

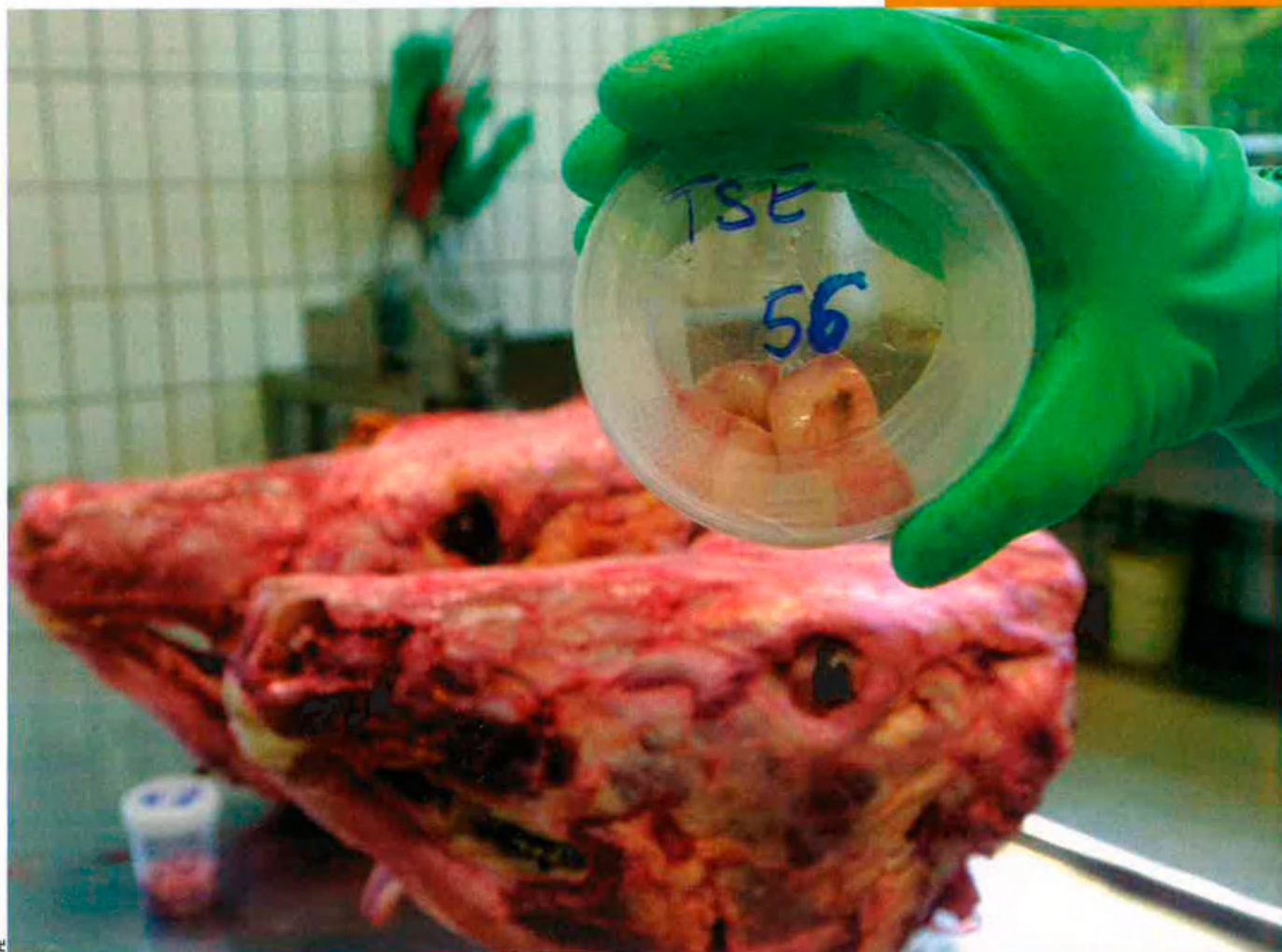
Del genoma al proteoma

Els estudis de genòmica han proporcionat una enorme quantitat de dades sobre la informació genètica dels organismes. La darrera publicació de l'esborrany complet del genoma humà sembla confirmar que les ordres per desenvolupar un individu humà només es troben en uns 35.000 gens. Aquesta dada no minva la nostra complexitat. El repte actual és comprendre com s'hi arriba amb només el doble de gens que un cuc. Tot apunta que les proteïnes, les molècules realment funcionals en els organismes vius, en són les responsables. I molts investigadors apunten al proteoma.

Què és un proteoma? La paraula "proteoma" deriva de *proteïnes* expressades per un *genoma*, i fa referència a totes les proteïnes produïdes per un organisme, de la mateixa manera que el genoma és tot el conjunt de gens. El cos humà pot contenir més de dos milions de proteïnes diferents, cada estirp amb una funció diferent.

Les proteïnes són molècules orgàniques complexes d'elevat pes molecular formades per diferents unitats que es repeteixen: els aminoàcids. D'aminoàcids només n'hi ha 20; la variació de les seqüències dóna desenes de milers de proteïnes, cada una amb forma i funció específiques. L'ordre dels aminoàcids ve determinat pel gen que codifica la proteïna. Ara bé, el pas de la informació genètica a la formació d'una proteïna no és una via directa. Perquè la informació passi dels gens a les proteïnes cal un mitjancer en el procés: l'àcid ribonucleic (RNA). A grans trets, el DNA és transcrit a RNA i l'RNA traduït a proteïna, però no és immediat. (Vegeu la figura de la pàgina 48.)

Després de la transcripció, l'RNA encara ha de madurar per a servir de motlle per a la síntesi de proteïnes, i aquesta maduració comporta una selecció de la informació. Només part de la cadena d'RNA transcrita serà traduïda a proteïna: el procés de se-



lecció de fragments d'RNA que sí que seran traduïts a proteïna en termes tècnics se'n diu empalmament (*splicing*). Un cop l'RNA és madur, es dirigeix als ribosomes, orgànuls intracel·lulars on té lloc la traducció: la síntesi de proteïnes segons la informació de l'RNA i les constriccions del codi genètic, que a un triplet de nucleòtids hi assigna un aminoàcid.

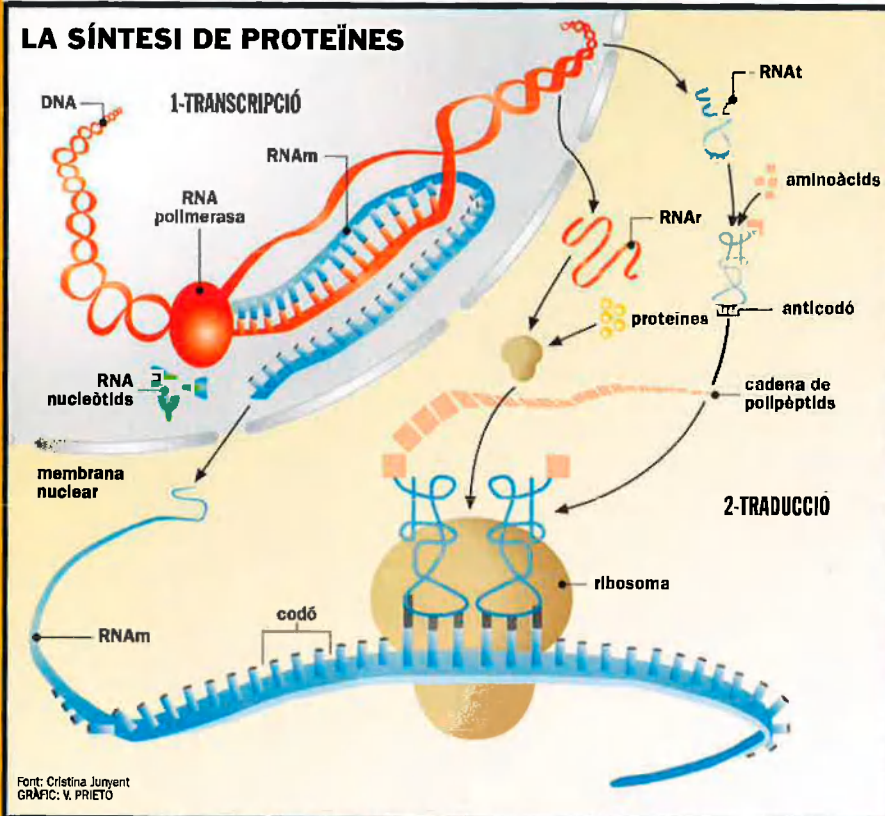
I, encara més, un cop s'ha traduït la seqüència d'una proteïna, que pot estar formada per centenars o milers d'aminoàcids, també ha de seguir un procés de maduració: es pot veure sornesa a una varietat de modificacions. En aquest procés hi pot haver l'addició d'àtoms metàl·lics, grups de carbohidrats o fosfats; també poden unir-se diverses proteïnes. I, per descomptat, les proteïnes són plegades per adoptar l'estructura funcional tridimensional. (Vegeu la figura de la pàgina 51.)

La manera com es pleguen les proteïnes per adoptar la seva forma funcional és coneguda com el segon codi genètic, i desxifrar-lo és un dels reptes més notables de la ciència actual. La forma de les proteïnes és molt important, tant com la composició. Això va quedar clar amb la malaltia de Creutzfeldt-Jacob, provocada per variació en la forma d'una molècula. Un cop la proteïna és madura, es dirigeix a la part concreta de la cèl·lula on és requerida, o bé, la cèl·lula on ha estat sintetitzada l'empaqueta per tal que pugui atribar a una altra cèl·lula o a una altra part de l'organisme.

L'exquisida estructura de les proteïnes és la que els permet de dur a terme nombroses tasques complexes, com ara la catàlisi de reaccions químiques com a enzims; com a missatgers, per exemple, neurotransmissors; elements clau que controlen la divisió cel·lular; factors tròfics que regulen el creixement i desen-

La malaltia de les vaques boges és un exemple de la importància de la forma de les proteïnes, ja que estava provocada per la variació en una molècula.

La manera com es pleguen les proteïnes per adoptar la seva forma funcional és coneguda com el "segon codi genètic".



bientals intra i extracel·lulars. El proteoma varia en estat de salut o de malaltia, el tipus de teixit, el moment del desenvolupament i els efectes de tractaments farmacològics. Per això, el proteoma sovint es defineix com "el conjunt de proteïnes presents en una mostra (teixit, organisme, cultiu cel·lular) en un moment determinat".

En molts sentits, l'estudi del proteoma va paral·lel al del genoma: la genòmica comença en el gen i infereix sobre els productes, les proteïnes. La proteòmica, al seu torn, comença estudiant la proteïna modificada per a ser funcional i busca el gen responsable de la seva producció. La seqüenciació del genoma humà ha fet créixer l'interès per la proteòmica, perquè mentre que la informació de la seqüència de DNA proporciona una instantània de les diferents maneres en què la cèl·lula pot fer servir les proteïnes, la vida de la cèl·lula és un procés dinàmic. Aquestes noves dades ofereixen noves promeses per les aplicacions de la proteòmica en ciència, medicina i, més notablement, farmacologia.

volupament de teixits; el transport d'oxigen a la sang, amb l'hemoglobina; la defensa de l'organisme davant infeccions, com ara els anticossos... i també en la formació d'estructures, com els cabells o les ungles, els músculs... fins a entorn de dos milions de proteïnes diferents que hi ha en un ésser humà.

Com s'hi arriba, a aquesta xifra? Si s'estima que els humans tenim entorn de 35.000 gens, potencialment tenim informació per codificar 35.000 proteïnes. Per l'empalmament (*splicing*) diferencial de les molècules d'RNA, la informació a partir d'un sol gen pot codificar més de 50 espècies de proteïnes. I per la maduració proteica que segueix la traducció, el nombre de proteïnes o fragments pot pujar fins a vora dos milions. Així doncs, el proteoma és molt més complex que el genoma.

Quina és la diferència entre el genoma i el proteoma? El nombre de molècules no és l'única diferència entre el genoma i el proteoma. A diferència del genoma, que és relativament estàtic, el proteoma canvia constantment en resposta a desenes de milers de senyals am-

Què estudia la proteòmica? Els investigadors treballen per desenvolupar el mapa del proteoma humà –de la mateixa manera que es fa amb el genoma humà– que identifiqui noves famílies de proteïnes, interaccions entre les proteïnes i rutes de senyalització. La proteòmica descriu com les proteïnes fan la feina necessària per tal de fer funcionar la cèl·lula viva. Òbviament, el nom va aparèixer després del de genòmica, i, de la mateixa manera que el genoma és la informació completa del DNA, el proteoma és el complement total de proteïnes d'una cèl·lula.

Una cèl·lula viva és un lloc molt actiu, amb moltes molècules actives. La major part del contingut d'una cèl·lula és format per proteïnes estructurals, enzims i transportadors moleculars. La proteòmica descriu l'estructura, la funció i les interaccions de les molècules proteiques i la interacció d'aquestes proteïnes amb altres molècules. No cal dir que sense potents ordinadors la proteòmica seria una ciència impossible d'estudiar, ja que els detalls de les operacions de la vida es poden descobrir amb models detallats

de les molècules i simular-ne les interaccions.

Les tecnologies associades a la proteòmica representaran un paper important en el descobriment de nous marcadors proteics amb finalitat diagnòstica i de noves dianes moleculars per a nous fàrmacs en el camp de la medicina molecular, perquè és l'enllaç entre gens, proteïnes i malalties. Com que els investigadors estudien les proteïnes defectuoses que provoquen malalties concretes, les seves troballes poden ajudar a desenvolupar nous fàrmacs que bé alterin la forma d'una proteïna defectuosa o bé en mimetitzin una que falti. Molts dels fàrmacs actualment més eficaços o són proteïnes o actuen sobre proteïnes diana. Els avenços en proteòmica poden ajudar els científics a crear medicacions que a la llarga es podran personalitzar amb característiques d'efectivitat superior i amb menys efectes secundaris.

La recerca en proteòmica. L'estudi de la composició, estructura i funció de les proteïnes ocupa un terreny entre la biologia i la química i fa servir mètodes d'ambdues. Entre els mètodes de laboratori hi ha, en primer lloc, l'accés als fluids i els teixits mostra, i l'obtenció de mostres. Actualment, les mostres més comunes en proteòmica són de malalties oncològiques, neurològiques cardiovasculars o endocrinològiques, com ara la diabetes. La separació, la identificació i la quantificació de totes les proteïnes expressades en cèl·lules, teixits i organismes és un dels grans reptes de l'era postgenòmica.

Composició de les proteïnes. Sovint la complexitat de la mostra emmascara els seus components minoritaris, en favor de les proteïnes més abundants; per això, abans de la identificació, cal separar les proteïnes, cosa que pot ser comparable a trobar una agulla en un paller. Però les tècniques de laboratori per separar proteïnes ho permeten. Per trobar la composició d'aminoàcids d'una proteïna o d'un pèptid, encara és vàlid el mètode Sanger, que consisteix bàsicament a trencar còpies d'una mateixa proteïna en fragments diferents, estudiar-ne la composició i refer el trencament.

ajuts Universitaris 2003



caixaManresa

Dotació de 300.000 euros

Informa-te'n a qualsevol oficina
o bé a www.caixamanresa.es/fundacio

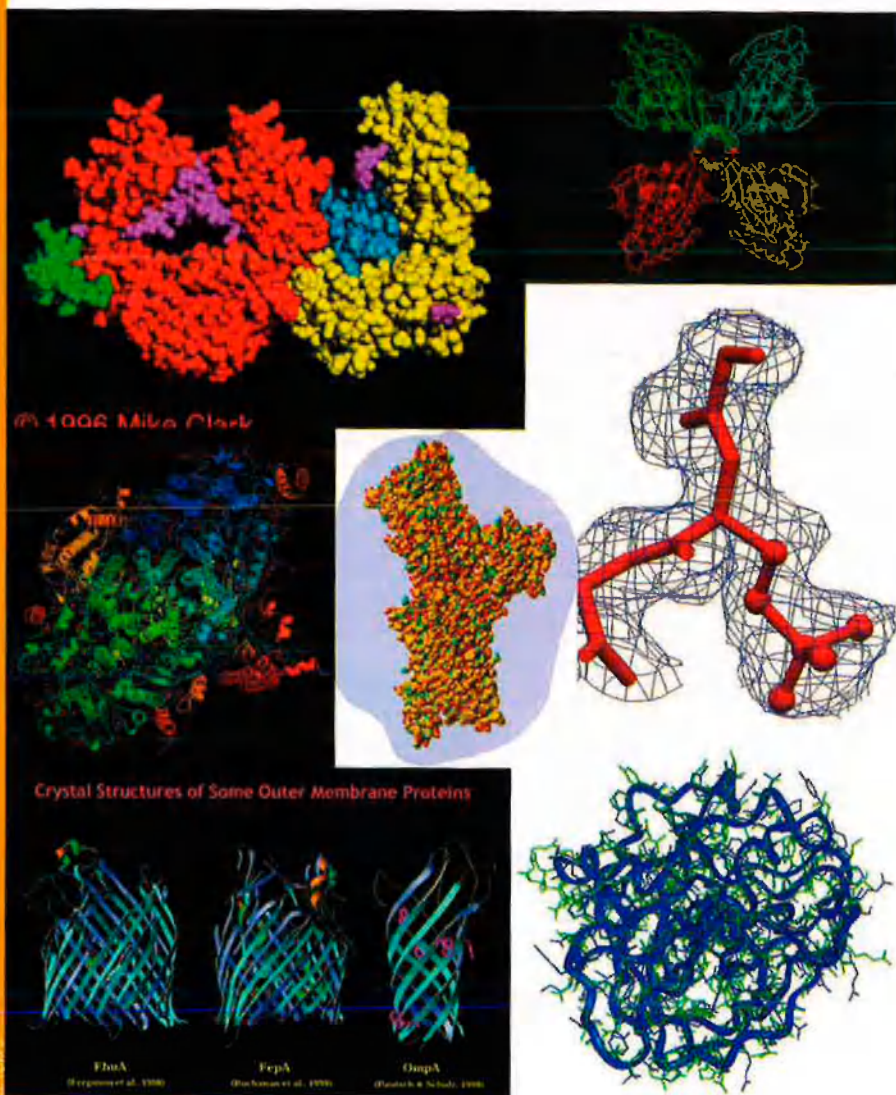
ORANGE ENERGY



Amb el suport de



Generalitat de Catalunya
Departament d'Universitats, Recerca
i Societat de la Informació



que es produeixen són carregats positivament (cations) i, quan són accelerats en el buit per l'acció d'un camp magnètic, poden ser separats segons la relació entre les seves massa i càrrega. Així, doncs, l'anàlisi d'un espectre de masses consisteix a extreure informació per a unir els diferents fragments i reconstruir la molècula original.

Un dels grups que estudia la composició d'aminoàcids d'una proteïna, i com d'un mateix fragment de gen per diferències en l'empalmament es poden obtenir diferents proteïnes, és el grup de Joaquín Abian, a la Unitat d'Espectrometria de Masses Estructural i Biològica del Departament de Bioanàlisi Mèdica de l'Institut d'Investigacions Biomèdiques de Barcelona (IIBB) CSIC-IDIBAPS. Conèixer la composició de les proteïnes pot ser molt valuós a l'hora de comparar una proteïna funcional amb una de patològica. Una diferència funcional reflecteix una diferència molecular, i una diferència molecular pot ser detectada per tècniques de separació de proteïnes, com ara l'electroforesi. Tenir separades les proteïnes permet identificar la diferència entre les molècules. Comparant amb bases de molècules conegudes es pot classificar la patologia, cosa necessària per determinar el possible tractament. Si el tractament és possible es basarà en inhibidors de funció, si la malaltia és per sobreexpressió; o en potenciadors, si és per defecte.

També cal recordar que, si bé tota la informació genètica és en cada una de les cèl·lules, de cada una de les estirps només se n'expressa una part. En la Unitat de Proteòmica i Química de Pèptids, creada en l'entorn Universitat Pompeu Fabra, Institut Municipal d'Investigació Mèdica i Centre de Regulació Genòmica, dirigit per David Andreu, s'anàliza l'expressió diferencial en diversos teixits humans i en microorganismes, com ara la bactèria *Leishmania*.

Estructura de les proteïnes.

Conèixer la composició i l'estructura d'una molècula és un progrés, però fins que no s'entén l'estructura, especialment en el cas de les proteïnes, no se sap com duu a terme la seva funció. L'es-

clusques per trobar la seqüència sencera. Enguany fa cinquanta anys que Frederick Sanger va aplicar el mètode per determinar la seqüència d'aminoàcids d'una proteïna: la insulina. Per això va rebre el premi Nobel el 1958. I és dels pocs científics que n'ha rebut dos: l'any 1980 va rebre el segon, per les seves contribucions en la determinació de la seqüència de les bases en els àcids nucleics.

Més recentment, una tècnica de laboratori associada a potent maquinari permet també conèixer la seqüència d'aminoàcids d'una proteïna de la qual es té informació prèvia: l'espectrometria de masses. En l'espectrometria de masses, la substància es bombardeja amb un feix d'electrons amb suficient energia per fragmentar les molècules. Els fragments

estructura és el determinant més important de les propietats d'una molècula. Conèixer-la permet plantejar i buscar resposta a determinades preguntes. Com es pleguen les proteïnes en el substrat natural? Què determina l'estabilitat de les proteïnes? Quines forces determinen l'associació de les proteïnes amb els substrats? Com transformen l'energia química en força?

Per comprendre l'estructura d'una proteïna es fan servir majoritàriament dues tècniques: ressonància magnètica i cristal·lografia. Les tècniques de ressonància magnètica nuclear, més modernes, proporcionen una informació més propera a la realitat, perquè les proteïnes es troben en solució gairebé fisiològica. El grup d'Ernest Giralt, de l'Institut de Recerca Biomèdica de Barcelona del Parc Científic de la Universitat de Barcelona, aplica aquesta tècnica per estudiar les propietats dinàmiques de les proteïnes i el disseny de molècules que s'hi lliguen i interactuen als seus centres reactius.

La tècnica més clàssica i més generalitzada per a estudiar l'estructura terciària (tridimensional) de les proteïnes és l'aplicació dels raigs X sobre proteïnes cristal·litzades. Els resultats trobats per aquesta tècnica i el seu estudi en sincrotrons han experimentat un creixement sense precedents en els darrers anys. La cristal·lografia permet determinar posicions atòmiques i, en conseqüència, observar el plegament de la cadena polipeptídica. A més, d'una proteïna que cristal·litza és fàcil d'estudiar-ne el centre de reacció, i dissenyar agonistes o antagonistes que potenciïn o frenin l'acció de la molècula. Amb l'ajut de potents ordinadors, esbrinar l'estructura funcional i el plegament de les proteïnes permet avui predir-ne la dinàmica funcional. Aquests estudis han fet de la biologia estructural una nova disciplina.


El desenvolupament de la informàtica proporciona una potent eina en la determinació de l'estructura terciària de les proteïnes: el seu plegament. El Protein Data Bank és el receptor mundial de les dades acumulades sobre estructures tridimensionals macromoleculares, el creixement exponencial de les quals és notable. A la Unitat de Biologia Computa-

ESTRUCTURA D'UNA PROTEÏNA


PRIMÀRIA

$$\begin{array}{ccccccc} & \text{H} & \text{O} & & \text{H} & \text{O} & & \text{H} & & & \\ & | & || & & | & || & & | & & & \\ \text{H}_2\text{N} & -\text{C} & -\text{C} & -\text{NH} & -\text{C} & -\text{C} & -\text{N} & -\text{C} & -\text{COOH} \\ & | & & & | & & & | & & & \\ & \text{R} & & & \text{R} & & & \text{H} & & & \\ & & & & & & & \text{M} & & & \\ & & & & & & & \text{R} & & & \end{array}$$

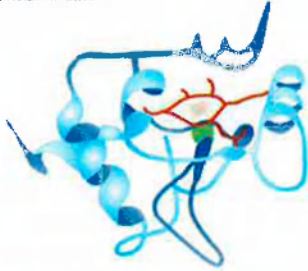
SECUNDÀRIA



TERCIÀRIA



QUATERNÀRIA



Font: Cristina Junyent
Gràfic: V. PRIETO

cional de l'Institut de Biotecnologia i de Biomedicina a la Universitat Autònoma de Barcelona, coordinat per Xavier Daura, desenvolupen metodologies per a estudiar els mecanismes de plegament de pèptids i proteïnes, bàsics per al funcionament.

A partir del coneixement de l'estructura d'una proteïna cristal·litzada també es poden buscar models computacionals per comprendre'n la dinàmica. Aquestes tècniques complexes i combinades pro-

Una proteïna es defineix per quatre nivells estructurals: les estructures primària, secundària, terciària i quaternària.

La primària ve donada pels aminoàcids que la formen i pel seu ordre.

Amb la síntesi, els aminoàcids es disposen de manera estable en l'espai: l'estructura secundària,

amb forma d'hèlix o de full plegat. L'existència d'enllaços entre els aminoàcids crea l'estructura terciària.

Diverses estructures terciàries (cadena polipeptídiques) enllacen les unes amb les altres i formen l'estructura quaternària



Enguany fa 50 anys que Frederick Sanger, a la dreta, va determinar la seqüència d'aminoàcids d'una proteïna, concretament la insulina. Va rebre el premi Nobel de química el 1958 i el 1980.

porcionen la comprensió dels processos que tenen lloc en els éssers vius. La simulació per ordinador dels sistemes biològics permet, per exemple, comparar molècules patològiques i molècules sanes; a partir d'aquesta informació es poden buscar substrats que actuin sobre el centre actiu. El grup de Modesto Orozco, al Departament de Bioquímica i Biologia Molecular de la Facultat de Química, de la UB, estudia la dinàmica de proteïnes per models moleculars, buscant el perfil de les proteïnes segons els seus centres actius, on té lloc la unió de les molècules i la seva interacció química.

Una altra aplicació de les tècniques de modelització enfocades a comprendre la dinàmica de les molècules és el disseny de noves molècules amb una funció determinada. El grup de Juan Jesús Pérez, del Departament d'Enginyeria Química de la Universitat Politècnica de Catalunya, fa recerca sobre pèptids biològicament actius i sobre el disseny de peptidomimètics.

L'aplicació de la tecnologia computacional obre el ventall d'estudi comparatiu de les estructures proteïques i una aproximació al desxiframent del codi estructural de les proteïnes. El codi que relaciona l'estructura primària d'una proteïna (la seqüència) i l'estructura terciària (la forma tridimensional) és lluny de ser trobat; però el grup de Gerard

Pujadas, de la Unitat de Biotecnologia Computacional del Departament de Bioquímica i Biotecnologia de la Universitat Rovira i Virgili de Tarragona, busca la correlació entre l'evolució molecular i l'estructura proteica de pèptids senzills.

Per acabar, un dels cristal·lografs que ha creat escola al nostre país és Joan Antoni Subirana, del Departament d'Enginyeria Química de la Universitat Politècnica de Catalunya. L'equip de Subirana ha estudiat l'estructura cristal·logràfica del DNA i les proteïnes que l'empaqueten, com també l'estructura cristal·lina de polímers peptídics sintètics i la seva relació amb la patogènia.

L'abundància d'informació que proporciona l'estudi del proteoma és complementària a la informació genètica generada per la recerca genòmica. De fet, la proteòmica serà una contribució clau per al desenvolupament de la genòmica funcional. I, recíprocament, la combinació dels coneixements acumulats per la genòmica i la proteòmica representarà un paper fonamental en la recerca biomèdica i, en un futur, arribarà a tenir un impacte molt significatiu en el desenvolupament de productes diagnòstics i terapèutics.

*Cristina Junyent
Universitat Pompeu Fabra
Societat Catalana de Biologia*