

ENGINYERIA DE PROTEÏNES: COMPLEMENTARIETAT DE LES APROXIMACIONS EXPERIMENTALS I COMPUTACIONALS EN EL DISSENY DE PROTEÏNES

E. QUEROL, J. A. PÉREZ-PONS, V. VILLEGAS, F. CANALS, B. OLIVA, J. VENDRELL I F. X. AVILÉS

*Institut de Biologia Fonamental i Departament de Bioquímica i Biologia Molecular.
Universitat Autònoma de Barcelona. 08193 Bellaterra.*

SUMMARY

Protein engineering originated in the development of site directed mutagenesis and heterologous expression methods that allow to incorporate different amino acid residues into a protein, in a conservative or even non-conservative fashion. The redesigned or new proteins are used for structure/function analysis and for biotechnological purposes. It is a fascinating mixture of genetic engineering, protein chemistry, biophysics, biocomputation and several other biochemical and molecular biology approaches. Straddling so many disciplines, it nevertheless has a form and style of its own. In this chapter both the methodology and some selected examples are described.

RESUM

Els orígens de l'enginyeria de proteïnes es troben en el desenvolupament dels mètodes de mutagènesi dirigida i d'expressió heteròloga. Aquestes metodologies ens permeten incorporar diferents aminoàcids en una proteïna, de forma conservada o no conservada. La proteïna redissenjada o nova serveix per a estudis d'estructura/funció i per a objectius biotecnològics. Es tracta d'una mescla fascinant d'enginyeria genètica, química de proteïnes, biofísica, biocomputació i altres aproximacions bioquímiques i de biologia molecular. Tot i que hi ha tantes disciplines, l'enginyeria de proteïnes té una estructura i un estil propis. En aquest capítol es presenten la metodologia i alguns exemples seleccionats.

INTRODUCCIÓ

L'enginyeria de proteïnes (EP) és el conjunt de les tècniques que permeten el redisseny d'una proteïna actual o el disseny d'una proteïna nova a partir del seu gen mitjançant tècniques de DNA recombinant. Es podria incloure dins d'aquesta definició la construcció de proteïnes mitjançant síntesi química. L'enginyeria de proteïnes neix els anys vuitanta amb l'aparició de la tècnica de la mutagènesi dirigida per oligonucleòtid, tècnica creada per Zoller i Smith. A causa que el coneixement de les lleis que determinen com una seqüència d'aminoàcids determina una estructura tridimensional i una funció és insuficient, el procés de redisseny proteic habitualment es fa de forma iterativa, seguint l'anomenat «cicle de l'enginyeria de proteïnes». La figura 1 mostra aquest cicle de forma simplificada. L'investigador inicia el cicle amb l'aïllament d'un gen corresponent a la proteïna que vol redissenyar (o produir) i amb el màxim nombre possible de coneixements estructurals previs sobre la proteïna. Hi ha diferents nivells de complexitat i aplicació de l'enginyeria de proteïnes. En aquest capítol hem escollit tres casos (i tres estratègies) per exemplificar-los, d'acord amb el nivell de coneixements estructurals previs de què es parteix.

Un dels primers passos que s'han de fer és decidir quina regió i quins aminoàcids cal mutar i per quins altres s'han de mutar. Aquest és el primer punt feble que fa que aquest procés no sigui, en sentit estricte, una veritable enginyeria per motiu de l'abans esmentada insuficiència de coneixement. Precisament un dels principals objectius de l'enginyeria de proteïnes és millorar aquest coneixement. El que a la pràctica es fa és establir hipòtesis per provar d'acord amb una sèrie de criteris bàsics, alguns dels quals es descriuran més endavant.

Màxim coneixement de l'estructura de la proteïna, donatge i seqüenciació del gen

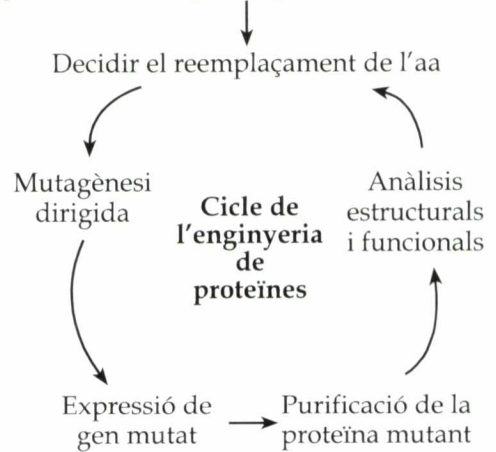


FIGURA 1. Cicle de l'enginyeria de proteïnes

vant. Una vegada decidit el mutant es crea mitjançant qualsevol dels procediments experimentals que es descriuen més endavant.

A continuació es tracta d'expressar, si es pot, el gen mutant o nou (normalment tot el procés s'intenta fer inicialment en *E. coli* pel coneixement del sistema i la facilitat de manipulació). S'aïlla i es purifica la proteïna mutant i es caracteritza (activitat biològica, conformació i camí del plegament, etc.). Si la proteïna no és expressada bé en *E. coli*, cal fer servir altres organismes hoste, com llevats (Reverter *et al.*, 1998), baculovirus o cultius de cèl·lules animals, entre d'altres (vegeu llibre Cleland i Craik a la bibliografia). A partir de les dades obtingudes sobre els mutants inicials es fan noves hipòtesis i es dissenyen nous mutants, i s'entra així de nou en el cicle. La gran quantitat de tècniques i especialitats necessàries (DNA recombinant, enzimologia, microbiologia, diverses espectroscòpies, etc.) fan de l'enginyeria de proteïnes una de les àrees de recerca més interdisciplinàries que existeixen.

Objectius de l'enginyeria de proteïnes

En EP és molt difícil de separar els aspectes de coneixement bàsic i els de coneixement aplicat. Entre els objectius bàsics habituals es poden considerar els següents: determinar el paper estructural o funcional d'un residu (o diversos), determinar la seva contribució energètica en interaccions, analitzar el procés i l'energètica del plegament, trobar els determinants de la funcionalitat, dissenyar variants funcionals, etc. Serien objectius aplicats, biotecnològics: estabilitzar una proteïna davant d'altres temperatures, o d'un medi no fisiològic, o de condicions oxidants; modificar el seu pH òptim, o l'activitat o l'especificitat; minimitzar estructures mantenint la seva funció, etc. Tanmateix, l'EP està molt relacionada amb el disseny de lligands a proteïnes i, per tant, amb el disseny de fàrmacs. Finalment, és obvi que un objectiu molt important, però encara futur, és el disseny de proteïnes completament noves amb estructures i funcions diferents de les actuals.

Alguns criteris o «regles» bàsics generals en EP

Al' hora de prendre una decisió sobre on i quin reemplaçament cal fer, l'investigador pot elegir entre dues estratègies: crear un mutant de pertorbació o un mutant de conservació. Això vol dir un mutant que alteri dràsticament o bé que conservi l'estructura o la funció de la proteïna. Com a regla general, al començament dels estudis amb una proteïna se solen fer mutants que conserven les estructures (p.e., substituir un residu per un altre amb propensió pel mateix tipus d'estructura secundària o d'exposició al dissolvent) i que pertorben la funció (p.e., substituint un residu per un altre de diferent caràcter químic). De totes maneres, una de les lliçons dels milers de mutants fets en proteïnes és que aquestes són molt més capaces

de suportar mutacions sense efectes massa deleteris del que abans es creia. Per exemple, s'han produït més de 1.700 diferents mutants artificials de l'enzim lisozim i la major part són proteïnes més o menys estables però expressables a la cèl·lula hoste (cosa que indicaria que es pleguen més o menys bé).

En tot cas, alguns dels criteris o «regles» bàsics generals que es poden donar per redissenyar una proteïna són els següents:

- En EP hi ha poques regles generals.
- Les proteïnes toleren més bé els reemplaçaments en els arcs (*loops*) que en altres estructures, tant les mutacions substitutòries com les insercions i delecions.
- No és convenient desestabilitzar les estructures secundàries regulars.

- Convé fer reemplaçaments isostèrics. No convé modificar el volum dels aminoàcids, especialment els del nucli hidrofòbic de la proteïna. Si s'ha de fer, millor fer ús d'un aminoàcid de volum més petit que a l'inrevés.

- Els canvis de càrrega o polaritat solen ser més pertorbadors i difícils d'acceptar que els que les mantenen.

- Les regions invariants (molt conservades) compartides per proteïnes homòlogues en un alineament són importants per a la conformació, el plegament o la funció. La seva modificació és delicada. Les regions més variables de l'alineament solen correspondre a regions externes, molt sovint arcs.

Alguns dels reemplaçaments més emprats, que trobem reiteradament a la literatura, són:

- Per valorar el paper de la càrrega negativa, per exemple la participació en processos àcid-base o en una interacció electrostàtica:

Asp ↔ Asn
Glu ↔ Gln

- Per determinar el paper d'un grup tiol:

Cys → Ser, Ala

Per determinar el paper d'un grup hidroxil o fenol:

Ser, Thr → Ala
Tyr → Phe

Per determinar el paper d'un grup imidazol:

His → Gln, Asn

Per analitzar el paper del grup R de qual-sevol residu:

aminoàcid → Ala

L'aminoàcid Ala es pot considerar un veritable comodí des del punt de vista estructural, ja que no sol pertorbar l'estructura secundària ni la polaritat de la regió de la proteïna en què s'introdueix. De fet existeix una estratègia que és l'escombrat d'alanines (*Alanine scanning*) per un rastreig sistemàtic, un per un, de tots els aminoàcids d'una regió d'una proteïna per valorar el seu paper estructural/funcional.

El procés pel qual un aminoàcid se substitueix pels altres dinou rep el nom de mutagènesi de saturació.

Finalment, com a criteri general, cal dir que hi ha casos en què el reemplaçament necessari pot ser identificable de forma fàcil (p.e., l'àcid general d'un centre actiu, si aquest centre es coneix), mentre que en d'altres és molt difícil de predir el lloc i per quin aminoàcid cal substituir-lo, sobretot quan hi ha contribucions múltiples (p.e., per termoestabilitzar una proteïna o per canviar la seva especificitat).

LA MUTAGÈNESI DIRIGIDA

La mutagènesi dirigida és la modificació del DNA per introducció d'una mutació es-

pecífica (un o més nucleòtids) en un punt concret de la seva seqüència. Existeixen diversos procediments per introduir mutacions selectives al DNA, entre els quals descriurem aquí tres dels més utilitzats (figura 2). En tots els casos, la mutagènesi dirigida requereix la síntesi prèvia d'un petit oligonucleòtid (generalment 15-20 nucleòtids) complementari de la seqüència on es vol introduir la mutació, però que presenta un desparellament, precisament a la base (o bases) que es vol canviar.

El primer procediment, que van descriure Zoller i Smith (com a revisió vegeu Smith, 1985), es basa en el clonatge del DNA que es vol mutar dins del bacteriòfag M13, que ens permet tenir el DNA en forma de monocadena o de doble cadena (forma replicativa). Aquest DNA monocadena s'hibrida amb l'oligonucleòtid mutagènic, que actua com a encebador de la DNA polimerasa Klenow. Amb una DNA lligasa, tal com la T4 DNA lligasa, lligarem l'extrem 3' del DNA sintetitzat amb l'extrem 5' de l'oligonucleòtid, i obtindrem un DNA de doble cadena circular, covalent i tancat (figura 2A). La introducció d'aquest DNA a l'interior d'*E. coli* originarà dues poblacions de bacteris, una infectada amb el genoma del bacteriòfag que conté la seqüència silvestre i una altra infectada amb el genoma del bacteriòfag que conté la seqüència mutada. Encara que ambdues seqüències es poden distingir per hibridació amb el mateix oligonucleòtid que havia estat utilitzat per a la mutagènesi (marcat radioactivament) i per la diferent temperatura de fusió (T_m), en general aquesta distinció es fa per seqüenciació de diversos clons. El lligament del fragment mutat a un vector d'expressió i la posterior introducció en un hoste adient permeten obtenir la proteïna mutant.

El rendiment en formes mutants del mètode bàsic és únicament del 0,5% dels clons. A causa d'això, a partir d'aquest mètode ini-

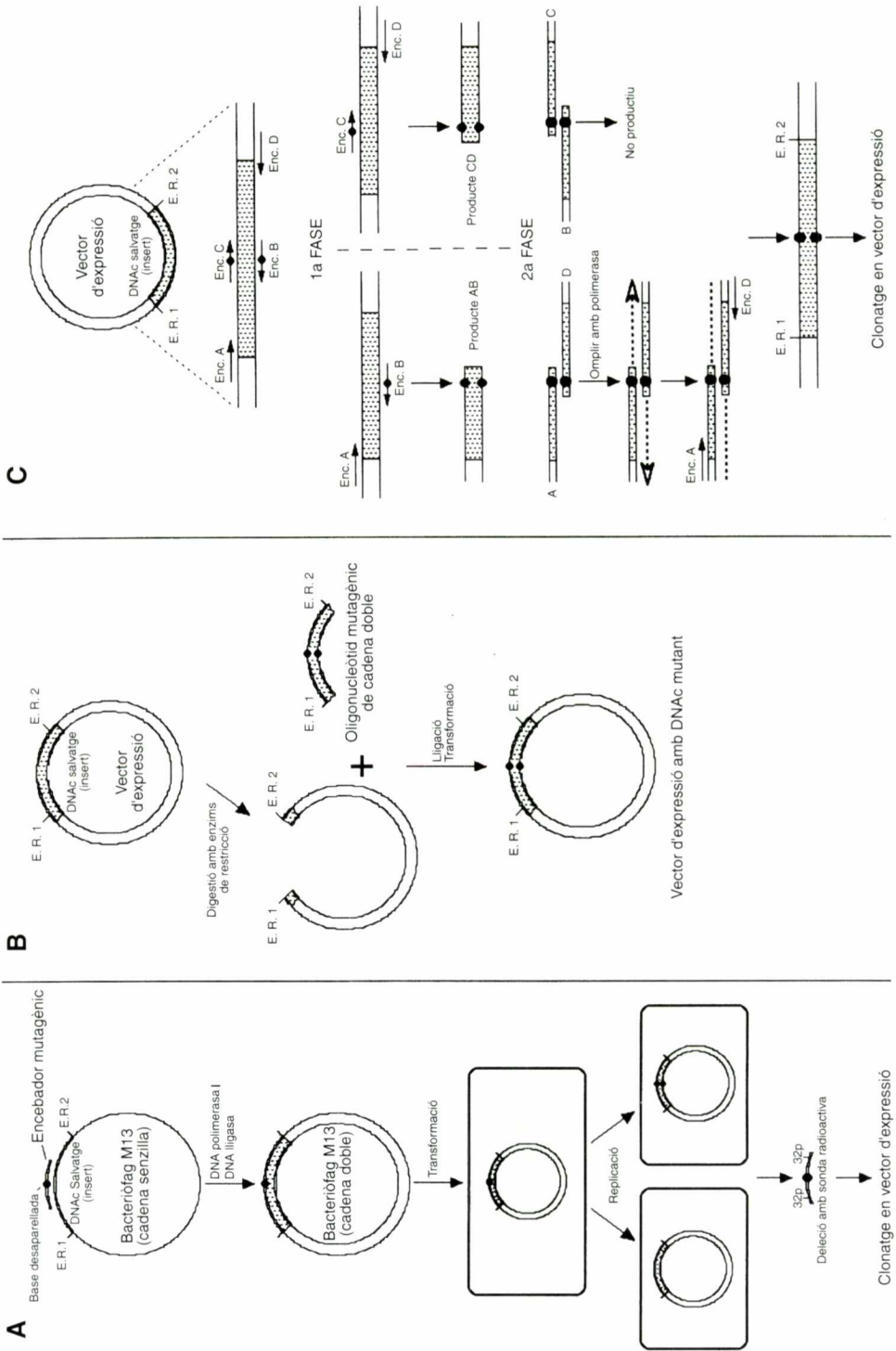


FIGURA 2. Tècniques de mutagènesi dirigida: la inicial, per *casset* i per PCR. Abreviatures: Enc. encebador; E. R., enzim de restricció; DNAC, DNA complementari.

cial s'han desenvolupat tota una sèrie d'estratègies per eliminar la cadena de DNA motlle, i s'ha evitat així la necessitat de seleccionar els clons mutants. Un bon exemple és el mètode de Nakayame i Eckstein (1986), en el qual la síntesi de la cadena complementària es realitza mitjançant una barreja de desoxiribonucleòtids, dels quals un (per exemple dCTP) porta un grup tiol inclòs. La presència d'una cadena de DNA amb grups tiol fa que molts enzims de restricció (per exemple, *Ava* I, *Ava* II, *Ban* II, *Hind* II, *Nci* I i *Pvu* I) no la reconeixin i només siguin capaços de tallar la cadena no derivatitzada que havia actuat com a motlle. Si a continuació afegim exonucleasa III, el motlle serà degradat i la cadena de DNA mutagènica restarà intacta. Aquesta cadena es pot convertir en doble cadena per addició de l'encebador mutagènic, de DNA polimerasa I i de DNA lligasa. Amb aquest mètode s'obtenen rendiments en mutants del 90% o més. Existeixen altres mètodes equivalents (p.e., el de Kunkel, el de les dianes *EcoA*/*EcoB* alternatives, etc.), que per limitacions d'espai no descriurem.

També es poden introduir mutacions específiques mitjançant el mètode de la casset. Si el nucleòtid que cal mutar està flanquejat per dianes de restricció úniques, es pot extreure la seqüència salvatge amb enzims de restricció i col·locar en el seu lloc, com si d'una casset es tractés, una seqüència mutant sintètica (fig 2B). Aquesta metodologia té l'avantatge que es pot realitzar directament sobre el vector d'expressió, però presenta els dos problemes següents: 1) la necessitat de dianes de restricció úniques flanquejant la zona que s'ha de mutar; i 2) els oligonucleòtids sintetitzats no poden superar unes desenes de nucleòtids per raons de fidelitat en la síntesi.

Finalment, la tècnica de la PCR (vegeu Mullis *et al.*, 1994) és actualment la més utilitzada per a la mutagènesi dirigida. En una

PCR, els oligonucleòtids actuen com a encebadors de l'amplificació d'un DNA motlle i generen un producte que en cicles posteriors actua, juntament amb el motlle, de seqüència complementària a partir de la qual es genera més producte. D'aquesta manera, es produeix una amplificació de 2^n molècules per molècula motlle inicial, on «n» és el nombre de cicles de la reacció. És a dir, a partir d'una única molècula se'n produeixen 3,5 milions en 30 cicles. Si l'oligonucleòtid incorpora una mutació, aquesta serà introduïda als productes de la reacció i, per tant, el producte final incorporarà la mutació. Si els encebadors dels extrems es dissenyen amb les seqüències de les dianes de restricció adients situades al seu extrem 5', el producte de la PCR es pot clonar directament sobre el vector d'expressió. Quan la mutació que s'ha de realitzar es troba en un dels extrems del DNA, un disseny adient dels encebadors dels extrems permet aconseguir el producte mutat amb una sola reacció d'amplificació (fig 2C). Si la mutació que s'ha de realitzar es troba al mig de la seqüència, és necessària l'amplificació per fases. En una primera fase, s'amplifiquen les dues meitats de la seqüència en dues reaccions separades, tenint en compte que han de contenir una seqüència encoberta, de tal manera que aquestes dues meitats puguin actuar com a motlle en una segona fase d'amplificació que originarà el producte complet de la reacció amb la mutació desitjada, utilitzant els encebadors dels extrems. Una variació d'aquesta metodologia és l'anomenat mètode del *megaencebador*, que implica una única reacció d'amplificació a la primera fase on s'incorpora la mutació desitjada i utilitza aquest producte com a encebador de la segona fase, juntament amb l'encebador de l'extrem oposat. Sigui quina sigui l'estratègia de la PCR mutagènica, és molt important la polimerasa utilitzada. La *Taq* polimerasa no mostra activitat correcto-

ra 3' → 5', que si bé no és crucial per a moltes aplicacions de la PCR, sí que ho és per a la incorporació d'una mutació específica, donada la importància de la fidelitat de la còpia. Algunes polimerases termofíliques tenen activitat correctora i originen productes amb una fidelitat de còpia deu vegades més gran. Entre d'altres, es pot citar la DNA polimerasa GB-D de les espècies de *Pyrococcus*, també anomenada «Deep Vent™ DNA polymerasa». Amb la *Vent* polimerasa s'ha demostrat una eficiència del 98-100% en la construcció d'una bateria de mutants (Junco *et al.*, 1994).

La mutagènesi dirigida no solament és útil per introduir mutacions puntuals, sinó que també permet la introducció de mutacions múltiples, d'insercions, la deleció d'un fragment determinat i, fins i tot, l'intercanvi de fragments dins d'una mateixa seqüència. Encara que la potencialitat de la tècnica és molt diversa, cada experiment ha de ser dissenyat racionalment i requereix unes condicions determinades. És important remarcar que cal seqüenciar el DNA dels clons transformants per assegurar la introducció de la mutació desitjada i la fidelitat de les còpies.

APLICACIONS COMPUTACIONALS A L'ENGINYERIA DE PROTEÏNES

L'enginyeria de proteïnes no seria possible sense l'ajut de la biocomputació. Les principals aplicacions d'aquesta tècnica en l'enginyeria de proteïnes són les següents:

Anàlisi de proteïnes amb estructura tridimensional coneguda

Una de les principals utilitats de la computació a l'enginyeria de proteïnes és la representació tridimensional d'una pro-

teïna mitjançant algun dels programes existents (FRODO, RASMOL, SIBYL, BIOSYM, PREPI, etc.). En particular, són útils per caracteritzar els entorns dels àtoms i residus, discernir les seves interaccions, o intuir la seva funcionalitat dins de la proteïna. Això permet suggerir vies per al seu redisseny, possibles mutacions, i observar com aquestes podrien afectar el conjunt d'una proteïna (estructura i funcionalitat).

Estudi de la seqüència d'una proteïna

En els casos en què tan sols tenim la seqüència d'una proteïna (obtinguda per determinació directa o bé per clonatge i seqüenciació del DNA, etc.) i no l'estructura 3D, també es poden fer diverses anàlisis:

Homologia seqüencial

Quan s'ha trobat un nou gen/proteïna i s'ha determinat la seva seqüència, el mètode més important per predir la seva funció potencial és la recerca de seqüències homòlogues amb funció coneguda a les bases de dades de seqüències (GenBank, EMBL, SwissProt, PIR). Els programes més utilitzats per esbrinar les seqüències homòlogues són el FASTA (Pearson i Lipman, 1988) i BLAST (Altschul *et al.*, 1990), que es poden usar a molts servidors a través de la xarxa d'Internet; un dels millors és a l'EBI (European Bioinformatics Institute: <http://www.embl-heidelberg.de>). El problema de la recerca d'homòlegs implica el de l'alineament de seqüències, que generalment es fa mitjançant un algoritme dinàmic (Needelman i Wunch, 1970) on s'alineen el màxim nombre possible de residus segons les identitats o les seves probables substitucions (Dayhoff, 1978; Henikoff i Henikoff, 1992), i s'afegeixen forats, produïts per insercions o delecions (amb les corresponents

penalitzacions), a una seqüència o a una altra. Els alineaments múltiples donen molta informació útil; per exemple, permeten identificar les regions estructuralment conservades, les quals solen correspondre a estructures secundàries regulars o a «motius» (patrons relacionats amb funció). Les regions més variables generalment corresponen a arcs. Els aminoàcids individuals conservats a totes les proteïnes solen ser residus importants per a la funció (catàlisi, etc.).

Predicció d'estructura secundària i d'hèlices transmembrana

Les prediccions d'estructura secundària de proteïnes globulars tenen per objectiu identificar les regions d'una seqüència de proteïna amb conformació en α -hèlix, estructura β , gir invers i desordenada. Els primers mètodes estadístics (Chou i Fasman, GOR, etc.) no han arribat a superar percentatges del 60-65% d'exactitud en la predicció. Actualment es fan servir algorismes basats en xarxes neuronals (Rost i Sander, 1993) o algorismes d'aprenentatge (King *et al.*, 1997) que arriben al 70-75% d'encert i són utilitzables a través de la xarxa d'Internet (<http://www.embl-heidelberg.de> i <http://www.icnet.uk>).

La predicció de l'estructura de proteïnes de membrana exigeix aproximacions diferents. Si bé són poques les proteïnes de membrana de les quals es coneix l'estructura tridimensional, el tipus de plegament del domini transmembrana pot ser de dos tipus: 1) d'hèlix transmembrana (p.e. bacteriorodopsina); i 2) de β -transmembrana (p.e. porines) insertades en la bicapa lipídica. Els mètodes utilitzats per a la predicció de les hèlices fan ús de xarxes neuronals (Rost i Sander, 1995; Aloy *et al.*, 1997), i obtenen percentatges del 90-95% d'encert.

Representació dels perfils d'hidrofobicitat/hidrofilicitat

Una de les anàlisis més senzilles que es poden fer sobre una seqüència proteica és la predicció de l'accessibilitat al dissolvent dels residus al llarg de la cadena polipeptídica basada en la predisposició hidrofòbica/hidrofíllica de cada residu. Això permet localitzar aquelles regions amb major probabilitat de ser enterrades dins del nucli de la proteïna o bé dins la doble capa lipídica de la membrana. Si bé els valors assignats a la hidrofobicitat/hidrofilicitat dels residus varien segons els mètodes, tots ells donen prediccions qualitativament similars. D'entre els mètodes més utilitzats, es poden esmentar els de Kyte i Doolittle, 1982, i Hopp i Woods, 1981.

Modelat de l'estructura tridimensional d'una proteïna

Modelat per homologia

El coneixement de l'estructura tridimensional és, com ja s'ha esmentat, de gran importància en el redisseny de proteïnes per al plantejament de possibles mutants. El principal problema és l'obtenció de l'estructura tridimensional detallada d'una proteïna de la mateixa família. Com que proteïnes homòlogues (amb percentatges d'identitat superiors al 30%) conserven tota o una bona part de l'estructura tridimensional, principalment les estructures secundàries, això ens permet proposar un model de l'estructura tridimensional d'una seqüència a partir del patró conformacional d'una o diverses proteïnes homòlogues (modelat per homologia). Les claus principals per al modelat per homologia són dues: 1) trobar els alineaments correctes entre la seqüència problema i les seqüències patró; i 2) construir un bon model per a aquelles regions on la seqüèn-

cia no és conservada. El model de l'estructura tridimensional es construeix per fragments, dividint la proteïna en regions estructuralment conservades (REC) i regions variables (RV). La construcció dels fragments REC es fa per substitució de la seqüència problema en l'estructura de la seqüència patró, per exemple amb el programa COMPOSER (Topham *et al.*, 1991). Una forma alternativa per construir aquests fragments és mitjançant l'aplicació de restriccions de distàncies definides a partir de les conformacions de les proteïnes patró, per exemple amb el programa MODELLER (Sali i Blundell, 1993).

Hom sol seguir tres mètodes per a la predicció de la conformació de la cadena principal dels fragments RV: 1) fent servir algorismes de construcció de cadenes polipeptídiques amb posicions restringides als extrems N i C terminals (*ring closure*); 2) cercant regions de conformació coneguda amb el mateix nombre de residus on les posicions en els extrems siguin similars a les desitjades per unir els fragments REC (Jones i Thirup, 1986); i 3) classificant el conjunt d'arcs d'estructura coneguda (Oliva *et al.*, 1997), determinant els residus conformacionalment importants per a l'estructura de l'arc, i cercant la classe a què pertany el fragment RV (Rufino *et al.*, 1997).

Algoritmes alternatius per al modelat: enfilat

Quan no existeixen proteïnes homòlogues amb estructura tridimensional coneguda, es fa servir el mètode de *threading* o enfilat (Jones *et al.*, 1992). Aquest mètode es basa en el fet que els tipus de plegament es repeteixen en proteïnes no homòlogues, postulant que la proteïna problema pot tenir un determinat plegament independentment de la seva homologia. El modelat es fa per substitució de la seqüència problema en

les estructures conegudes, determinant el plegament més adient. Aquest plegament s'obté sobre la base de potencials estadístics, generats a partir de bases de dades tridimensionals (PDB), escollint el plegament que mostra l'energia òptima d'entre tots els provats (Sippl i Weitckus, 1992).

Estudi de les característiques fisicoquímiques d'una proteïna

Mètodes de simulació. Dinàmica molecular

Un dels principals atractius de la biocomputació per a l'enginyeria de proteïnes es basa en la seva capacitat per predir l'efecte de mutacions. Per aquest motiu, és bàsic l'estudi de l'espai conformacional de la forma nativa de la proteïna i dels seus mutants. Aquest estudi permet predir les variacions en flexibilitat, estabilització i reactivitat del sistema. Per estudiar aquest espai es fan servir els mètodes de simulació i, concretament, les simulacions per mecànica/dinàmica molecular són les que ens poden donar més informació del sistema en les regions properes al seu estat natiu. La dinàmica molecular es basa en l'aplicació directa de les lleis de Newton a un sistema multidimensionat. D'aquesta manera, l'equació $F = -m \cdot \nabla E$, on E és l'energia potencial, resol les equacions de l'espai i de la velocitat en funció del temps per a cadascun dels àtoms constituents del sistema (proteïna més dissolvent i contracions). Com a resultat s'obtenen les posicions i les velocitats de cada àtom del sistema en les condicions requerides, generalment les del sistema isotèrmic-isobàric, que defineixen el conjunt estadístic del qual es poden extreure les propietats termodinàmiques (Oliva *et al.*, 1995, Daura *et al.*, 1996).

En particular, les propietats més interessants per analitzar són les variacions conformacionals (plasticitat) i la seva capacitat de retorn (flexibilitat) que permeten identificar

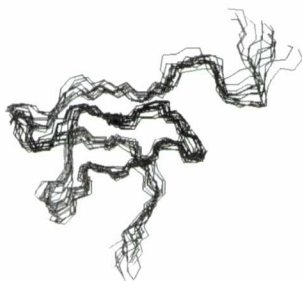


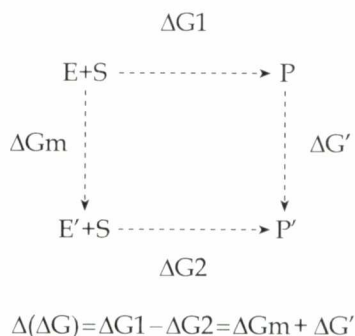
FIGURA 3. Anàlisi de la dinàmica molecular (DM) d'una proteïna de l'inhibidor de carboxipeptidasa de patata (PCI). Es representen superposades les diferents conformacions de les cadenes de carboni alfa del PCI al llarg d'una simulació en presència d'aigua (vegeu l'exemple III d'aquest capítol per a més detalls).

els punts més febles i els més estables dins de la proteïna. Això pot ser igualment utilitzat per comprovar la idoneïtat d'un model d'estructura quan no ha estat possible obtenir dades experimentals de la proteïna (o d'una regió d'aquesta proteïna). L'anàlisi de les variacions conformacionals es fa mitjançant el càlcul del RMSD (desviació estàndard) de les posicions atòmiques amb relació al seu estat original (vegeu figura 3), i al de la flexibilitat del sistema mitjançant les fluctuacions (factors B de temperatura o RMSD respecte a les posicions atòmiques mitjanes) (Oliva *et al.*, 1995).

Variacions d'energia lliure. Pertorbacions

La forma més natural de diferenciar l'estabilitat d'una forma mutant en relació amb la forma nativa és mitjançant el càlcul de la variació d'energia lliure (ΔG), la qual es pot predir mitjançant simulacions de dinàmica molecular amb pertorbacions. Aquestes simulacions consisteixen a canviar les equacions que descriuen l'energia del sistema

(hamiltonià) per una combinació de les equacions que descriuen les del sistema final i inicial. D'aquesta manera, per conèixer l'efecte d'una mutació es fa ús d'un cicle termodinàmic, on coneixem experimentalment el resultat per a la forma nativa i obtenim la diferència d'energia lliure respecte d'aquesta per a la forma mutada, fet que ens permet determinar si el mutant és més o menys estable o reactiu:



Aquestes simulacions presenten un gran cost computacional i només poden ser aplicades quan l'espai conformacional de la forma mutada és prou similar al de la forma nativa (això no és factible en casos de desplegament local o total de la proteïna) (Mark i Van Gunsteren, 1994).

Reactivitat química

L'aproximació teòrica més acurada per estudiar el mecanisme enzimàtic d'una proteïna és l'estudi de la seva reactivitat mitjançant mecànica quàntica. Fer aquest càlcul amb un sistema proteic complet és una tasca avui en dia computacionalment inviable. És per aquest motiu que només la part corresponent al lloc actiu de la proteïna i la part reactiva del o dels substrats és tractada quànticament, mentre que la resta del sistema és tractada per mecànica/dinàmica molecular. La combinació dels dos mètodes constitueix avui en dia la forma més innovadora per

atacar aquest problema (tècnica QM/MM) (Liu *et al.*, 1996). D'aquesta forma es poden identificar els diferents passos i els elements que intervenen en els mecanismes de reacció d'un enzim i proposar possibles mutacions que canviïn la seva reactivitat (Álvarez-Santos *et al.*, 1996).

Encaix proteïna/proteïna i proteïna/l·ligand

El problema de conèixer la conformació de complexos proteïna/proteïna i proteïna/l·ligand a partir de les conformacions de les estructures no complexades, també conegut com a encaix o *docking*, constitueix un repte de la biocomputació. La via més òbvia per resoldre aquest problema implica provar totes les possibles formes de disposar els elements que cal encaixar entre si, tenint en compte al mateix temps la deformació que es produeix en el contacte. Això fa que avui en dia sigui pràcticament impossible resoldre aquest problema de forma rigorosa per motiu de l'alt cost computacional, de manera que s'intenta resoldre el problema amb aproximacions, com per exemple sense considerar la deformació produïda en l'encaix. En aquest sentit, els algorismes que han donat els millors resultats s'han basat en la recerca de l'àrea complementària amb les millors interaccions electrostàtiques (Gabb *et al.*, 1997). Cal remarcar que aquesta àrea té un gran interès biotecnològic en el disseny de fàrmacs.

DIFERENTS NIVELLS D'APLICACIÓ DE L'ENGINYERIA DE PROTEÏNES

Exemples escollits

Les estratègies emprades per caracteritzar una proteïna mitjançant EP, i per redissenyar-la, s'han d'adaptar als nivells de coneixements estructurals i funcionals de

què es disposa sobre la proteïna. A continuació descriurem tres tipus d'investigacions realitzades pels nostres grups de recerca, amb objectius específics diferents i que han partit d'un nivell de coneixement del sistema molt diferent, fet que condiciona la velocitat d'avanç en la recerca portada a terme.

EXEMPLE I. Es coneix únicament la seqüència de la proteïna: glicosidases

Les glicosidases són enzims que catalitzen la hidròlisi estereoespecífica d'enllaços glicosídics en oligosacàrids i polisacàrids. Es tracta d'enzims amb gran aplicabilitat industrial i agroalimentària. Actualment hi ha dipositades en els bancs de dades més de 900 seqüències de glicosidases que han estat agrupades per Bairoch i Henrissat en més de seixanta famílies mitjançant l'anàlisi de *clusters* hidrofòbics (HCA) i el multialineament de seqüències, especialment aquelles relacionades amb el domini catalític (<http://expasy.hcuge.ch/cgi-bin/lists.glycosid.txt>). L'HCA permet examinar l'entorn de cada aminoàcid al llarg de tota la seqüència i ha demostrat ser particularment útil per establir relacions entre famílies que haurien passat desapercubudes amb la sola comparació entre seqüències (Henrissat, 1991). Des del punt de vista estructural, actualment es coneix l'estructura tridimensional d'almenys un representant de vint-i-tres famílies de glicosidases. Quant al plegament, les glicosidases resoltes per raigs X mostren topologies del tipus barril ($(\beta/\alpha)_s$), barril β , barril α/α i sandvitx β . Només s'han observat tres topologies de centre actiu: butxaca (*pocket*), escletxa (*cleft*) i túnel (*tunnel*). La funció bàsica de cada enzim, la hidròlisi de l'enllaç glicosídic, està determinada per la topologia del centre actiu, mentre que el posicionament de les cadenes laterals dels aminoàcids dins del centre actiu determina tant l'estereoquímica com

l'especificitat de substrat i d'enllaç. En conseqüència, el fet que dos enzims pertanyin a la mateixa família no implica necessàriament que duguin a terme la mateixa funció.

Pel que fa al seu mecanisme, les glicosidases es divideixen en dos tipus bàsics: a) enzims que utilitzen un mecanisme de doble desplaçament per catalitzar la hidròlisi de l'enllaç glicosídic amb retenció de configuració en el carboni anomèric; i b) enzims que utilitzen un mecanisme de desplaçament únic que dona lloc a la inversió de configuració en el centre anomèric (McCarter i Withers, 1994) (figura 4). En ambdós mecanismes, hi participen un parell d'àcids carboxílics. En els enzims amb retenció, un carboxilat amb càrrega actua com a nucleòfil i un grup carboxílic protonat com a catalitzador àcid-base. La reacció inclou la formació d'un intermedi covalent enzim-substrat per atac del nucleòfil sobre el carboni anomèric (etapa de glicosidació) i la posterior formació d'un nou extrem reductor per atac sobre el carboni anomèric d'una molècula d'aigua, la nucleofília de la qual està augmentada per la base conjugada de l'àcid general

(etapa de desglicosidació). En els enzims amb inversió, un grup carboxílic actua com a àcid i l'altre com a base que augmenta la nucleofília de la molècula d'aigua que desplaça el grup aglicó sortint.

Estem estudiant dues glicosidases: una 1,3-1,4- β -glucanasa i una β -glucosidasa. El treball dut a terme amb aquests enzims és un bon exemple de com, a partir de la seqüència i combinant mètodes predictius i anàlisi d'homologia, es poden formular hipòtesis que, un cop contrastades experimentalment, permetin esbrinar la relació estructura/funció i, fins i tot, tractar de millorar-ne les propietats enzimàtiques.

1,3-1,4- β -Glucanasa (EC 3.2.1.73)

La 1,3-1,4- β -glucanasa (endo-1,3-1,4- β -D-glucan 4-glucanohirolasa) hidrolitza enllaços β -1,4 dels β -glucans, polisacàrids lineals formats per residus de glucosa units per enllaços mixtos β -1,3 i β -1,4. El gen de la glucanasa de *Bacillus licheniformis* es va clonar en *E. coli* (Lloberas *et al.*, 1991). En absència d'informació sobre l'estructura tridimensional, els intents inicials per identificar els residus essencials del centre actiu es van basar en la comparació de la seqüència amb les corresponents a lisozims eucariòtics i virals, més conegudes des del punt de vista mecànic. Així es van predir dos possibles setins actius: 1) un segment de vint-i-cinc aminoàcids que presentava un 70% de similitud (40% d'identitat) amb el setí actiu del lisozim del fag T4 i que contenia els residus Glu i Asp en les mateixes posicions que els assignats com l'àcid general (Glu-11) i el nucleòfil (Asp-20) a l'enzim de T4; i 2) un segment de vint-i-un aminoàcids amb un 20% de similitud amb la regió que inclou els residus catalítics Glu-35 i Asp-52 en el lisozim de clara d'ou. D'una banda, aquest segon possible setí actiu no estava conservat en 1,3-1,4- β -glucanases d'altres espècies de *Bacillus* i, de l'altra, la predicció d'estructura

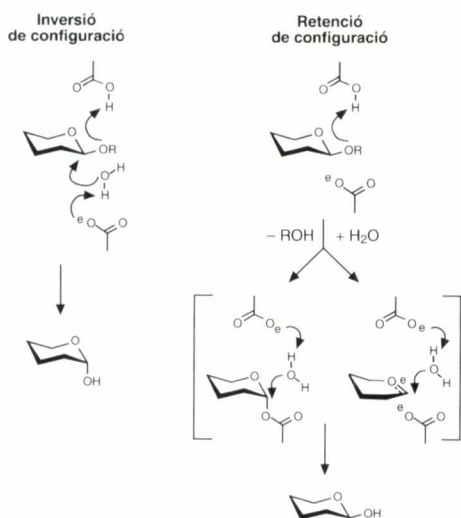


FIGURA 4. Mecanismes proposats per a les glicosidases.

secundària estava més d'acord amb la del T4 lisozim (Querol *et al.*, 1992). Es va procedir a substituir per mutagènesi dirigida tots els catorze residus àcids conservats a les dues regions pel corresponent aminoàcid isostèric sense càrrega (Asn i Gln). Així es va demostrar que la mutació Glu-138-Gln donava lloc a un enzim totalment inactiu; a aquest residu se li va atribuir el paper d'àcid general, així com que el Glu-134 actuava com a nucleòfil (Juncosa *et al.*, 1994). La figura 5 mostra el resultat de l'assaig d'activitat en placa fet directament sobre clons portadors de les diferents mutacions puntuals. El paper dels residus essencials ha estat posteriorment confirmat mitjançant diverses anàlisis cinètiques.

L'estructura cristal·logràfica de l'enzim posteriorment obtinguda (Hahn *et al.*, 1995) mostra que la glucanasa presenta un plegament del tipus β sandvitx. També s'ha posat en evidència l'existència d'un pont disulfur entre les dues úniques cisteïnes que presenta l'enzim i que connecta un llaç que cobreix parcialment la cavitat del centre actiu i una cadena β del nucli hidrofòbic de la proteïna. Per tal d'estudiar el paper del pont disulfur i del llaç del centre actiu en la catàlisi i per

tractar de redissenyar l'especificitat i l'activitat enzimàtiques, es va procedir a fer el que s'anomena «escombrat d'alanines» amb què tots els residus són substituïts «un per un» per Ala. Aquesta estratègia de mutagènesi en combinació amb prediccions per dinàmica molecular i *docking* de substrats ha proporcionat el mutant Met-58-Ala, que és set vegades més actiu que l'enzim salvatge. Així mateix, s'ha comprovat que la substitució de l'Asn-57 per Ala augmenta la termoresistència de l'enzim en un 70% (Pons *et al.*, 1995 i 1997). En aquest moment, s'estan aplicant regles teorícopredictives (Querol *et al.*, 1996) en el disseny racional d'estratègies encaminades a millorar tant la termostabilitat com la termoresistència, propietats importants des del punt de vista de la utilització industrial de l'enzim.

β -Glucosidasa (EC 3.2.1.21)

La β -glucosidasa (β -glucòsid glucohidrolasa) catalitza la transferència de grups glucosídics entre oxígens nucleofílics. En condicions fisiològiques, aquesta reacció de transferència generalment dona com a resultat la hidròlisi d'un enllaç glicosídic. El gen de la β -glucosidasa Bgl3 (479 aminoàcids, 53 kDa) de *Streptomyces* sp. es va clonar inicialment en *S. lividans* i posteriorment en *E. coli* (Pérez-Pons *et al.*, 1994). Per anàlisi d'homologia, l'enzim es va classificar dins de la família 1 de glicosilhidrolases. En el moment de la comparació, formaven part d'aquesta família almenys catorze enzims d'origen bacterià, animal i vegetal, amb la qual cosa la identificació de residus altament conservats per comparació de seqüències es veia molt simplificada. A més, s'havia localitzat per RMN del complex covalent enzim-inhibidor (Withers i Street, 1988) i per mutagènesi dirigida (Trimbur *et al.*, 1992) el residu Glu que actua de nucleòfil catalític a l'enzim d'*Agrobacterium faecalis*. Aquest residu (Glu-383 en la glucosidasa de *Streptomyces* com hem

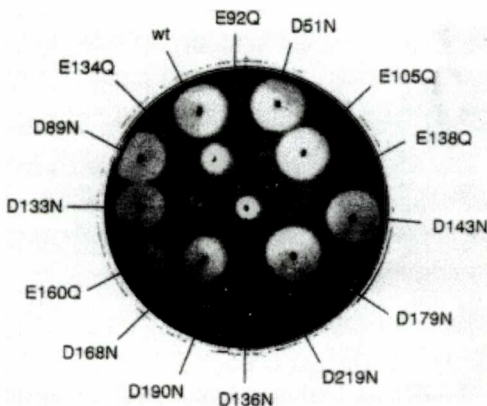


FIGURA 5. Assaig d'activitat en placa utilitzant β -glucà com a substrat i tinció amb roig Congo dels clons portadors de les diferents mutacions puntuals indicades.

comprovat substituint-lo per Gln) està totalment conservat entre tots els membres de la família 1 que actualment inclou almenys vint proteïnes. Quedava per identificar l'aminoàcid que fa el paper de catalitzador àcid/base i tampoc no es coneixia l'estructura cristal·logràfica de cap representant de la família. Per tant, igual que amb la glucanasa, l'estratègia que es va seguir va ser la d'aplicar mutagènesi de rastreig sobre tots els residus Glu i Asp més conservats, i es va trobar que el Glu-178 és l'altre residu essencial. Posteriorment va sortir publicada l'estructura tridimensional de la glucosidasa de *Trifolium repens* (Barrett *et al.*, 1995), cosa que va permetre aplicar tècniques de modelat i dinàmica molecular sobre la glucosidasa de *Trifolium*, amb la qual Bgl3 mostra un 36% d'identitat per seqüència d'aminoàcids. Amb l'estructura modelada, s'han dissenyat racionalment mutants del centre actiu que han permès establir la seva contribució a la catalisi (p.e., Arg-89) i, fins i tot, alterar-la (per exemple, el mutant Cys-181-Ser que, a diferència de l'enzim salvatge, no presenta inhibició per substrats del tipus aril-glucòsid). Actualment ja es disposa de cristalls de la Bgl3 (A. Guasch, manuscrit en preparació) i s'està refinant l'estructura amb dades d'espectres de difracció recollits a 1,8 Å. Els resultats preliminars mostren que aquesta glucosidasa és una proteïna globular, molt compacta i amb una topologia del tipus barril (β/α)₈. La regió catalítica adopta la forma de «butxaca», a l'entrada de la qual es trobarien els residus que intervenen en la unió de la part aglicona de la molècula de substrat. També determina l'especificitat de l'enzim; la Cys-181 està precisament localitzada en aquesta regió. Es disposa així de les bases estructurals que permetran futures investigacions dirigides en aquest enzim per tractar de millorar especialment la seva capacitat de transglicosidació tot adaptant-la a la síntesi de glicoconjugats.

EXEMPLE II. Es disposa d'informació 3D parcial: el domini d'activació de la procarboxipeptidasa A2 humana; la utilitat d'un modelat i del coneixement del procés de plegament

Les carboxipeptidases pancreàtiques se sintetitzen com a precursors inactius o zimògens gràcies a una extensió N-terminal anomenada segment *pro*, que és alliberada al duodè per acció de la tripsina. El segment *pro* està format per un domini globular de vuitanta residus (li direm domini *pro*), també anomenat domini d'activació, connectat a l'enzim per una hèlix α C-terminal. Experiments de calorimetria (DSC) demostren que les carboxipeptidases presenten una desnaturalització totalment irreversible, mentre que el domini *pro* presenta una desnaturalització reversible (Conejero-Lara *et al.*, 1991; Sánchez-Ruiz, *et al.*, 1991). D'altra banda, l'expressió heteròloga de les carboxipeptidases produeix proteïna insoluble, i l'única manera d'aconseguir el plegament correcte de la proteïna recombinant és expressant-la en forma de proenzim. Això suggereix que el domini *pro* actua com una proteïna de plegament interna i que pot ésser un model interessant per estudiar el procés de plegament de les proteïnes.

El fragment de DNA que codifica la procarboxipeptidasa A2 (PCPA2) humana es va obtenir a partir d'una genoteca de DNAC (Catasús *et al.*, 1995). L'estructura primària de la proteïna codificada es va modelar sobre les coordenades de l'estructura terciària de proteïnes homòlogues, com la procarboxipeptidasa A1 de porc i la mateixa forma en bou (Catasús *et al.*, 1995; Aloy *et al.*, 1997). El resultat mostra que el domini *pro* de la PCPA2 està constituït per dues hèlixes α que s'empaqueten contra una fulla- β de quatre cadenes (figura 6). Les hèlixes α es troben exposades al solvent i la fulla- β és la que contacta parcialment amb el centre actiu de l'enzim. Observant les estructures secundàries

a)

```

Sec.St      TT SSSSS  THHHHHHHHHHTT 333  SSSS  TTTT SSSSS 333HHHHHHHHHTT  SSSSST
Prot  MRSLETFVGDQVLEIVPSNEEQIKNLLQLEAQEHLQLDFWKSPTTPGETAHVRVVFVNVQAVKVFLESQGIAYSIMIEDVQ
H1wt      VPSNEEQIKNLLQLEAQEHLQY
H1Mt      VPSNEEQIKKLLLEAKKHLQY
H2wt
H2Mt      FVNVQAVKVFLESQGIAY
          FVNVEAVKAFLEAHGIAY
    
```

b)

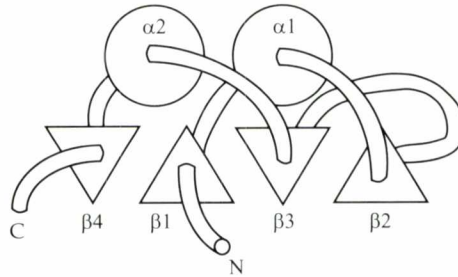


FIGURA 6. Esquema de l'estructura secundària i patró general de conformació del domini d'activació de la PCPA2 humana.

de la proteïna modelada ens vàrem adonar que els aminoàcids constituents de les hèlixes α a la proteïna silvestre són deficientes com a formadors d'aquest tipus d'estructura i que el dipol elèctric de les hèlixes és gran i desestabilitza el plegament global. A partir d'aquestes observacions vàrem dissenyar mutacions per tal d'incrementar la capacitat de mantenir-se en conformació de hèlix α , a la vegada que disminuïem el dipol elèctric i creàvem interaccions estabilitzants de la conformació α ($i, i+3; i, i+4$) (figura 6).

Utilitzant l'algoritme AGADIR (Muñoz i Serrano, 1997), es va predir un percentatge de conformació α del 16% i del 5% per a les seqüències de les hèlixes 1 i 2, respectivament. La predicció per a les seqüències mutants va ser del 63% i del 44%, cosa que implica un augment de l'estabilitat de 2,5 i 1,6 kcal mol⁻¹, respectivament. Finalment, es van construir tres mutants: 1) hèlix 1 estabilitzada, 2) hèlix 2 estabilitzada, i 3) el doble mutant. Els fragments que codifiquen el domini *pro* silvestre i els dominis mutats es van clonar en un vector d'expressió d'*E. coli* (Villegas *et al.*, 1995b).

A la figura 7 es mostren alguns dels resultats experimentals obtinguts en l'estudi de la proteïna salvatge i dels seus mutants. L'estudi de la variació de la intensitat de fluorescència en variar la concentració d'un agent desnaturant com la urea (figura 7) permet deduir dades termodinàmiques. En ser mesures fetes en l'equilibri, i pel fet d'ésser el punt d'inflexió de la corba aquell en el qual la constant d'equilibri és 1 (i l'energia lliure, per tant, 0), és possible extrapolar l'energia lliure del plegament global, és a dir, la diferència en estabilitat entre les formes nativa i desnaturitzada. Aquesta estabilitat és de 4,2 kcal mol⁻¹ per al domini *pro*. El pendent de la corba és una mesura indirecta de la diferència en el grau d'exposició al dissolvent entre les formes nativa i desnaturitzada i, per tant, del grau de compactació de la proteïna.

Ara bé, l'estudi en condicions d'equilibri poques vegades permet la detecció de formes intermediàries en el procés de plegament. Si tenim tres estats, desnaturitzat, intermediari i natiu, la cinètica de conversió de l'estat desnaturitzat en l'intermediari

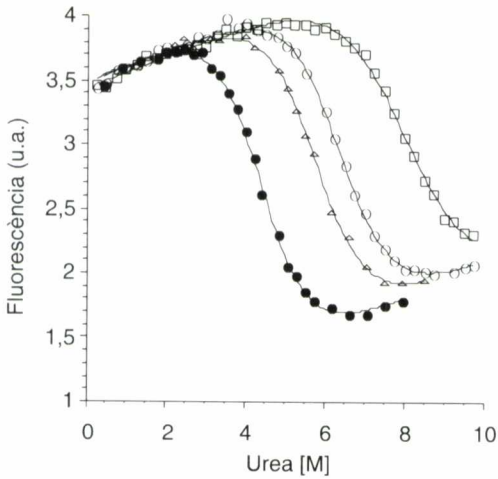


FIGURA 7. Variació de la fluorescència durant la desnaturalització del domini ADA2H i les formes redissenyades, segons la concentració d'urea. (●) ADA2H, forma silvestre; (Δ) mutant amb hèlix 1 estabilitzada; (○) mutant amb hèlix 2 estabilitzada; (□) doble mutant.

és distinta de la cinètica de conversió de l'intermediari en la forma nativa, ja que són processos diferents. La cinètica d'una reacció es pot mesurar amb un utilatge anomenat de flux retint (stopped-flow) que consisteix en dues xeringues, una que conté la proteïna desnaturalitzada i l'altra, l'amortidor que serveix per renaturalitzar. La mescla es realitza a una concentració final de desnaturalitzant determinada, i un detector de fluorescència connectat en sèrie ens dona la lectura de la variació de la fluorescència durant el repliegament. D'aquesta manera, podem conèixer la velocitat de la reacció a una concentració determinada de desnaturalitzant. Si es representa el logaritme natural de la velocitat de la reacció de repliegament davant de la concentració de desnaturalitzant (figura 8), la gràfica obtinguda és lineal si no hi ha intermediaris del procés de repliegament. Per contra, si l'intermediari existeix, la seva aparició provocarà una desviació de la linealitat a la concentració de desnaturalitzant on l'intermediari és estable.

Amb l'aparell de flux retint també podem estudiar la reacció de desplegament si una xeringa conté la proteïna nativa i l'altra una dissolució amb desnaturalitzant. La representació del logaritme natural de les constants cinètiques de desplegament davant de la concentració de desnaturalitzant és sempre una recta. A partir de la combinació lineal d'ambdues parts del procés, repliegament i desplegament, podem obtenir informació sobre la velocitat de reacció a diferents concentracions de desnaturalitzants i, de la mateixa manera que en l'estudi en condicions d'equilibri, sobre l'estabilitat i el grau de compactació del plegament. Els valors obtinguts van ser els mateixos que van obtenir-se a l'estudi en l'equilibri i que van ésser també corroborats per estudis de CD (Viguera *et al.*, 1996). Aquests estudis varen indicar també que el domini *pro* de la PCPA2 humana es pot plegar i desplegar en uns pocs mil·lisegons, un dels temps més ràpids descrits per a proteïnes.

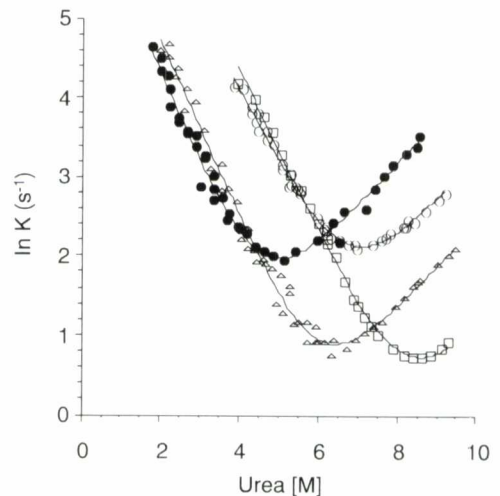


FIGURA 8. Representació (semilogarítmica) de la velocitat de la reacció de repliegament del domini ADA2H i les formes redissenyades, a diferents concentracions d'urea. (●) ADA2H, forma silvestre; (Δ) mutant amb hèlix 1 estabilitzada; (○) mutant amb hèlix 2 estabilitzada; (□) doble mutant.

Tant l'estudi de la cinètica del procés com de la seva calorimetria revelen que el domini *pro* es plega sense la formació d'un intermediari. Sembla que l'aparició d'intermediaris en un procés de plegament depèn de la mida de la cadena polipeptídica que s'ha de plegar i del tipus de plegament, i són necessaris quan la proteïna excedeix del centenar de residus (Villegas *et al.*, 1995a). L'absència d'intermediari i la ràpida velocitat de plegament del domini *pro* sembla que són la causa que aquest domini faciliti el plegament del proenzim i que actuï com una proteïna de plegament interna.

En comparar els paràmetres termodinàmics i cinètics obtinguts en els quatre casos estudiats, s'observa un increment de l'estabilitat d' $1,5 \text{ kcal mol}^{-1}$, $1,7 \text{ kcal mol}^{-1}$, i de $3,2 \text{ kcal mol}^{-1}$ per als mutants de l'hèlix 1, de l'hèlix 2 i per al doble mutant, respectivament (Viguera *et al.*, 1996). L'efecte d'estabilització aconseguit amb el redisseny independent de les hèlices esdevé, doncs, additiu. A més, l'estabilització pot originar un augment en la velocitat del plegament de fins a quatre vegades, com s'observa amb el mutant de l'hèlix 2. L'exemple de l'estabilització del domini *pro* de la procarboxipeptidasa A2 demostra que el plegament de les proteïnes no està optimitzat evolutivament pel que fa referència a l'estabilitat i a la velocitat del plegament, i que aquestes propietats són millorables per l'enginyeria de proteïnes.

Però la potencialitat de les tècniques d'EP no s'acaba aquí: també ens poden servir per entendre quina és l'estructura d'un estat efímer, no aïllable ni caracteritzable per cap altra metodologia, com pot ser l'estat de transició en la via de plegament d'una proteïna. Per fer-ho, es parteix del disseny d'una bateria de mutants de pertorbació, disruptius, repartits per tota la seqüència codificant. Els mutants disruptius són aquells que eliminen un grup funcional sen-

se introduir noves interaccions amb la resta de la proteïna. Un bon exemple és la mutació d'Ala a Gly, amb la qual podríem mesurar l'efecte desestabilitzant de l'eliminació d'un grup metil i, com a conseqüència d'això, la importància d'aquest metil en l'energètica del plegament. Mitjançant la determinació de les cinètiques de les reaccions de replegament i desplegament podem deduir si la interacció en la qual està implicat el grup metil està formada o no a l'estat de transició. Si ho està, la desestabilització de l'estat de transició serà la mateixa que l'observada al mutant plegat, i això es reflectirà a la cinètica de desplegament. Si la interacció no està formada a l'estat de transició, la desestabilització d'aquest estat serà la mateixa que l'observada al mutant desplegat, i això es reflectirà a la cinètica de replegament. Amb aquesta metodologia s'ha observat que les proteïnes monodomini (és a dir, petites) es pleguen segons el que s'ha anomenat el model de nucleació i condensació, on la formació d'estructura secundària és paral·lela a la formació d'estructura terciària. Abans del desenvolupament de les tècniques d'enginyeria de proteïnes era inimaginable poder conèixer quines són les interaccions que es donen a l'estat de transició i, per tant, entendre quin és el camí de plegament de les proteïnes.

Cal afegir que darrerament hem aconseguit determinar experimentalment l'estructura tridimensional detallada de la PCPA2 humana completa per difracció de raigs X (García-Saez *et al.*, 1997). La comparació de la conformació derivada per al domini *pro* i per a tota la proteïna amb les obtingudes prèviament per modelat computacional indica que aquests modelats varen ser molt acurats, amb desviacions significatives només en les regions d'unió entre estructures secundàries regulars, i concretament en els arcs.

En disposar d'una estructura 3D acura-

da s'obren moltes possibilitats de redisseny d'aquesta proteïna amb objectius diferents. Per exemple, per redisseny de la regió *pro* o de la regió enzimàtica es pot facilitar l'expressió recombinant i la renaturabilitat de la proteïna.

EXEMPLE III. Es coneix bé l'estructura 3D de la proteïna: l'inhibidor de carboxipeptidases de patata (PCI)

A molts vegetals s'hi han trobat inhibidors de diverses proteases que poden servir com a sistema de defensa de les plantes contra insectes o fongs (Ryan, 1989). L'inhibidor proteic de carboxipeptidases que es troba a la patata (PCI) és una proteïna de només trenta-nou aminoàcids. Es coneix l'estructura tridimensional del PCI en dissolució, determinada per ressonància magnètica nuclear (Cloure *et al.*, 1987) i l'estructura del complex que forma el PCI amb la carboxipeptidasa A (CPA), determinada per difracció de raigs X (figura 9) (Rees i Lipscomb, 1982). De l'estructura d'aquesta proteïna destaca la presència dels tres ponts disulfur que estableixen el seu nucli globular, i també l'absència d'estructures secundàries

regulars, a excepció d'un curt segment en hèlix 3₁₀. De l'estructura del complex amb la carboxipeptidasa A, s'observa que el PCI interacciona amb el centre actiu de l'enzim a través de la seva cua C-terminal. Hi ha també una zona de contacte secundari entre els residus 27 a 30 del PCI i l'enzim.

Al nostre laboratori vàrem triar aquest inhibidor com a model d'estudi per a enginyeria de proteïnes per diverses raons. D'una banda, el fet que es conegués l'estructura del seu complex amb la carboxipeptidasa el feia un bon candidat per estudiar més a fons la interacció entre les dues proteïnes. Una gran part de mecanismes funcionals de les proteïnes es basen en la interacció específica proteïna-proteïna. Entendre els factors que expliquen l'especificitat i l'estabilitat d'aquestes interaccions és, doncs, bàsic per arribar a comprendre aquests mecanismes des del punt de vista molecular. D'altra banda, el PCI, en ser una proteïna molt petita, presenta el gran avantatge de reduir els temps de càlcul necessaris per efectuar simulacions per dinàmica molecular o per aplicar altres mètodes de simulació computacional. Des d'aquest punt de vista, el PCI és un bon model tant per a l'ús dels mètodes computacionals per al disseny d'experiments com per a la contrastació experimental de la validesa o la utilitat de nous mètodes computacionals en l'enginyeria de proteïnes.

Des del punt de vista experimental, la base per iniciar estudis d'enginyeria de proteïnes és disposar d'un bon sistema d'expressió de la proteïna en forma recombinant, que permeti la construcció i l'obtenció de mutants específics. En el cas del PCI, això es va dur a terme mitjançant el desenvolupament d'un sistema molt eficient d'expressió en *E. Coli* (200 mg/L), en bioreactor, a partir d'un gen sintètic codificant per a una de les formes de l'inhibidor trobades a la patata, en un vector de secreció al periplas-

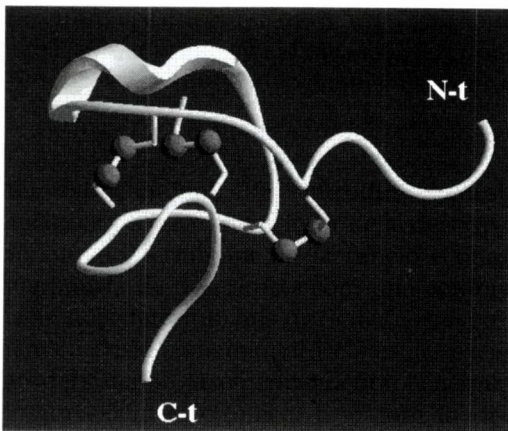


FIGURA 9. Estructura 3D del PCI.

ma (Molina *et al.*, 1992; Marino *et al.* 1994).

Per tal d'estudiar la interacció entre la regió de contacte primària entre el PCI i el centre actiu de la carboxipeptidasa A, que comprèn els residus 36 a 39 de la cua C-terminal (ProTyrValGly), es varen dissenyar diversos mutants. Posteriorment es va determinar la constant d'inhibició K_i de la CPA per a cada un d'ells, constant que està directament relacionada amb l'estabilitat del complex enzim-inhibidor, mesurat com l'increment d'energia lliure de Gibbs de dissociació del complex (ΔG_d°). En primer lloc, es va estudiar la contribució dels dos darrers residus del PCI a l'estabilitat del complex amb la CPA (Molina *et al.*, 1994). Aquests residus, d'acord amb l'estructura del complex, són els que ocuparien els llocs S1 i S1' al centre actiu de l'enzim. Per al mutant delectiu delG39 es va observar una K_i pràcticament idèntica a la del PCI natural (PCIwt). Aquest resultat implica que la presència de la Gly39 al lloc S1' de l'enzim no contribueix pràcticament a l'estabilitat del complex. Així, la hidròlisi d'aquesta Gly, que té lloc quan l'inhibidor s'uneix a l'enzim, no modifica pràcticament la seva capacitat inhibidora. La contribució de la Val38 es va analitzar per mitjà de la producció i l'anàlisi de les K_i dels mutants del V38G39, V38G, V38A, V38L i V38I. Els resultats obtinguts van permetre «disseccionar» la contribució a l'estabilitat del complex del residu Val38, de la manera que es mostra a la figura 10.

La contribució total del residu Val-38, estimada a partir de la diferència entre els valors de ΔG_d° dels mutants del G39 i del V38G39, és de 5,2-5,8 Kcal mol⁻¹. Aquest valor representa el 45-48% de l'estabilitat total del complex PCI-CPA (11,6-12,1 kcal mol⁻¹), i correspon a un increment en la constant d'inhibició pel mutant delV38G39 d'un factor de 10⁴, una reducció dramàtica de la capacitat inhibidora respecte al PCIwt.

Aquests resultats demostren que aquest residu Val38 és clau per a l'estabilitat del complex PCI-CPA. Com es mostra a la figura 10, la contribució de la cadena lateral d'aquest residu, per comparació amb l'estabilitat del mutant V38G, és de 3,4-4,0 kcal mol⁻¹. Si se suposa que aquesta contribució és deguda a interaccions hidrofòbiques de la cadena alifàtica amb el centre actiu de l'enzim, es pot subdividir en aportacions individuals de cada grup d'aquesta cadena per comparació dels valors de ΔG_d° del PCIwt i dels diversos mutants.

Com es mostra a la figura 10, els resultats suggereixen que cada un dels grups metil o metilè de la cadena alifàtica contribueixen a l'estabilitat del complex amb una contribució similar, al voltant d'1,0-1,5 kcal mol⁻¹, fins a la posició γ de la cadena, a partir de la qual ja no hi hauria interacció hidrofòbica amb l'enzim (Molina *et al.*, 1994). Per tal de reforçar aquesta hipòtesi, es va dur a terme un estudi d'aquesta sèrie de mutants per mètodes computacionals. S'obté una bona correlació entre les variables experimentals i les calculades, cosa que reforça tant la interpretació dels resultats experimentals com la validesa dels models construïts i els algorismes de càlcul utilitzats. Una altra aproximació computacional que s'ha assajat per intentar la predicció teòrica de les diferències entre l'energia d'estabilització per diferents complexos CPA-PCI mutants és l'anomenada teoria de perturbacions que s'ha descrit a la pàgina 52.

Un altre aspecte referent a la interacció entre la regió C-terminal del PCI i la CPA que ha estat objecte d'estudi, tant des del punt de vista experimental com computacional, és el que fa referència a l'orientació espacial i a la mobilitat conformacional d'aquesta cua C-terminal, i el seu paper en l'afinitat de l'inhibidor per la carboxipeptidasa. El resultat més destacat és que la cua C-terminal, és a dir, la zona responsable de la in-

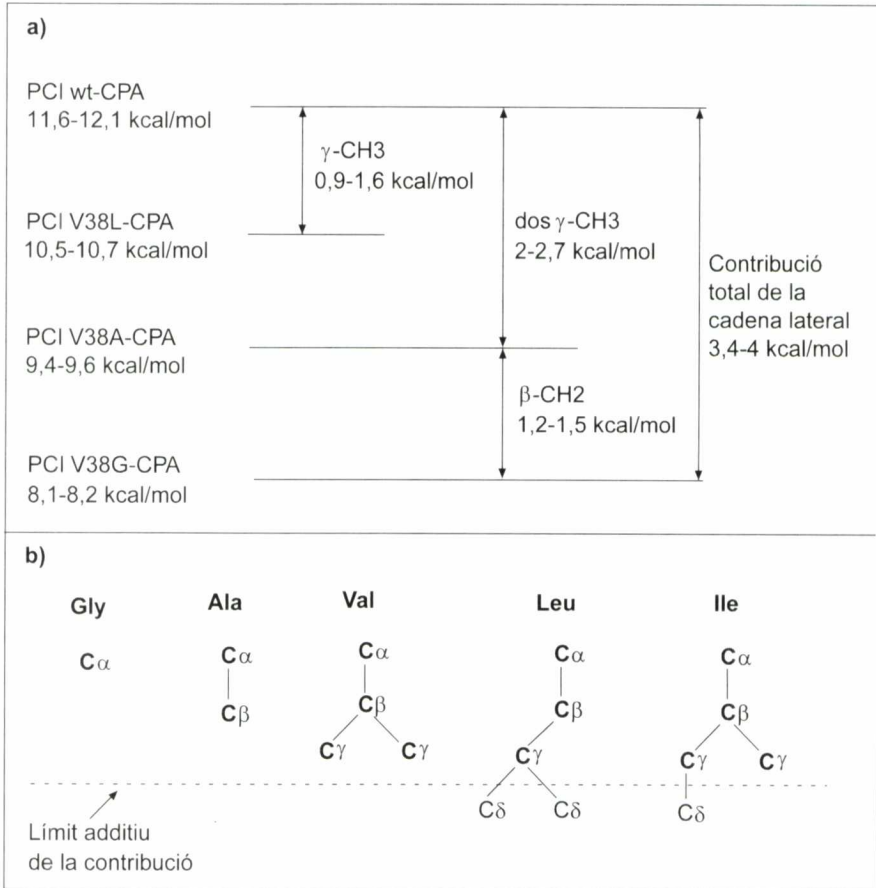


FIGURA 10. a) Diferències en les energies lliures de dissociació (ΔG°) dels complexos de CPA i PCI silvestre i mutats a la posició 38. b) Esquema dels grups R a la posició 38 del PCI.

teracció primària amb l'enzim, té una mobilitat força restringida, i adopta en estat aïllat una orientació a l'espai semblant a la que s'observa quan es troba formant el complex amb la CPA. Aquesta restricció conformational seria en part la responsable de l'elevada afinitat de l'inhibidor per la CPA (Oliva *et al.*, 1995).

D'altra banda, s'han dut a terme estudis sobre el camí del plegament del PCI. En tenir tres ponts disulfur a la seva estructura, el camí del plegament de l'inhibidor va poder ser estudiat mitjançant l'anàlisi dels intermediaris que es formen al llarg del procés

de formació dels ponts disulfur simultani al plegament. Els resultats mostren que el plegament de la molècula de PCI segueix un camí que s'inicia amb un ràpid col·lapse, en què es formen de manera seqüencial un gran nombre d'intermediaris amb 1, 2 i 3 ponts disulfur, seguit d'una etapa lenta de reorganització dels aparellaments dels tres ponts disulfur formats, per conduir a l'estructura nativa (Chang *et al.*, 1994). La caracterització experimental del mecanisme de plegament del PCI és la base per a estudis de simulació computacional del plegament que s'estan portant a terme. D'altra banda,

l'estudi d'aquest mecanisme, tant per al PCIwt com per a diversos mutants, ha proporcionat informació sobre zones o residus de la molècula que resulten clau per a un plegament eficient, informació útil de cara a projectes de redisseny de l'inhibidor.

PERSPECTIVES FUTURES DE L'ENGINYERIA DE PROTEÏNES

A més de la seva aplicabilitat per aprofundir en el coneixement d'una proteïna, l'EP té un gran futur per dos motius, entre d'altres. El primer perquè des del punt de vista industrial es farà cada vegada més ús d'enzims en processos de síntesi química, de manipulació biològica, de transformació d'aliments, etc., donat que els enzims, contràriament a la major part d'operacions de la química orgànica, treballen en condicions de medis més compatibles amb el medi ambient i, a més, són estereoespecífics. El segon és que els avenços de la genòmica i de la proteòmica estan donant lloc a milers d'ORF, proteïnes potencials a la recerca de la seva funció. Per descomptat que l'EP i la biocomputació tenen molt a dir en aquesta situació.

Des d'un punt de vista més concret, els reptes que té l'EP es poden resumir en els punts següents:

- Ser capaços de predir el lloc i l'efecte d'una mutació (o de diverses) per arribar a un objectiu determinat (millora d'activitat, estabilitat, especificitat, etc.).

- Partint d'estructures 3D de proteïnes diana i dades de mutants específicament dissenyats, ser capaços de dissenyar lligands específics (poden ser inhibidors, fàrmacs, etc.). Això implica la capacitat de simular interaccions macromoleculares en el computador o d'assajar experimentalment i ràpidament milions d'interaccions diferents (operació que ja es fa avui en dia amb les biblioteques combinatòries o fàgiques).

- Del creixent coneixement de biologia estructural, fruit dels avenços en cristal·lografia, NMR i la mateixa EP, entre ells la millora de l'algorítmica d'alineaments i la predicció computacional, arribar a comprendre com es creen a la naturalesa noves proteïnes i noves funcions a partir de les existents. Aquest coneixement permetria dissenyar noves proteïnes i noves funcions *ab initio*. Ara per ara no som capaços (almenys de manera rutinària o variada) ni de dissenyar estructures supersecundàries proteiques suficientment compactes i estables perquè puguin ser expressades heteròlogament o cristal·litzades, com moltes proteïnes naturals.

Tot això s'inscriu en el que s'anomena un disseny racional de proteïnes i consisteix no tan sols a millorar el nostre coneixement de les regles segons les quals les proteïnes arriben a una estructura estable amb una funció biològica, sinó també per la millora dels mètodes computacionals en l'ús. Això vol dir augment de capacitat i velocitat de càlcul dels computadors, però sobretot del disseny de millors algorismes de predicció i simulació de l'estructura de proteïnes. En qualsevol cas, és d'esperar que la interfase i l'ús combinat d'aproximacions experimentals i computacionals en EP creixin de manera contínua i es facin cada vegada més útils biotecnològicament.

AGRAÏMENTS

Una part dels resultats aquí presentats ha estat recollida per diversos membres no citats dels nostres grups de recerca, i una altra part s'ha obtingut en col·laboració amb els grups de recerca dels doctors Luis Serrano (EMBL), Antoni Planas (IQS) i Miquel Coll (CID-CSIC) entre d'altres. Agraïm a tots la seva disponibilitat i el seu interès en els nostres estudis, com també el suport del

Centre de Referència en Biotecnologia (Generalitat de Catalunya) i de la Fundació Maria Fca. de Roviralta.

BIBLIOGRAFIA GENERAL

- REES, A.; STERNBERG, M. J. E.; WETZEL, R. *Protein engineering*. R. IRL Press, 1992.
- CLELAND, J.; CRAIK, C. *Protein Engineering: principles and practice*. Wiley Liss, 1996.

BIBLIOGRAFIA CITADA

- ALTSCHUL, S. F.; GISH, W.; MILLER, W.; MYERS, E. W.; LIPMAN, D. J. (1990). «Basic local alignment search tool». *J. Mol. Biol.*, núm. 215, pàg. 403-410.
- ALOY, P.; CEDANO, J.; OLIVA, B.; AVILÉS, F. X.; QUEROL, E. (1997). «TransMem: A neural network implemented in Excel spreadsheets for predicting transmembrane domains in proteins». *Comput. Appl. Biosci.*, núm. 13, pàg. 231-234.
- ALVAREZ-SANTOS, S.; GONZÁLEZ-LAFONT, A.; LLUCH, J. M.; OLIVA, B.; AVILÉS, F. X. (1996). «Theoretical Study of the mechanism of Carboxypeptidase A inhibition by Zinc ions». *New J. Chemistry*, núm. 20, pàg. 979-984.
- BARRETT, T.; SURESH, C. G.; TOLLEY, S. P.; DODSON, E. J.; HUGHES, M. A. (1995). «The crystal structure of a cyanogenic β -glucosidase from white clover, a family 1 glycosyl hydrolase». *Structure*, núm. 3, pàg. 951-960.
- CATASÚS, LL.; VENDRELL, J.; AVILÉS, F. X.; CARREIRA, S.; PUIGSERVER, A.; BILLETTER, M. (1995). «The sequence conformation of human pancreatic procarboxypeptidase A2. cDNA cloning, sequence analysis, and three-dimensional model». *J. Biol. Chem.*, núm. 270, pàg. 6651-6657.
- CHANG, J.-Y.; CANALS, F.; SCHINDER, P.; QUEROL, E.; AVILÉS, F. X. (1994). «The disulfide folding pathway of potato carboxypeptidase inhibitor». *J. Biol. Chem.*, núm. 269, pàg. 22087-22094.
- CLORE, G. M.; GRONEBORN, A. M.; NILGES, M.; RYAN, C. A. (1987). «Three-dimensional structure of potato carboxypeptidase inhibitor in solution. A study using nuclear magnetic resonance, distance geometry and restrained molecular dynamics». *Biochemistry*, núm. 26, pàg. 8012-8023.
- CLOTET, J.; CEDANO, J.; QUEROL, E. (1994). «An Excel spreadsheet computer program for prediction of structural characteristics of proteins». *Comput. Appl. Biosci.*, núm. 10, pàg. 495-500.
- CONEJERO-LARA F.; SANCHEZ-RUIZ, J. M.; MATEO, P. L.; BURGOS, F. J.; VENDRELL, J.; AVILÉS, F. X. (1991). «Differential scanning calorimetry study of carboxypeptidase B, procarboxypeptidase B and its globular activation domain». *Eur. J. Biochem.*, núm. 200, pàg. 663-670.
- DAURA, X.; OLIVA, B.; QUEROL, E.; AVILÉS, F. X.; TAPIA, O. (1996). «On the sensitivity of MD trajectories to changes in the water-protein interaction parameters. The potato carboxypeptidase inhibitor in water as a test case for the GROMOS force field». *Proteins: Struct., Funct. Genet.*, núm. 25, pàg. 89-103.
- DAYHOFF, M. O. (1978). «Atlas of Protein Sequence and Structure» *Natl. Biomed. Res. Found. Washington*, vol 5, supl. 3, pàg. 345-358.
- GABB, H. A.; JACKSON, R. M.; STERNBERG, M. J. E. (1997). «Modelling protein docking using shape complementarity, electrostatics and biochemical information». *J. Mol. Biol.*, núm. 272, pàg. 106-120.
- GARCÍA-SAEZ, I.; REVERTER, D.; VENDRELL, J.; AVILÉS, F. X.; COLL, M. (1997). «The three-dimensional structure of human procarboxypeptidase A2. Deciphering the basis of the inhibition, activation and intrinsic activity of the zymogen». *EMBO J.*, núm. 16, pàg. 6906-6913.
- HAHN, M.; PONS, J.; PLANAS, A.; QUEROL, E.; HEINEMANN, U. (1995). «Crystal structure of *Bacillus licheniformis* 1,3-1,4- β -D-glucan 4-glucanohydrolase at 1.8Å resolution». *FEBS Letters*, núm. 374, pàg. 221-224.
- HENIKOFF, S.; HENIKOFF, J. G. (1992). «Amino acid substitution matrices from protein blocks». *Proc. Natl. Acad. Sci. USA*, núm. 89, pàg. 10915-10919.
- HENRISSAT, B. (1991). «A classification of glycosyl hydrolases based on amino acid sequence similarity». *Biochem. J.*, núm. 280, pàg. 309-306.
- HOJ, P. B.; CONDRON, R.; TRAEGER, J. C.; MCAULIFFE, J. C.; STONE, B. A. (1992). «Identification of glutamic acid 105 at the active site of *Bacillus amyloliquefaciens* 1,3-1,4- β -D-glucan 4-glucanohydrolase using epoxy based inhibitors». *J. Biol. Chem.*, núm. 267, pàg. 25059-25066.
- HOOP, T. P.; WOODS, K. R. (1981). «Prediction of protein antigenic determinants from amino acid sequence». *Proc. Natl. Acad. Sci. USA*, núm. 78, pàg. 3824-3828.
- JONES, T. H.; THIRUP, S. (1986). «Using known substructures in protein model building and crystallography». *EMBO J.*, núm. 5, pàg. 819-822.
- JONES, D. T.; TAYLOR, W. R.; THORNTON, J. (1992). «A new approach to protein fold recognition». *Nature*, núm. 358, pàg. 86-89.
- JUNCOSA M.; PONS, J.; DOT, T.; QUEROL, E.; PLANAS, A. (1994). «Identification of active site carboxylic residues in *Bacillus licheniformis* endo-1,3-1,4-D-glucan 4-glucanohydrolase by site-directed mutagenesis». *J. Biol. Chem.*, núm. 269, pàg. 14530-14535.

- JUNCOSA, M.; PONS, J.; PLANAS, A.; QUEROL, E. (1994). «Improved efficiency in site-directed mutagenesis by PCR using a *Pyrococcus* sp GB-D polymerase». *Bio-Techniques*, núm. 16, pàg. 820-821.
- KING, R. D.; SAQI, M.; SAYLE, R.; STERNBERG, M. J. E. (1997). «DSC: Public domain protein secondary structure prediction». *Comput. Appl. Biosci.*, núm. 13, pàg. 473-474.
- KYTE, J.; DOOLITTLE, R. F. (1982). «A simple method for displaying the hydropathic character of a protein». *J. Mol. Biol.*, núm. 157, pàg. 105-132.
- LIU, H.; MÜLLER-PLATHE, F.; VAN GUNSTEREN, W. F. (1996). «A combined Quantum/Classical Molecular Dynamics Study of the catalytic mechanism of HIV protease». *J. Mol. Biol.*, núm. 261, pàg. 454-469.
- LLOBERAS, J.; PÉREZ-PONS, J. A.; QUEROL, E. (1991). «Molecular cloning, expression and nucleotide sequence of the endo- β -1,3-1,4-D-glucanase gene from *Bacillus licheniformis*. Predictive structural analyses of the encoded polypeptide». *Eur. J. Biochem.*, núm. 197, pàg. 337-343.
- MARINO, C.; MOLINA, M. A.; CANALS, F.; AVILÉS, F. X.; QUEROL, E. (1994). «Overproduction of a recombinant carboxypeptidase inhibitor protein by optimization of the fermentation conditions». *Appl. Microbiol. Biotechnol.*, núm. 41, pàg. 632-637.
- MARK, A. E.; VAN GUNSTEREN, W. F. (1994). «Decomposition of the free energy of a system in terms of specific interactions. Implications for theoretical and experimental studies». *J. Mol. Biol.*, núm. 240, pàg. 167-176.
- MCCARTER, J. D.; WITHERS, S. G. (1994). «Mechanisms of enzymatic glycoside hydrolysis». *Curr. Op. Struct. Biol.*, núm. 4, pàg. 885-892.
- MOLINA, M. A.; AVILÉS, F. X.; QUEROL, E. (1992). «Potato carboxypeptidase inhibitor synthetic gene cloned and expressed in an active form in *Escherichia coli*». *Gene*, núm. 116, pàg. 129-138.
- MOLINA, M. A.; MARINO, C.; OLIVA, B.; AVILÉS, F. X.; QUEROL, E. (1994). «The C-tail Valine is a key residue for the stabilization of the complex between potato inhibitor and carboxypeptidase A». *J. Biol. Chem.*, núm. 269, pàg. 21467-21472.
- MULLIS, K.; FERRE, F.; GIBBS, R., eds. (1994). «The polymerase chain reaction». Birkhäuser, Boston.
- MUÑOZ, V.; SERRANO, L. (1997). «Development of the multiple sequence approximation within the AGADIR model of α -helix formation. Comparison with the Zimm-Bragg and Lifson-Roig formalisms». *Biopol.*, núm. 41, pàg. 495-509.
- NAKAYAME, K. L.; ECKSTEIN, F. (1986). «Inhibition of restriction endonuclease NciI cleavage by phosphothiolate groups and its application to oligonucleotide-directed mutagenesis». *Nucl. Acids Res.*, núm. 14, pàg. 9679-9698.
- NEEDLEMAN, S. B.; WUNSCH, C. D. (1970). «A general method applicable to the search for similarities in the aminoacid sequence of two proteins.» *J. Mol. Biol.*, núm. 48, pàg. 443-453.
- OLIVA, B.; DAURA, X.; NILSSON, O.; QUEROL, E.; AVILÉS, F. X.; TAPIA, O. (1995). «Structure and fluctuation patterns of potato carboxypeptidase A inhibitor protein in aqueous solution. A molecular dynamics study». *Eur. Biophys. J.*, núm. 24, pàg. 1-12.
- OLIVA, B.; BATES, P.; QUEROL, E.; AVILÉS, F. X.; STERNBERG, M. J. E. (1997). «An automatic classification of the structure of protein loops». *J. Mol. Biol.*, núm. 267, pàg. 1-17.
- PEARSON, W. R.; LIPMAN, D. J. (1988). «Improved tools for biological sequence comparison». *Proc. Natl. Acad. Sci. USA*, núm. 85, pàg. 2444-2448.
- PÉREZ-PONS, J. A.; CAYETANO, A.; REBORDOSA, X.; LLOBERAS, J.; GUASCH, A.; QUEROL, E. (1994). «A β -glucosidase gene from *Streptomyces* sp ATCC 11238. Molecular cloning, nucleotide sequence, purification and characterization of a new member of family 1 glycosyl hydrolases». *Eur. J. Biochem.*, núm. 223, pàg. 557-565.
- PÉREZ-PONS, J. A.; PADRÓS, E.; QUEROL, E. (1995). «Prediction and Fourier transform infrared spectroscopy estimation of the secondary structure of a recombinant β -glucosidase from *Streptomyces* sp (ATCC 11238)». *Biochem. J.*, núm. 308, pàg. 791-794.
- PLANAS, A.; JUNCOSA, M.; CAYETANO, A.; QUEROL, E. (1992a). «Studies on the *Bacillus licheniformis* endo- β -1,3-1,4-glucanase: characterization and kinetic analyses». *Appl. Microbiol. Biotechnol.*, núm. 37, pàg. 583-589.
- PLANAS, A.; JUNCOSA, M.; LLOBERAS, J.; QUEROL, E. (1992b). «Essential catalytic role of Glu134 in endo- β -1,3-1,4-D-glucan 4-glucanohydrolase from *Bacillus licheniformis* as determined by site-directed mutagenesis». *FEBS Letters*, núm. 308, pàg. 141-145.
- PONS, J.; QUEROL, E.; PLANAS, A. (1995). «Contribution of the disulfide bridge to the stability of *Bacillus licheniformis* endo-1,3-1,4-D-glucan 4-glucanohydrolase». *Prot. Engineer.*, núm. 8, pàg. 939-945.
- PONS, J.; QUEROL, E.; PLANAS, A. (1997). «Mutational analysis of the major loop of *Bacillus* 1,3-1,4- β -D-glucan 4-glucanohydrolase: effects on protein stability and substrate binding». *J. Biol. Chem.*, núm. 272, pàg. 13006-13012 (1997).
- QUEROL, E.; PADRÓS, E.; JUNCOSA, M.; PLANAS, A.; LLOBERAS, J. (1992). «Secondary structure of a *Bacillus* endo- β -1,3-1,4-D-glucanase: a study by Fourier Transform Infrared Spectroscopy and predictive methods». *Biochem. Biophys. Res. Commun.*, núm. 184, pàg. 612-617.
- QUEROL, E.; PÉREZ-PONS, J. A.; MOZO-VILLARIAS, A. (1996). «Analysis of conformational characteristics related to protein thermostability». *Prot. Engineer.*, núm. 9, pàg. 265-271.

- REES, D. C.; LIPSCOMB, W. N. (1982). «Refined Crystal Structure of the Potato Inhibitor Complex of Carboxypeptidase A at 2,5 Å of Resolution». *J. Mol. Biol.*, núm. 160, pàg. 475-498.
- REVERTER D.; VILLEGAS, V.; VENTURA, S.; VENDRELL, J.; AVILÉS, F. X. (1998). «Human procarboxypeptidase A2: overexpression in *Pichia pastoris* and detailed characterization of its activation pathway». *J. Biol. Chem.*, núm. 273, pàg. 3535-3541.
- ROST, B.; SANDER, C. (1993). «Prediction of protein secondary structure at better than 70% accuracy». *J. Mol. Biol.*, núm. 232, pàg. 584-599.
- ROST, B.; CASADIO, R.; FARISELLI, P.; SANDER, C. (1995). «Transmembrane helices predicted at 95% accuracy». *Prot. Science*, núm. 4, pàg. 521-533.
- RUFINO, S. D.; DONATE, L. E.; CANARD, L. H. J.; BLUNDELL, T. L. (1997). «Predicting the Conformational Class of Short and Medium Size Loops Connecting Regular Secondary Structures: Application to Comparative Modelling». *J. Mol. Biol.*, núm. 267, pàg. 352-367.
- RYAN, C. A. (1989). «Proteinase Inhibitor Gene Families, Strategies for Transformation to Improve Plant Defenses Against Herbivores». *BioEssays*, núm. 10, pàg. 20-24.
- SÁNCHEZ-RUIZ, J. M.; LÓPEZ, J. L.; MATEO, P. L.; VILANOVA, M.; SERRA, M. A.; AVILÉS, F. X. (1991). «Analysis of the thermal unfolding of porcine carboxypeptidase A and its functional pieces by differential scanning calorimetry». *Eur. J. Biochem.*, núm. 176, pàg. 225-230.
- SALL, A.; BLUNDELL, T. L. (1993). «Comparative Protein Modelling by satisfaction of spatial restraints». *J. Mol. Biol.*, núm. 234, pàg. 779-815.
- SIPPL, M. J.; WETTKUS, S. (1992). «Detection of native-like models for amino acid sequences of unknown three-dimensional structure in a data base of known protein conformations». *Proteins: Struct. Func. Genet.*, núm. 13, pàg. 258-271.
- SMITH, M. (1983). «In vitro mutagenesis». *Ann. Rev. Genet.*, núm. 19, pàg. 423-462.
- TOPHAM, C. M.; THOMAS, P.; OVERINGTON, J. P.; JOHNSON, M. S.; EISENMENGER, F.; BLUNDELL, T. L. (1991). «An assesment of COMPOSER: a rule based approach to modelling protein structure». *Biochem. Soc. Symp.*, núm. 57, pàg. 1-9.
- TRIMBUR, D. E.; WARREN, R. A. J.; WITHERS, S. G. (1992). «Region-directed mutagenesis of residues surrounding the active site nucleophile in β -glucosidase from *Agrobacterium faecalis*». *J. Biol. Chem.*, núm. 267, pàg. 10248-10251.
- VIGUERA, A. R.; VILLEGAS, V.; AVILÉS, F. X.; SERRANO, L. (1996). «Native-like favourable helical local interactions can accelerate protein folding». *Fold. Des.*, núm. 2, pàg. 23-33.
- VILLEGAS, V.; AZUAGA, A.; CATASÚS, LL.; REVERTER, D.; MATEO, P. L.; AVILÉS, F. X.; SERRANO, L. (1995a). «Evidence for a two-state transition in the folding process of the activation domain of human procarboxypeptidase A2». *Biochemistry*, núm. 34, pàg. 15105-15110.
- VILLEGAS, V.; VIGUERA, A. R.; AVILÉS, F. X.; SERRANO, L. (1995b). «Stabilization of proteins by rational design of α -helix stability using helix-coil transition theory». *Folding and Design*, núm. 1, pàg. 29-34.
- WITHERS, S. G.; STREET, I. P. (1988). «Identification of a covalent α -D-glucopyranosyl enzyme intermediate formed on a β -glucosidase». *J. Am. Chem. Soc.*, núm. 110, pàg. 8551-8553.