestes dades, juntament amb la baixa eficiència del plegament *in vitro* de la proteïna madura suggerien que el pro -pèptid N-terminal del PCI podria actuar ajudant en el procés de plegament. Aquesta hipòtesi de treball s'analitza al capítol V d'aquesta tesi, on s'estudia el possible paper de les regions precursoras del PCI en el plegament de la proteïna.

II.A.5.2 L inhibidor de carboxipeptidasa de sangonera (LCI)

El LCI és el primer inhibidor de metal·locarboxipeptidases aïllat de sangonera (*Hirudo medicinalis*). L'estructura del LCI lliure en solució s'ha determinat per RMN; i complexat amb la CPA s'ha determinat per difracció de raigs X (Reverter i col., 2000). La proteïna de 66 residus està organitzada en un nucli central globular de 52 residus format per una fulla β de 5 cadenes antiparal·leles i una petita hèlix α ; una cua N-terminal de 9 residus i una cua C-terminal de 5 residus. Els elements d'estructura secundària estan connectats entre si i estabilitzats per quatre ponts disulfur. Els últims quatre aminoàcids del LCI constitueixen la regió primària d'interacció amb la CPA, els quals se situen en el seti actiu de l'enzim. La regió secundària d'interacció està constituïda per residus dels llaços β 1- β 2, β 3- α 1 i de l' α 1; que interaccionen a través de 3 ponts d'hidrogen i per forces de van der Waals.

La funció més probable del LCI a la sangonera és contribuir al manteniment de la sang en estat líquid, tant durant la ingesta com al seu interior. La CPB de plasma o TAFI, en la seva forma activa, actua proteolitzant residus de lisina i arginina de la fibrina. La presència d'aquests residus d'arginina i lisina permet l'ancoratge del plasminogen sobre la superfície de la fibrina, que és un pas essencial en l'inici de la reacció de fibrinòlisi. La inhibició de l'enzim per part del LCI impedeix la proteòlisi d'aquests residus, afavorint la reacció de fibrinòlisi i per tant, participant en la eliminació dels coàguls sanguinis (Bajzar i col., 1995).

A l'últim treball d'aquesta tesi es compara el plegament del LCI per bescanvi D/H amb els resultats prèviament obtinguts al nostre grup mitjançant la tècnica de captura en medi àcid seguit per RP-HPLC (Salamanca i col., resultats no publicats).

II.A.5.3 La hirudina

La hirudina constitueix una família d'inhibidors específics de trombina que contenen de 64 a 66 aminoàcids i que foren aïllats de les glàndules salivals de les sangonera (Harvey i col., 1986).

La hirudina conté un domini globular N-terminal constituït per dues fulles β antiparal·leles que formen el nucli de la molècula i que està estabilitzat per tres ponts disulfur. La cua C-terminal (residus 49 a 65), que constitueix el domini funcional, presenta una configuració en β (Rydel i col., 1991). La principal interacció entre la hirudina i la peptidasa no es dona en el seti actiu de la trombina, sinó en el seti d'unió secundari. Els residus de la cua C-terminal estableixen interaccions iòniques amb el seti d'unió secundari i els 5 últims aminoàcids constitueixen una volta d'hèlix 3_{10} que interacciona amb una zona hidrofòbica de la superfície de la trombina. Finalment, l'extrem N-terminal de la hirudina, que presenta una conformació globular, s'uneix al seti actiu de la trombina.

La hirudina actua inhibint específicament la trombina mitjançant la formació d'un complex no covalent d'unió forta (*tight binding*). La trombina és una serín-proteasa que té un paper central en la coagulació sanguínia, ja que hidrolitza el fibrinogen convertint-lo en monòmers de fibrina, que constitueixen la base dels coàguls. La interacció entre la hirudina i la trombina impedeix la hidròlisi del fibrinogen i d'aquesta manera la hirudina actua impedit l'acció pro-coagulant de la trombina. En el cas concret de la sangonera, l'acció de la hirudina seria semblant a la del LCI, intervindria en el manteniment de la sang en estat líquid, tant durant la ingesta com a l'interior de la sangonera.

El plegament oxidatiu de la hirudina, prèviament estudiat pel mètode de captura en medi àcid seguit per RP-HPLC (Chatrenet i Chang, 1993), s'estudia en el darrer treball d'aquesta tesi mitjançant la tècnica de bescanvi D/H seguida per MALDI-TOF.

II.B L expressió heteròloga i la producció recombinant de proteïnes

L'expressió heteròloga ens permet produir quantitats importants de proteïnes mitjançant les tècniques d'ADN recombinant, per tal de realitzar recerca bàsica i també amb fins comercials. A través de l'enginyeria de proteïnes, les proteïnes diana poden ser redissenyades per tal de millorar la seva estabilitat o solubilitat, modificar la seva activitat o crear proteïnes multifuncionals, i d'aquesta manera es poden obtenir mutants o altres variants de la molècula que no es troben *in vivo* (Oline i Lee, 1993). Actualment, en la era post-genòmica, una de les principals tasques a realitzar és determinar les funcions de totes les proteïnes codificades en els genomes dels organismes model seqüenciats. Això implica obtenir informació sobre la seva regulació espacial i temporal, la seva activitat bioquímica, les seves interaccions i modificacions post-traduccional. L'expressió heteròloga i purificació d'aquestes proteïnes permet elucidar la seva activitat biològica i estudiar les interaccions proteïna-proteïna que són aspectes centrals per a la comprensió de la funció.

Des d'un punt de vista comercial, s'han desenvolupat aplicacions agrícoles, industrials i mèdiques per un gran nombre de proteïnes. En el camp de la medicina és on s'ha produït una màxima incorporació de les tècniques d'expressió heteròloga, que han permès l'ús d'un ampli ventall de proteïnes com a elements terapèutics. Vacunes recombinants, anticossos monoclonals, factors de creixement, enzims, fibrinolítics i altres proteïnes involucrades en múltiples desordres mèdics, es

troben en el mercat o bé s'estan provant clínicament. La possibilitat de poder produir i purificar les proteïnes d'una manera reproduïble, amb un cost acceptable i que puguin ser redissenyades, ha convertit l'expressió heteròloga de proteïnes en una tècnica destacada dins la indústria farmacèutica (Swartz, 2001).

II.B.1 Expressió en organismes procariotes. *Escherichia coli*

Escherichia coli és el microorganisme procariota més emprat per a l'expressió heteròloga de proteïnes. És un sistema senzill, força econòmic i la seva fisiologia i genètica han estat a bastament estudiades. No obstant, presenta alguns inconvenients, com la poca capacitat de secretar proteïnes i la seva incapacitat de realitzar modificacions post-traduccional.

Les peculiaritats de cada gen, mRNA o proteïna fan que no es pugui parlar d'un procés general d'expressió, sinó que en cada cas l'estratègia a seguir sigui diferent. En termes generals es poden definir una sèrie de components dels vectors d'expressió i de les soques d'*E.coli* que influeixen de manera important en l'expressió de la proteïna d'interès:

II.B.1.1 Promotor

El promotor consisteix en una seqüència de ADN situada en direcció 5' (*upstream*) del lloc d'unió del ribosoma al ADN (RBS) i controla l'inici de la transcripció quan se li uneix una RNA polimerasa dependent de ADN. El promotor també es troba sota el control de gens reguladors, que poden trobar-se en el vector o bé estar integrats en el cromosoma de l'hoste.

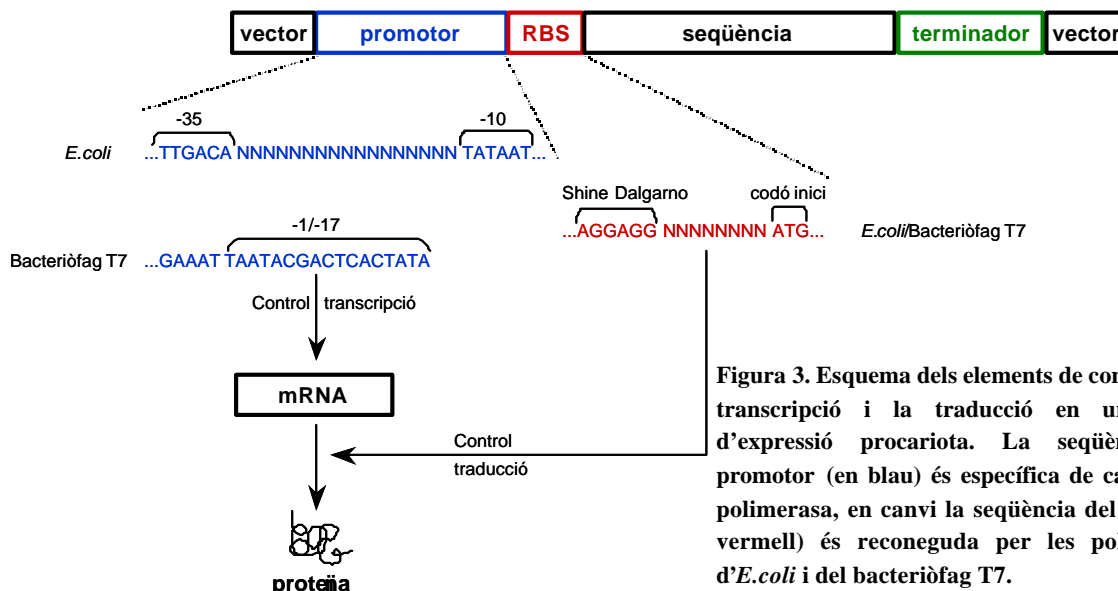


Figura 3. Esquema dels elements de control de la transcripció i la traducció en un vector d'expressió procariota. La seqüència del promotor (en blau) és específica de cada RNA polimerasa, en canvi la seqüència del RBS (en vermell) és reconeguda per les polimerases d'*E.coli* i del bacteriòfag T7.

El promotor consens reconegut per la RNA polimerasa d'*E.coli* està format per una regió rica en adenines i timidines situada a -10 bp i per un hexanucleòtid de seqüència específica situat a -35 bp de l'inici de la transcripció. Ambdues regions estan separades per un ADN separador d'uns 17 bp (Figura 3). El promotor consens reconegut per la RNA polimerasa del bacteriòfag T7 està format

per una seqüència de 17 bp (de -1 a -17) que està flanquejada per altres seqüències també definides, però que contribueixen poc en l'inici de la transcripció (Figura 3).

Les principals característiques d'un bon promotor són:

- Promotor fort (producció proteica entre un 10-30% del pes sec).
- Transcripció basal mínima (sobretot si la proteïna és tòxica). Regulable.
- Fàcilment induïble i a baix cost (inducció tèrmica/IPTG o altres inductors).

Per tal de reduir la transcripció basal s'usen sistemes regulats, un dels més usats és el basat en el promotor-operó d'utilització de lactosa (*lac*). En aquest cas s'ha observat que la localització de l'operó de 17 bp dins el promotor afecta el nivell de repressió. Si l'operó de 17 bp està situat entre les regions -10 i -35 la repressió és entre 50 i 70 cops superior que si se situa en direcció 5' de la regió -35 o en direcció 3' (*downstream*) de la regió -10 (Lanzer i Bujard, 1988).

Els promotors que s'han usat en aquesta tesi són:

- *lpplac* (vector pINIII OmpA3): està regulat pel repressor *lac* i s'indueix per addició d'IPTG.
- *T7 RNA polimerasa* (vector pBAT4) (Peranen i col., 1996): la T7 RNA polimerasa és altament específica pel seu promotor, el qual no és usat per la polimerasa d'*E.coli*. La seva eficiència d'inici de la transcripció és unes 5 vegades superior a la de la RNA polimerasa d'*E.coli*. Per a l'expressió recombinant del gen d'interès cal emprar una soca d'*E.coli* compatible (ex: BL21(DE3)), que contingui una còpia del gen de la T7 RNA polimerasa. En el cas de la soca BL21(DE3), el gen es troba sota el control del promotor *lacUV5* i la seva expressió està controlada pel repressor *lac* (codificat en el vector pBAT4), de manera que la seva expressió s'indueix per addició d'IPTG.

A la taula de la pàgina següent es mostra una llista dels promotors més usats per a la sobreexpressió de gens a *E.coli* (Taula 1).

II.B.1.2 Terminadors de la transcripció

Un terminador de la transcripció situat en direcció 3' de la seqüència codificant aborta la transcripció a través de la formació d'una estructura en forma de llaç en el transcrit. Això potencia l'estabilitat del mRNA i augmenta la seva vida-mitja, ja que el protegeix de la degradació exonucleòtica i evita la transcripció més enllà de l'extrem 3' del gen recombinant (Makrides, 1996).

II.B.1.3 Estabilitzadors dels transcrits

No s'han detectat elements universals per a la estabilització dels transcrits, però es coneixen algunes seqüències que afavoreixen aquesta estabilitat (Hannig i Makrides, 1998):

- Seqüències situades a la regió no traduïda (UTR) 5'. Per exemple la seqüència 5'UTR del transcrit *OmpA* augmenta la vida-mitja de molts mRNA's heteròlegs a *E.coli* (Chen i col., 1991).
- Llaços a la zona 3'UTR que protegeixen el mRNA de la degradació exonucleòtica.

Taula 1. Promotors emprats a *E. coli* per a la sobreexpressió proteica (Hannig i col., 1996).

Promotor	font	regulació	inducció
lac	<i>E. coli</i>	lacI, lacI ^q	IPTG
lacUV5	<i>E. coli</i>	lacI, lacI ^q	IPTG
tac (híbrid)	<i>E. coli</i>	lacI, lacI ^q	IPTG
trc (híbrid)	<i>E. coli</i>	lacI, lacI ^q	IPTG
P _{svn} (sintètic)	<i>E. coli</i>	lacI, lacI ^q	IPTG
trp	<i>E. coli</i>		falta de trp
araBAD	<i>E. coli</i>	araC	l-arabinosa
lpp ^a	<i>E. coli</i>		IPTG, lactosa
lpp-lac (híbrid)	<i>E. coli</i>	lacI	IPTG
phoA	<i>E. coli</i>	phoB (positiu) phoR (negatiu)	falta de fosfat
recA	<i>E. coli</i>	lexA	àcid nalidixic
proU	<i>E. coli</i>		osmolaritat
cst-1	<i>E. coli</i>		falta de glucosa
tetA	<i>E. coli</i>		tetraciclina
cadA	<i>E. coli</i>	cadR	pH
nar	<i>E. coli</i>	fnr	condicions anaeròbiques
p _r	λ	λ cIts857	tèrmic (canvi a 42°C)
cspA	<i>E. coli</i>		tèrmic (canvi a T ^a < 20°C)
T7	T7	λ cIts857	tèrmic
T7-lac operó	T7	lacI ^q	IPTG
T3-lac operó	T3	lacI ^q	IPTG
T5-lac operó	T5	lacI, lacI ^q	IPTG
T4 gene 32	T4		infecció per T4
nprM-lac operó	<i>Bacillus</i>	lacI ^q	IPTG
VHb	<i>Vitreoscilla</i>		oxigen

^a El promotor constitutiu lpp es va convertir a un promotor induïble mitjançant la inserció del promotor-operó lac UV5 a la regió downstream.

II.B.1.4 Elements que afecten la traducció

El RBS és l'element que controla l'eficiència d'inici de la traducció i està situat en posició immediatament 5' del transcrit (Figura 3). El RBS està format per la seqüència Shine-Dalgarno que possiblement interacciona amb la subunitat 16S del ribosoma i per un ADN separador (*spacer*) que conté de 4 a 8 nucleòtids. És preferible que l'espaiador estigui enriquit amb adenines i timidines per evitar la formació d'estructures secundàries, que aparentment redueixen l'eficiència de la traducció al impedir la interacció de la subunitat ribosomal 30S.

II.B.1.5 Terminadors de la traducció

La presència d'un codó de terminació és imprescindible per a la terminació de la traducció. *E. coli* presenta una preferència clara pel codó de terminació UAA, però l'eficiència de terminació augmenta si s'usa el tetranucleòtid UAAU (Poole i col., 1995).

II.B.1.6 Ús de codó

L'expressió de proteïnes heteròlogues que contenen un nombre important de codons que són usats en baixa freqüència per a *E.coli* condueix a una baixa eficiència de producció, per a solucionar el problema s'han seguit dues estratègies:

- Alterar a nivell d'ADN els codons poc freqüents per un d'ús ampli, sense modificar la seqüència aminoacídica.
- Coexpressió de la proteïna d'interès amb gens que codifiquen pels t-RNA's poc freqüents (Andrews i Asenjo, 1996; Brinkmann i col., 1989).

II.B.1.7 Lloc d'expressió

Per tal de decidir si l'expressió d'una proteïna es realitzarà al citoplasma, periplasma o a l'espai extracel·lular cal tenir en compte els avantatges i inconvenients de cada procediment i escollir, per a cada cas, el sistema més favorable.

Taula 2. Avantatges i inconvenients de l'expressió en *E.coli* en funció del lloc d'expressió.

	Avantatges	Inconvenients
Citoplasma	<ul style="list-style-type: none"> formació cossos d'inclusió -fàcil allament -protecció vers proteases -proteïna inactiva (no toxicitat) producció elevada 	<ul style="list-style-type: none"> formació cossos d'inclusió -cal replegament proteïna -disminució producció -elevat cost recuperació formació ponts disulfur desafavorida aminoàcid N terminal és metionina possible proteòlisi purificació més complexa
Periplasma	<ul style="list-style-type: none"> fàcil purificació poca proteòlisi millora del plegament aminoàcid N terminal adequat 	<ul style="list-style-type: none"> El pèptid senyal no sempre funciona Saturació maquinària exportació
Extracel·lular	<ul style="list-style-type: none"> fàcil purificació (respecte al nombre de proteïnes presents al medi) poca proteòlisi millora del plegament aminoàcid N terminal adequat 	<ul style="list-style-type: none"> purificació més complexa (respecte al volum a tractar) problemes de secreció, no s'obtenen grans quantitats

II.B.1.9 Proteòlisi

La degradació proteolítica és una de les causes principals de la pèrdua de producció proteica. S'ha observat que *E.coli* segueix la regla de l'N-terminal (Tobias i col., 1991). Un test realitzat amb la mateixa proteïna amb diferents aminoàcids N -terminals va mostrar que la presència de: Arg, Lys, Leu, Phe, Tyr i Trp, donen lloc a una vida mitja 300 vegades inferior a la resta.

També s'han seguit altres estratègies per a minimitzar la proteòlisi (Murby i col., 1996):

- Expressió a periplasma/medi extracel.lular.
- Soques deficientes en proteases.
- Creixement a baixa temperatura.
- Proteïnes de fusió/còpies en tàndem.
- Coexpressió amb chaperones moleculars.
- Mutagènesi per a canviar llocs de tall de proteases.

II.B.1.10 Proteïnes de fusió

L'expressió de proteïnes recombinants a *E.coli* en forma de proteïna de fusió ha estat una estratègia a bastament usada que permet disminuir la formació de cossos d'inclusió, millorar el plegament, minimitzar la proteòlisi o facilitar la purificació (Makrides, 1996; La Vallie i McCoy, 1995; La Vallie i col., 1993). Aquesta estratègia es pot emprar per a facilitar la purificació de la proteïna estudiada, per exemple fusionant-la a una cua d'histidines. Això ens permetrà separar la proteïna d'interès de la resta mitjançant una única etapa cromatogràfica, usant una cromatografia d'afinitat amb una columna amb ions Ni^{2+} immobilitzats, que aprofita l'afinitat que presenten les histidines per aquest metall.

II.B.2 Expressió en organismes eucariotes. *Pichia pastoris*

Els sistemes procariotes representen el sistema més senzill d'expressió d'una proteïna recombinant, però de vegades el sistema no és adequat per la proteïna; principalment si es tracta d'una proteïna gran (de diversos dominis), de plegament difícil, d'origen eucariota o quan requereix modificacions post-traduccionals. En aquests casos cal usar sistemes eucariotes, entre els més emprats trobem els dels llevats *Saccharomyces cerevisiae* i *Pichia pastoris*.

P. pastoris, encara que de genètica molt menys coneguda que *S. cerevisiae*, ha estat àmpliament usat per a produir quantitats elevades d'un gran nombre de proteïnes (Cereghino i Cregg, 2000; Cregg i col., 2000). És un llevat metilotròfic, fàcilment manipulable i generalment dóna expressions més elevades que altres sistemes eucariotes, tipus *S. cerevisiae*, baculovirus... Les estratègies anteriorment descrites per a millorar l'expressió proteica a *E.coli* no són exclusives d'aquest sistema, ni d'organismes procariotes; sinó que en molts casos són també aplicables a organismes eucariotes. En altres casos, la falta d'estudis exhaustius en la resta de microorganismes, com ara en *P.pastoris*, fa que encara no s'hagin determinat quines són les seqüències estabilitzadores del transcrit o els elements que afecten la traducció, tot i que la seva existència sembla clara.

L'ús de *P. pastoris* presenta diversos avantatges:

- Integració genòmica del gen d'interès -major estabilitat-.
- Fàcil inducció i baix cost -addició de metanol-.
- Creixement a elevades densitats òptiques (DO) en flascó -fins a 20 UDO-.
- Alts rendiments de producció proteica.

No obstant, hi ha una gran varietat de paràmetres que afecten el nivell d'expressió proteica en *P.pastoris*, com per exemple: el nombre de còpies del gen, el tipus de vector d'expressió, el fenotip d'utilització de metanol, etc. Això provoca una elevada variació clonal en l'expressió i fa necessària la realització d'un triatge ampli per tal de seleccionar el clon que presenta una producció més elevada. En el cas de proteïnes petites riques en ponts disulfur s'han descrit expressions de 1.54 g/l per la hirudina (Rosenfeld i col., 1996), 1.7 g/l pel TAP (Laroche i col., 1994) i 10 mg/l per la gilatina (Brankamp i col., 1995), usant en tots els casos un sistema d'expressió extracel·lular. Això suggereix que la secreció extracel·lular en *E.coli* és un bon sistema per a l'expressió d'aquest tipus de proteïnes.

II.C El procés de plegament de proteïnes

En aquests últims anys s'ha avançat molt en els estudis de plegament *in vitro* i *in silico* i això ha permès començar a comprendre el procés de plegament proteic *in vivo*, sobretot en el cas de les proteïnes monodomini. La comprensió del plegament proteic a nivell atòmic té una gran importància biològica, ja que hi ha un gran nombre de malalties causades per un mal-plegament proteic, agregació o inestabilitat, que poden ser produïts per mutacions o per canvis en l'entorn de la proteïna. L'estudi del desplegament proteic també presenta molta rellevància, ja que trobem molts exemples de processos cel·lulars on es produeix un desplegament; per exemple, en el pas a través de les membranes, en la degradació via proteasoma o en la amiloidosi.

II.C.1 Evolució històrica de la visió del plegament de proteïnes

Als principis dels anys 60 el grup de Christian B. Anfinsen va demostrar que la RNasaA podia plegar-se de manera reversible, la qual cosa implicava que les estructures natives de les proteïnes eren estats termodinàmicament estables i representaven un mínim d'energia lliure (Anfinsen i col., 1961). Uns anys més tard, a finals dels 60, Cyrus Levinthal va apuntar un nou argument que posteriorment seria conegut com la paradoxa de Levinthal, segons el qual el nombre de possibles conformacions que pot adoptar una proteïna és massa gran per a poder trobar l'agulla (l'estructura nativa) en el paller (l'espai conformacional) mitjançant una búsqueda a l'atzar (Figura 6a) i va concloure que les proteïnes s'havien de plegar a través de camins de plegament (Figura 6b) (Levinthal, 1968).

Dins aquesta visió del plegament calia tenir en compte dos aspectes que segons Levinthal estaven contraposats: El control termodinàmic, que representa la capacitat de la proteïna d'assolir el mínim energètic global i això s'aconsegueix independentment del camí de plegament; és a dir, que l'estructura nativa (o mínim energètic global) tan sols està determinat per les condicions finals (natives) i no per les condicions desnaturalitzants inicials (Figura 4). Això requeriria un temps molt llarg perquè implica una búsqueda extensa. El control cinètic representa la capacitat que tenen les proteïnes de plegar -se ràpidament (en una escala de temps biològica) i és dependent del camí de

plegament; és a dir, que l'estructura final difereix segons les condicions desnaturalitzants inicials i per tant, la proteïna només pot assolir mínims locals (Figura 4).

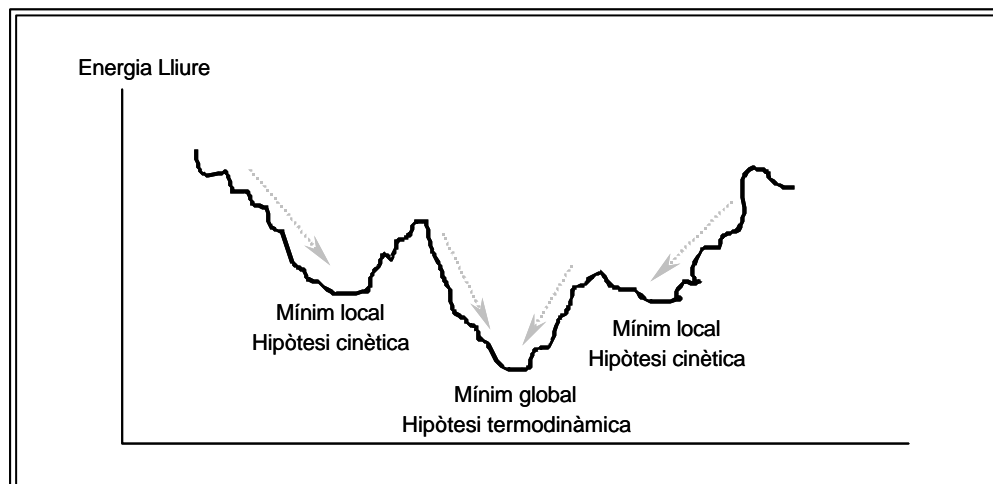


Figura 4. Representació esquemàtica de les hipòtesis cinètica i termodinàmica del plegament proteic en funció de l'energia lliure assolida per les formes finals.

Aquesta visió del plegament va conduir durant molts anys a una búsqueda de camins de plegament, centrant-se en models cinètics fenomenològics. La cerca d'intermediaris de plegament com a principals referents del camí de plegament d'una proteïna van conduir cap a una nova qüestió: Els intermediaris de plegament es troben dins-el-camí o bé fora-del-camí de plegament? (*on-pathway* versus *off-pathway*) (Tsong i col., 1971; Ikai i Tanford, 1971). Aquells intermediaris que es trobin dins-el-camí ens indicaran el camí de plegament de la proteïna, en canvi els altres són formes no productives.

En aquesta visió clàssica del plegament les dades que s'empraven provenien del seguiment dels canvis que s'observaven en diverses propietats òptiques de les proteïnes quan aquestes passaven de l'estat plegat al desplegat o viceversa. Quan les dades s'ajusten a una exponencial simple es parla d'una cinètica de dos estats, l'estat natiu (N) i l'estat desnaturalitzat (U); és a dir, que no s'observa la presència d'intermediaris de plegament. Si les dades s'ajusten a una exponencial múltiple, llavors s'assumeix que la proteïna presenta més de dos estats de plegament, i cada estat addicional correspon a un intermediari de plegament (I)(Figura 5).

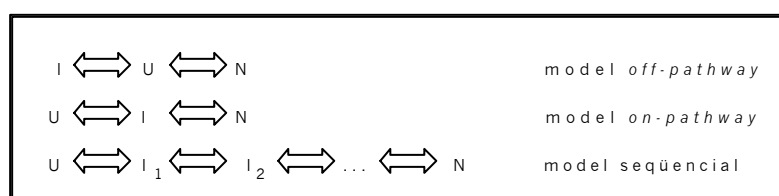


Figura 5. Els 3 models clàssics de plegament de proteïnes. U indica la molècula desplegada, N la molècula en estat natiu i I un intermediari de plegament.

En aquests models clàssics s'obté una visió de la mitjana del comportament d'una proteïna, però no del comportament a nivell microscòpic. La nova visió (*the new view*) del plegament de les proteïnes es basa en models estadístics i a diferència de la visió clàssica distingeix entre els estats macroscòpics i els microscòpics. Els estats macroscòpics són una distribució de conformacions individuals. No tenim una entitat única, sinó una barreja d'estructures. De fet, dos estats macroscòpics diferents poden tenir la majoria de les seves configuracions moleculars en comú.

Tot i que els nous models pretenguin descriure els detalls microscòpics de les conformacions individuals, les dades experimentals tan sols ens donen una idea de la mitjana global per sobre dels detalls. Així doncs, de la mateixa manera que en la visió clàssica, s'estan buscant maneres d'expressar els resultats dels models de plegament per etapes. Es reorganitzen aquells trets en comú que es poden identificar entre les trajectòries de plegament de les molècules individuals i, d'aquesta manera, es pot descriure quins esdeveniments dins el plegament d'una proteïna tenen lloc primer que uns altres.

Per tant, el concepte de camí de plegament a través d'esdeveniments seqüencials és substituït pel concepte d'embut de plegament a través d'esdeveniments paral·lels. Cada molècula individual segueix una determinada sèrie seqüencial de canvis conformacionals i finalment totes elles es retrobaran en l'estructura nativa. En aquesta visió d'embut de plegament l'eix vertical de l'embut representa l'energia lliure interna d'una determinada configuració d'una cadena (ponts d'hidrogen, interaccions electrostàtiques, energies dels angles de torsió...). Aquest valor depèn de les condicions externes; per tant, el valor de l'eix vertical està en funció de la temperatura i el solvent. Els múltiples eixos laterals representen coordenades conformacionals (per exemple els angles dihèdrics $\phi_1, \psi_1, \phi_2, \psi_2...$) (Figura 6).

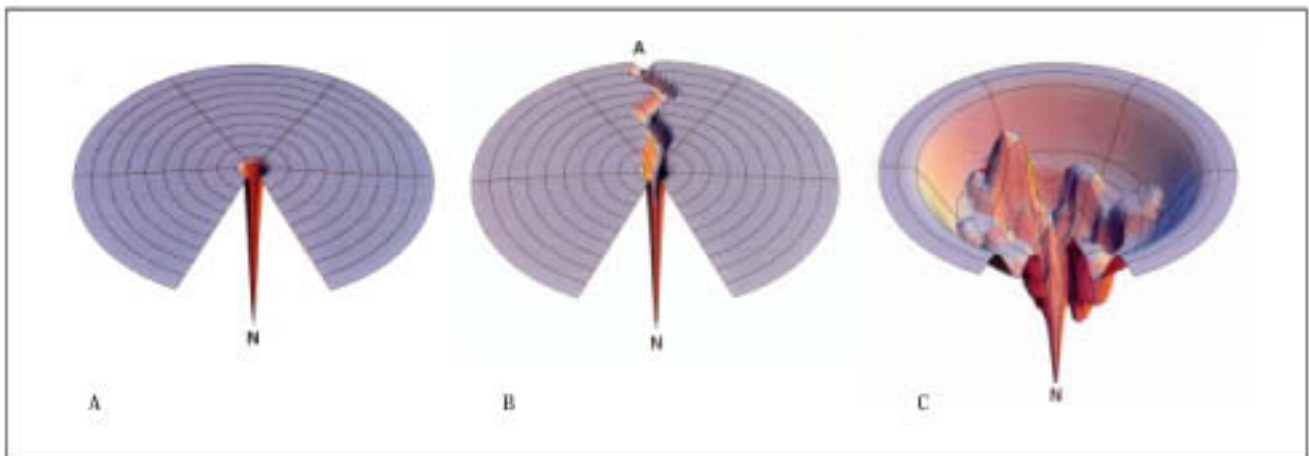


Figura 6. Embuts de plegament, l'eix vertical representa l'energia lliure interna i els eixos horitzontals representen les coordenades conformacionals. 6A: embut de plegament corresponent a la "paradoxa de Levinthal". 6B: embut de plegament corresponent al concepte de plegament. 6C: embut de plegament corresponent a un paisatge energètic abrupte amb trampes cinètiques i barreres energètiques.

També s'incorpora el terme paisatges d'energia (*energy landscapes*) que ens indica l'energia lliure de cada conformació en funció dels graus de llibertat. Cada conformació està representada per un

punt en la superfície d'energia multidimensional i aquelles conformacions geomètricament similars estan properes entre si en el paisatge d'energia. En aquests paisatges hi podem trobar muntanyes, que corresponen a conformacions d'alta energia i valls, que corresponen a configuracions energèticament més favorables que les del seu voltant.

La paradoxa de Levinthal, que es generà quan es va afrontar el problema del plegament de les proteïnes a través del concepte de camí de plegament, és eliminada en aquesta nova visió del plegament. En la nova visió, quan comença el procés de plegament, l'estat de plegament d'una proteïna no correspon a una única forma (U), sinó a tot el conjunt de conformacions existents en el paisatge energètic, cada una de les quals seguirà la seva pròpia trajectòria per arribar al mínim global.

II.C.2 Mètodes de seguiment del procés de plegament de proteïnes aplicats a les proteïnes riques en ponts disulfur

Les proteïnes monodomini han sigut les més estudiades per tal de dilucidar el procés de plegament a nivell atòmic. En aquestes proteïnes el plegament *in vitro* i *in vivo* és molt semblant, ja que el plegament no es produeix fins que els residus C-terminals no estan lliures per interaccionar amb la resta de la proteïna (Neira i Fersht, 1999 i Neira i col., 1997). Per tant, el plegament d'aquestes proteïnes no es produeix fins que han abandonat completament el canal del ribosoma i a causa del seu ràpid plegament i dèbil interacció amb les chaperones, es pot assumir que el seu plegament *in vivo* és molt similar al plegament *in vitro* (Corrales i Fersht, 1996).

En el cas concret de les proteïnes petites riques en ponts disulfur, la formació dels ponts disulfur dirigeix el plegament; per tant, caracteritzant les formes amb diferent nombre o aparellament de ponts que es formen al llarg del procés podem caracteritzar el plegament. Per tal d'estudiar a bastament el procés de plegament d'una proteïna, les formes que apareixen en el transcurs del procés cal que siguin aïllades i caracteritzades estructuralment. Idealment s'haurien de poder aïllar totes les formes detectables, fossin predominants o minoritàries, per tal d'analitzar tant les seves estructures com la seva evolució durant la resta del procés de plegament (experiments de plegament *stop/go*). Per tal d'aconseguir això, els diversos intermediaris cal que siguin capturats d'una manera reversible. Existeixen diversos mètodes per a caracteritzar aquests intermediaris: la captura en medi àcid fou un dels primers sistemes utilitzats, però la tècnica de bescanvi H/D seguida per RMN o per MS també dona uns resultats molt acurats.

II.C.2.1 Captura en medi àcid

La captura de disulfurs (*disulfide quenching*) és una tècnica que ens permet atrapar els intermediaris de plegament, separar-les i purificar-les (Creighton i col., 1978; Creighton i col., 1986). Inicialment aquesta tècnica s'aplicava d'una manera irreversible, ja que la captura de disulfurs es duia a terme mitjançant la derivatització química dels tiols lliures (alquilació).

Posteriorment es va desenvolupar la captura de disulfurs en medi àcid, que és un mètode reversible i ens permet realitzar experiments de plegament *stop/go* de les formes que s'han aïllat i purificat (Weissman i Kim, 1991; Konischi i Scheraga, 1982). Aquest mètode de captura de disulfurs ens permet aturar el procés de plegament d'una proteïna de manera senzilla, mitjançant l'acidificació de la mostra fins a pH:3-4 (Konischi i Sheraga, 1982). A pH àcid s'evita la formació, trencament o reorganització dels ponts disulfur i conseqüentment podem aturar el procés de plegament d'una manera ràpida i reversible. La mostra pot ser seguidament analitzada cromatogràficament (per exemple, per HPLC de fase reversa (RP-HPLC)) i d'aquesta manera es pot fer un estudi quantitatiu de les variants conformacionals presents a cada estadi del procés de plegament. Aquestes mostres separades cromatogràficament s'usen per a realitzar estudis estructurals (DC, RMN...) o bé, tornant a pujar el pH de la mostra, per seguir el procés de plegament de la forma aïllada (experiments *stop/go*).

El mètode de captura en medi àcid seguit per RP-HPLC s'ha utilitzat per estudiar un gran nombre de proteïnes i ha permès determinar el procés de plegament i/o desplegament del BPTI (Weissman i Kim, 1991), la hirudina (Chatrenet i Chang, 1993), el PCI (Chang i col., 1994), l'EGF (Chang, 1995), el TAP (Chang, 1996), i la RNasaA (Chang, 1999) entre altres.

II.C.2.2 Bescanvi H/D

La tècnica del bescanvi H/D s'ha usat durant molts anys per estudiar l'estructura i el plegament de proteïnes i pèptids (Englander i col., 1997). Aquesta tècnica es basa en la capacitat que tenen alguns hidrògens de les proteïnes de bescanviar amb el solvent que els envolta (Figura 7). Aquests

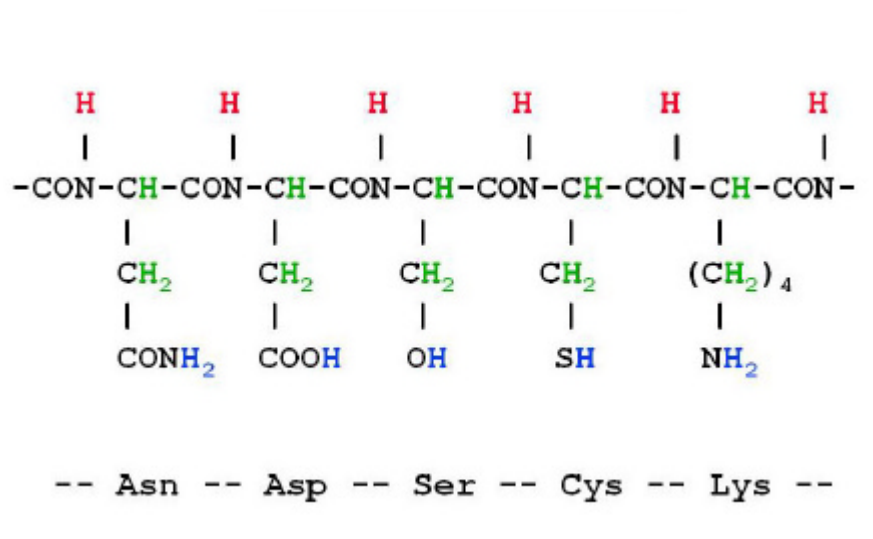


Figura 7. Esquema dels diferents tipus d'hidrògens presents en una proteïna. Els hidrògens units covalentment a un àtom de carboni (verd) no bescanvien, els que pertanyen a cadenes laterals (blau) bescanvien molt ràpid i normalment no es poden detectar i els hidrògens amida de la cadena principal (vermell) són els que bescanvien a una velocitat que pot ser mesurada. Aquesta velocitat variarà si els hidrògens estan involucrats en ponts d'hidrogen o no accessibles al solvent.

hidrògens làbils o bescanviables poden estar protegits vers el bescanvi per la conformació nativa de la proteïna, a causa de la seva participació en la formació de ponts d'hidrogen o bé perquè es troben soterrats a l'interior de la molècula i per tant, no estan accessibles al solvent (Englander i col., 1996).

El bescanvi d'aquests protons protegits requereix un procés de desplegament local o global de la proteïna que els exposi al solvent. D'aquesta manera es perturba l'entorn on es troba el protó làbil, provocant un canvi estructural en la proteïna i per tant, l'estudi d'aquesta cinètica de bescanvi ens permet obtenir informació conformational (Huyghues-Despointes i col., 1999).

La cinètica de bescanvi H/D depèn, en gran part, del pH del medi i presenta un mínim de bescanvi al voltant d'un pH de 2-3, mentre que la velocitat de la reacció augmenta tant a pH bàsic com a pH més àcid (Figura 8). Per tant, cal treballar sempre amb sistemes ben tamponats.

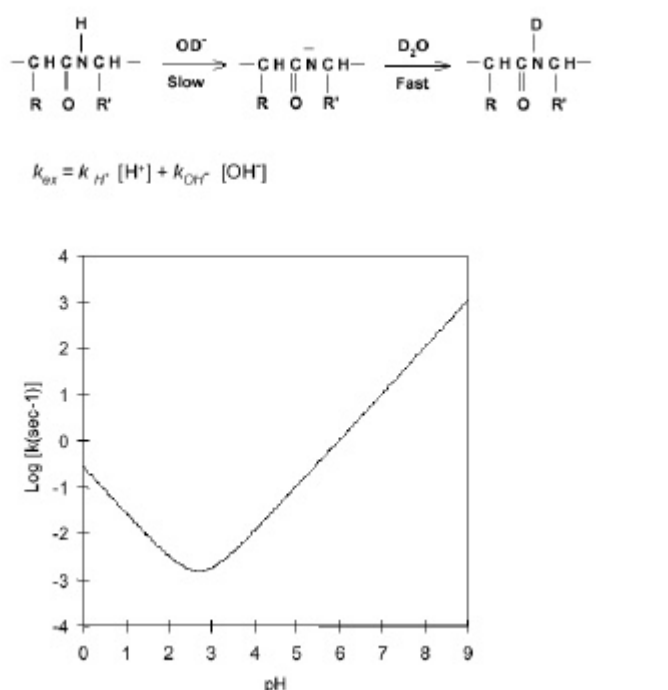


Figura 8. Cinètica del bescanvi H/D en funció del pH.

El bescanvi H/D s'ha monitoritzat mitjançant diverses tècniques experimentals (IR, UV, RMN, ESI MS i MALDI-TOF MS), entre les quals la RMN i l'espectrometria de masses són les que han tingut una aplicació més extensa.

II.C.2.2.1 Bescanvi D/H seguit per RMN

La ressonància magnètica nuclear (RMN) és una tècnica que ens permet obtenir dades estructurals detallades a nivell atòmic de proteïnes en solució. L'espectroscopia de RMN usa radiació de baixa freqüència, en el rang de la radiofreqüència, que al ser absorbida per les molècules produeix una reorientació de l'spin nuclear. Existeixen diversos nuclis susceptibles a aquest efecte, especialment els de ^1H i els de ^{13}C , que al ser exposats a un camp magnètic intens es poden alinear a favor (α) o

en contra (β) del camp magnètic. Aquesta transició entre α i β es pot donar mitjançant la radiació de radiofreqüència, donant lloc a una ressonància de protó i a un espectre d'absorció característic. La posició d'un determinat senyal a l'espectre, o desplaçament químic, depèn de l'ambient en què es troba el nucli d'hidrogen, produint-se una variació en el desplaçament químic en funció de la densitat electrònica circumdant. Quan es treballa amb macromolècules hi ha un gran solapament de senyals i cal usar mètodes bidimensionals per tal de mesurar les constants d'acoblament spin-spin. Els experiments COSY i TOCSY s'utilitzen per a identificar els sistemes d'spin, ja que ens permeten obtenir senyals seqüencials d'acoblament spin-spin separats fins a 3 enllaços (NH-C $^{\alpha}$ H, C $^{\alpha}$ H-C $^{\beta}$ H) i d'aquesta manera es poden assignar sistemes d'spin a determinats aminoàcids. Posteriorment cal comparar aquests resultats amb els dels experiments NOESY, que ens permeten obtenir senyals no seqüencials (NH-C $^{\beta}$ H, NH-C $^{\alpha}_{i-1}$ H) i d'aquesta manera podem determinar la seqüència aminoacídica de la nostra proteïna. Els espectres NOESY contenen un gran nombre de NOEs (Nuclear Overhauser Effects), que corresponen a distàncies entre protons de fins a 4Å i per tant, també trobem senyals entre aminoàcids no seqüencials, sinó propers en l'estructura tridimensional. Aquests NOEs seran de gran importància alhora de derivar un conjunt de possibles estructures tridimensionals compatibles amb les dades de partida.

De manera habitual, en el bescanvi H/D seguit per RMN, la proteïna és inicialment desplegada a en una solució D₂O-desnaturalitzant, de manera que tots els protons amida (NH) bescanviables es deuterin i llavors s'inicia el repliegament per dilució de la mostra. A diferents temps, alíquotes de la mostra són marcades amb un pols de 50 ms de H₂O a pH:9 (*pulse labeling*). En aquestes condicions la cinètica de bescanvi dels NH bescanviables que es troben lliures és molt alta (aproximadament 1 ms), en canvi els protons amida involucrats en ponts d'hidrogen o bé amb poca accessibilitat al solvent presenten una constant de bescanvi molt més lenta. D'aquesta manera, els NH de les zones no estructurades seran bescanviats, però no els involucrats en estructures secundàries o bé enterrats en el nucli hidrofòbic de la molècula. Després d'aquest pols, es canvia a unes condicions de bescanvi lent (pH àcid) on el repliegament pot continuar normalment, obtenint com a resultat una molècula nativa que porta imprès un patró de marcatge per H corresponent a les variants conformacionals presents en el moment en què ha tingut lloc el pols de H₂O. Així doncs, el bescanvi D/H ens permet seguir el plegament d'una proteïna a través de la identificació dels NH no protegits al llarg del procés (Roder i col, 1988; Udgaonkar i Baldwin, 1988).

La RMN també ens permet calcular la constant de bescanvi dels protons amida dissolent la proteïna en D₂O i seguint la pèrdua d'intensitat de senyal. De manera general, els protons amida es classifiquen en dos grups en funció de la seva velocitat de bescanvi: de bescanvi ràpid i de bescanvi lent. Els últims protons amida a bescanviar estan involucrats en la formació del nucli hidrofòbic i es coneixen com a nucli de bescanvi lent (*slow exchange core*). S'ha suggerit que aquest nucli està format pels elements d'estructura secundària que estan més empaquetats en l'estructura nativa de la proteïna (Kim i col., 1993; Woodward, 1993) .