

BIOLOGIA ESTRUCTURAL: TRENTA ANYS FILANT PRIM

LLUÍS RIBAS I DE POUPLANA

The Scripps Research Institute. La Jolla, EUA.

Adreça per a la correspondència: The Scripps Research Institute BCC-379. 10550 North Torrey Pines Rd. La Jolla, CA 92117, EUA. Adreça electrònica: Lluis@hermes.scripps.edu

INTRODUCCIÓ

Ara fa exactament trenta anys (el 1968) es proclamava la fi de la primera era daurada de la biologia molecular (Stent, 1968). Aquest període havia començat aproximadament a la fi dels anys trenta, amb els primers estudis sobre la biologia del bacteriòfag lambda (Luria i Delbrück, 1943), i les anàlisis de les característiques físiques de proteïnes i DNA (Franklin i Gosling, 1953; Kendrew, 1950). El descobriment de l'estructura del DNA el 1953 (Watson i Crick, 1953) va representar la culminació d'aquesta fase de la biologia molecular, que va acabar amb la determinació del codi genètic universal i la definició aproximada dels mecanismes de replicació, transcripció i traducció genètics, a finals dels anys seixanta (Sagdopal, 1968).

Al cap de dos anys altres autors anunciaven el començament de la segona edat d'or de la biologia molecular (Lewin, 1970). Aquest segon renaixement va coincidir amb

la publicació de les tertúlies dedicades al futur de la biologia moderna, organitzades per la Societat Catalana de Biologia, el 1970 (Margalef *et al.*, 1971). Llegida avui, la transcripció de les tertúlies reflecteix aquest estat de transició en què es trobava la comunitat d'investigadors. Per una banda, en les converses es detecta la convicció general que els principis fonamentals de la vida havien estat descoberts. Per l'altre costat, la manca d'informació estructural i funcional detallada sobre aquests mateixos principis dóna a les converses un aire preliminar. Sembla clar que un capítol acabava de tancar-se, i que la tasca de refinar els descobriments assolits fins al moment encara no havia començat.

Els trenta anys que han seguit han generat una explosió de coneixements biològics sense precedents, atida pel desenvolupament de les tècniques d'enginyeria genètica. Entre les disciplines que han experimentat un creixement més accelerat cal incloure-hi aquella dedicada a l'estudi de l'estructura i

de les interaccions físiques entre biomolècules, que aquí anomenarem biologia estructural.

Una revisió completa dels avenços de la biologia estructural en els últims trenta anys és una tasca excessiva, tant per al revisor com per al lector. Aquest article és, per tant, una perspectiva personal d'alguns dels aspectes que (potser) caracteritzen millor els últims trenta anys de recerca en aquest camp.

He dedicat una primera part a l'evolució tecnològica de la qual s'ha beneficiat l'estudi estructural de les biomolècules. En particular descriuré breument les millores en el camp de la cristal·lografia i la ressonància magnètica nuclear. Faré menció de l'evolució dels mètodes computacionals per a l'anàlisi de proteïnes i àcids nucleics, però per una discussió més detallada és recomanable llegir la revisió dedicada a la biocomputació, en aquest mateix volum.

La segona part de l'article és una limitada revisió de dues qüestions que representen bé l'evolució de la investigació estructural en els últims trenta anys. Donaré, sobretot, exemples de molècules relacionades amb els mecanismes de replicació i expressió gènica. Aquest camp de recerca resulta especialment adequat, perquè els avenços en aquesta disciplina marquen els límits dels dos períodes de la biologia molecular mencionats abans. La seva trajectòria encaixa amb el període de temps que ens interessa, i podria ser un model per predir (si el lector s'hi veu amb cor) la futura evolució d'altres camps de la biologia.

AVENÇOS TECNOLÒGICS

La cristal·lografia de raigs X, la ressonància magnètica nuclear (RMN), i la biocomputació són les tècniques més utilitzades en recerca estructural. Altres mètodes, com la

difracció d'electrons (Dorset, 1996), la ultracentrifugació analítica (Hansen *et al.*, 1994), i diverses tècniques d'espectroscòpia (Ferreira i Coelho-Sampaio, 1996; Kelly i Price, 1997), són també útils, però no generen informació tan detallada com les tres primeres.

L'increment en la quantitat d'informació estructural generada per aquestes tècniques en els últims trenta anys queda reflectit pel creixement del banc de dades d'estructures tridimensionals de macromolècules biològiques (PDB). Aquest magatzem digital de coordenades tridimensionals de proteïnes i àcids nucleics va ser fundat el 1971, amb les estructures de cinc proteïnes diferents (Meyer, 1997). Vint-i-set anys després el PDB conté 7.373 estructures obtingudes per cristal·lografia de raigs X, 1.388 estructures obtingudes per RMN, i 201 estructures construïdes amb mètodes computacionals. Actualment, el volum del PDB augmenta a un ritme aproximat de 100 estructures al mes (vegeu <http://pdb.pdb.bnl.org>)

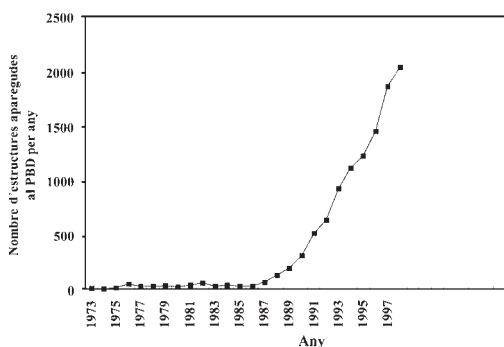


FIGURA 1. Creixement del banc de coordenades tridimensionals de biomolècules (PDB). Nombre d'estructures (coordenades) dipositades per any, des de 1973.

Abans d'entrar en detall en les millores que aquestes tècniques han experimentat en els últims trenta anys cal destacar que un dels principals factors que han contribuït a aquests avenços ha estat el desenvolupa-

ment dels mètodes d'enginyeria genètica. La revolució del DNA recombinant ha permès, no només aïllar i seqüenciar informació genètica, sinó també expressar gens i purificar-ne els productes en quantitats i amb nivells de puresa molt superiors al que era possible abans. Aquests avenços han permès també la generació de proteïnes modificades i marcades (ja sigui amb isòtops o amb aminoàcids poc freqüents, com la selenocisteïna). Com veurem més endavant, aquestes molècules mutades o marcades faciliten en gran manera els seu estudi estructural.

La cristal·lografia de raigs X

La cristal·lografia de raigs X és l'única de les tres tècniques que discutiré que ja havia proporcionat les primeres estructures de proteïnes a escala molecular fa trenta anys. La primera estructura tridimensional obtinguda per una proteïna va ser la de la mioglobina, resolta per Kendrew i col·laboradors, el 1958 (Kendrew *et al.*, 1958). La primera estructura tridimensional d'una molècula d'àcid nucleic (l'RNA de transferència), va ser obtinguda el 1973 (Kim *et al.*, 1973). Des d'aleshores l'eficiència i la sensibilitat de la tècnica han augmentat espectacularment.

Els principis teòrics de la cristal·lografia es basen en la relació matemàtica entre l'estructura d'una molècula i el patró de difracció que la mateixa molècula produeix en ser exposada a una font de radiació. L'observació d'un objecte en un microscopi segueix el mateix principi, en aquest cas la difracció de la llum provocada per l'objecte observat és captada (després de ser enfocada a través de prismes òptics) per l'ull humà.

L'anàlisi per difracció d'estructures moleculars requereix radiacions de longitud d'ona suficientment petita, i d'energia suficientment alta, per provocar difracció a escala molecular. En el cas de la cristal·lografia

convencional la radiació utilitzada són els raigs X, i els elements moleculars que generen la difracció són els electrons de les molècules irradiades. Per poder obtenir un patró de difracció amb suficient energia per ser detectat, la cristal·lografia de raigs X requereix que la molècula que s'estudiï hagi, prèviament, format un cristall uniforme. D'aquesta manera les difraccions de tots els àtoms equivalents en totes les molècules del cristall es consoliden, i formen un senyal que pot ser observat i mesurat amb un detector.

Gràcies a les relacions matemàtiques entre el patró de difracció i l'estructura molecular de l'objecte irradiat, és teòricament possible calcular la distribució espacial d'electrons en la molècula cristal·litzada. Un cop la distribució dels electrons (o densitat electrònica) de la molècula ha estat calculada, és possible reconstruir l'estructura atòmica de la molècula fent-ne encaixar les unitats (aminoàcids, o nucleòtids) dins de la densitat electrònica.

El procés de cristal·lització de proteïnes o d'àcids nucleïcs és el primer escull que cal superar a l'hora d'estudiar aquestes molècules per difracció de raigs X. La qualitat dels cristalls obtinguts és el principal factor que determina la resolució de l'estructura final (McPherson, 1976). En essència, el procés de cristal·lització d'una macromolècula és, encara avui, un procés empíric molt sovint limitat per la quantitat i la puresa del material que s'ha de cristal·litzar. En aquest sentit, com hem dit abans, les noves tècniques de clonatge, expressió i purificació de proteïnes i àcids nucleïcs han suposat un gran avenç, perquè permeten l'obtenció de grans quantitats de material en un estat de moltíssima puresa (McPherson, 1976).

Un altre aspecte de la cristal·lografia que ha estat millorat en els últims trenta anys és el referent a la producció i detecció dels patrons de difracció de les molècules cristal·litzades. Moltes d'aquestes millores són degu-

des a avenços procedents de la recerca en física de partícules, que ha permès obtenir fonts de radiació més potents i uniformes (Moffat i Ren, 1997). Una altra millora substancial ve de l'increment en el poder computacional, que ha permès accelerar el gran nombre de càlculs requerits per la construcció dels mapes de densitat electrònica a partir dels patrons de difracció. Per altra banda, la millora i l'abaratiment dels sistemes de visualització gràfica en tres dimensions (desenvolupats a partir dels anys setanta) han facilitat en gran manera la construcció i la difusió dels models cristal·logràfics. El 1972, per exemple, els únics aparells disponibles per dur a terme aquestes operacions (que avui estan a l'abast de qualsevol ordinador personal) tenien un preu aproximat de set-cents mil dòlars de l'època (al canvi actual, uns cent milions de pessetes. Meyer, 1997).

La cristal·lografia ha experimentat també importants avenços metodològics i teòrics. La principal limitació del càlcul de densitats electròniques és el fet que, a partir de la difracció que es recull d'un cristall, és impossible obtenir un dels paràmetres essencials per a les posteriors operacions matemàtiques. Aquest paràmetre és la fase de cada difracció, que no es pot mesurar, perquè cada difracció prové de tots els àtoms estructuralment equivalents en el cristall (Tivol, 1995). L'estratègia habitual per obtenir les fases d'un patró de difracció es basa en la introducció, en els cristalls, d'àtoms de gran densitat electrònica, que són detectats per la seva potent difracció, i utilitzats per calcular les fases de les difraccions dels àtoms del seu voltant.

Aquest mètode, conegut com a reemplaçament isomorf, requereix que els àtoms de gran densitat s'incorporin a un nombre limitat de posicions específiques en la molècula, sense afectar ni la conformació d'aquesta ni l'estructura del cristall. En els últims quinze anys han aparegut diversos mètodes alternatius que permeten obtenir les

fases de difracció amb més facilitat en utilitzar els avenços en els mètodes d'enginyeria genètica. Un elegant exemple ve donat pel mètode desenvolupat per Hendrickson i col·laboradors, el 1985, per incorporar àtoms pesants en la seqüència de proteïnes a mesura que aquestes són sintetitzades dins la cèl·lula, substituint residus específics de cisteïnes per l'aminoàcid selenocisteïna (Hendrickson *et al.*, 1985). La selenocisteïna funciona, en essència, com un àtom pesant amb gran potència de difracció. La mutació específica de gens per incorporar aquest tipus d'aminoàcids a posicions determinades permet, doncs, la incorporació d'un àtom pesant directament al cristall.

La ressonància magnètica nuclear

La ressonància magnètica nuclear (RMN) és el segon mètode més important a l'hora d'obtenir estructures moleculars de macromolècules (Wagner, 1997). A grans trets, la tècnica consisteix a definir la posició relativa de cada nucli de la molècula d'interès, a partir de la identificació del altres nuclis de la mateixa molècula amb els quals interacciona. La RMN va començar a ser aplicada per la resolució d'estructures tridimensionals de macromolècules a mitjan dècada dels setanta (Wagner, 1997). Les primeres estructures detallades de proteïnes resoltes per RMN no van ser publicades fins a principis dels vuitanta (Williamson *et al.*, 1985). Avui una cinquena part de totes les estructures tridimensionals disponibles al PDB provenen d'experiments de RMN. Les millores tècniques en aquest camp han estat també molt importants, i han incrementat tant el grau de resolució de la RMN com el pes molecular de les molècules a l'abast de la tècnica (Clare i Gronenborn, 1998).

Tots els nuclis atòmics posseeixen un cert moment magnètic que ve determinat per la seva distribució de càrregues. Sota la

influència d'un camp magnètic extern, el moment nuclear s'orienta respecte al camp extern en la direcció més energèticament estable. Aquest estat pot ser alterat amb radiacions electromagnètiques. L'energia requerida per alterar l'orientació inicial dels moments magnètics dels nuclis és específica per a cada tipus de nucli, i depèn del seu entorn molecular (Roberts, 1993).

L'objectiu dels experiments de RMN en biologia estructural és identificar cada nucli individual de la molècula que s'estudia, els nuclis que l'envolten i la posició tridimensional relativa de tots. Aquesta identificació es basa en les característiques energètiques del moment magnètic de cada nucli. Els principals avantatges de la tècnica són tres. En primer lloc, les estructures obtingudes corresponen als estats en solució, sense els possibles artefactes de les interaccions entre molècules cristal·litzades. En segon lloc, l'anàlisi de les interaccions físiques entre els nuclis proporciona una imatge dinàmica, que permet identificar les característiques de mobilitat de les molècules. Finalment, les tècniques de RMN són molt sensibles a interaccions moleculars de baixa afinitat, i resulten molt útils per a l'estudi de les interaccions intermoleculars, ja siguin entre inhibidors moleculars i les seves dianes, o en processos de reconeixement molecular de dues biomolècules (Roberts, 1993).

D'una manera equivalent a la cristal·lografia, el desenvolupament de mètodes d'expressió de proteïnes en bacteris ha permès també millorar molt els resultats dels experiments de RMN. Un primer avenç ha estat la conseqüència de les millores en quantitat i qualitat del material disponible. D'altra banda, la possibilitat d'expressar proteïnes o àcids nucleïcs en bacteris crescuts en medis enriquits amb isòtops químics (especialment ^{15}N i ^{13}C) ha permès purificar molècules marcades amb aquests isòtops, la qual cosa facilita el procés d'i-

dentificació dels seus nuclis (Kainosho, 1997).

La biocomputació

Els principis teòrics de la biocomputació estructural comencen a establir-se a principis dels anys setanta, quan les primeres anàlisis estadístiques de les estructures cristal·logràfiques de proteïnes es duen a terme (Bajorath *et al.*, 1993; Rost i Sander, 1994; Rost *et al.*, 1993). Les tècniques de biocomputació estructural es basen en la utilització de dades experimentals per a la predicció de l'estructura de biomolècules a partir de la seva seqüència. Per aquesta raó, el seu desenvolupament es troba necessàriament lligat al creixement dels bancs de dades de seqüències i estructures, tant de proteïnes com d'àcids nucleïcs. D'altra banda, les tècniques de modelatge s'han pogut beneficiar de la constant millora de les prestacions dels ordinadors disponibles, i de l'aparició de noves tècniques de programació.

La importància de la biocomputació rau, sobretot, en la velocitat amb què permet l'obtenció de models estructurals, útils per a la generació d'hipòtesis experimentals. Aquesta velocitat fa que les tècniques computacionals siguin els únics mètodes viables per a l'anàlisi dels bancs de dades de seqüències, que estan experimentant creixements exponencials. Aquest creixement es deu, sobretot, a la proliferació de projectes de seqüenciació de genomes (al moment d'escriure aquest article el nombre de genomes complets seqüenciats és de divuit. Vegeu <http://www.tigr.org>). El principal banc de dades de seqüències genètiques, el GenBank, conté més de 2.800.000 seqüències, 1.200.000 de les quals han estat incorporades des de l'agost de 1997 (Benson *et al.*, 1998). A aquest ritme de creixement, la mida de GenBank es duplica cada quinze mesos.

Els mètodes biocomputacionals, i el de-

desenvolupament de la xarxa Internet, han posat a l'abast de la majoria d'investigadors la capacitat de *a)* identificar les famílies estructurals a les quals pertany qualsevol nova seqüència, *b)* predir la funció de la molècula i determinar quins residus poden ser importants, i *c)* predir el plegament terciari de noves proteïnes, si l'estructura d'una altra proteïna suficientment homòloga ja ha estat resolta per cristal·lografia o RMN.

La biologia estructural ha deixat de ser la disciplina minoritària, extremadament complicada i difícil de valorar i/o d'utilitzar que devia ser fa trenta anys. Les generacions futures de biòlegs moleculars seran educades en l'ús i l'anàlisi rutinària de dades estructurals. Els experiments biològics, des de l'anàlisi biofísica de macromolècules fins a estudis de genètica de poblacions, es duran a terme en el context d'informació molecular detallada. Aquesta base molecular permetrà una integració del coneixement nova, i una gran interacció de disciplines diferents, tal i com demanava Margalef en el col·loqui de 1970 (Margalef *et al.*, 1971).

TRENTA ANYS DE DESCOBRIMENT

Ara fa tres dècades, els descobriments del codi genètic i dels seus mecanismes de replicació i expressió van marcar un canvi en la història de la biologia. Els principis més fonamentals de la vida havien estat delimitats, i l'esforç dels investigadors es començava a centrar en els detalls d'aquests processos. Lògicament, trenta anys de recerca han incrementat el nostre grau de comprensió de molts aspectes de la biologia molecular i han proporcionat imatges atòmiques de moltes de les interaccions que regulen la divisió i el creixement de les cèl·lules.

Un aspecte d'aquest desenvolupament, impossible de predir a finals dels seixanta, ha estat l'explosió d'aplicacions mèdiques i

tecnològiques derivades del progrés de la biologia molecular. Hi ha hagut, també, sorpreses i descobriments que han canviat qualitativament la nostra percepció de la vida, i de la seva evolució. D'entre aquest grup nombrós de resultats en discutiré dos amb un cert detall: la natura essencialment física dels processos de regulació de la vida i el paper de l'RNA en l'origen de la vida. Tractem doncs d'interaccions moleculars que són física i només física.

Al llarg dels segles la complexitat dels sistemes vius ha sorprès i fascinat tots aquells que han volgut estudiar-los. Aquesta complexitat és prou intimidatòria com per convèncer més d'un il·lustre pensador que la seva base fonamental no podria ser explicada per simples principis físics. Aquestes teories vitalistes han estat una de les víctimes conceptuals de la biologia estructural.

La noció que la vida, com a fenomen, no podria ser explicada per lleis físiques va arribar a ser defensada fins a mitjan segle. El 1932 Niels Bohr («pare» de la física quàntica) argumentava: «la qüestió, de fet, és saber si ens manquen encara per descobrir trets fonamentals de la vida, que haurem de determinar abans de poder explicar-la en termes físics» (traduït a partir de Bohr, 1933). Max Delbrück, deixeble de Bohr i un dels fundadors de la biologia molecular, reafirmava aquesta opinió en escriure el 1949: «és possible que trobem característiques de les cèl·lules vives que no són reductibles a principis físics, i que representen lleis complementàries a les de la física atòmica» (traduït a partir de Delbrück, 1935). Naturalment aquesta opinió tenia, ja aleshores, importants crítics, entre els quals destacava un altre notable físic: Erwin Schrödinger. En el seu llibre *Què és la vida?* (Schrödinger, 1945) Schrödinger escriu: «l'evident incapacitat de la física i de la química d'avui per explicar els fenòmens que tenen lloc en un ésser

viu no és raó per dubtar que aquests processos puguin ser inclosos dins d'aquestes ciències».

El descobriment de l'estructura del DNA i les implicacions mecanicistes per a la comprensió de la transmissió de l'herència van ser els primers revessos importants que va rebre el «vitalisme molecular» (Watson i Crick, 1953). Els models de manteniment de la fidelitat del codi genètic en la seva replicació, regulació i expressió, que van començar a aparèixer a principis dels seixanta, eren ja purament mecanicistes, basats en interaccions físiques entre components cel·lulars (Pardee *et al.*, 1959).

Els primer model de regulació descrit, el de l'operó Lac d'*Escherichia coli*, va demostrar el paper de les interaccions moleculars en la regulació gènica, i va evidenciar la necessitat d'obtenir informació atòmica per arribar a una comprensió total dels sistemes de control biològics (Pardee *et al.*, 1959). En particular, era necessari explicar quin era el mecanisme que permetia a les proteïnes reconèixer seqüències específiques de DNA i RNA.

Aquesta informació a escala atòmica no va començar a arribar fins vint anys més tard. Les primeres estructures tridimensionals de proteïnes d'unió al DNA van ser resoltes a principis dels vuitanta (Pabo i Sauer, 1984). Al mateix temps, tècniques de modelatge computacional i RMN van ser utilitzades per construir els primers models atòmics de les interaccions proteïna-DNA (Pabo i Sauer, 1984). El 1988 les cinc primeres estructures de complexos entre proteïnes reguladores i DNA van ser publicades, una d'elles resolta per RMN (Pabo i Sauer, 1992). El 1989 la primera estructura cristal·logràfica de la interacció entre una proteïna i una molècula de RNA va ser publicada (Rould *et al.*, 1989). A la fi de 1998 el banc de dades PDB conté unes cent vint estructures de complexos entre DNA i proteïnes, i

unes cinquanta de complexos entre proteïnes i RNA.

Amb el detall a escala atòmica va arribar la confirmació que els mecanismes reguladors dels processos biològics estan basats en interaccions atòmiques clàssiques, de tipus electroestàtic i hidrofòbic. La seva complexitat només és la conseqüència del gran nombre total d'interaccions i del complicat equilibri termodinàmic entre cada biomolècula, el conjunt total de molècules presents dins la cèl·lula i el solvent (Pabo i Sauer, 1992).

L'estudi experimental d'aquestes forces (utilitzant tècniques biofísiques, bioquímiques i genètiques), i el disseny de mètodes computacionals per a la simulació molecular, han generat un grau de comprensió relativament bo dels principis fonamentals de l'estructura de biomolècules. La utilització racional d'aquests principis per manipular aquestes interaccions és, avui en dia, un procés rutinari per al disseny de fàrmacs, nous biocatalitzadors, etc. (Hoog *et al.*, 1995; Hunter, 1995; Nagase *et al.*, 1998).

Del món de RNA al món de DNA i proteïnes: nissaga de poder

El descobriment del sistema genètic va generar el famós dogma de la biologia: «del DNA se'n fa RNA, de l'RNA se'n fa proteïna». El dogma donava al DNA el paper de sistema documental, o llibre d'instruccions, i a les proteïnes el rol de catalitzadors, o maquinària. El paper que l'RNA tenia dins d'aquest esquema es limitava a ser sistema de transferència, o intermediari sense cap funció important de control.

Aquesta visió, però, generava una paradoxa sense solució: si el DNA fa les proteïnes i les proteïnes fan el DNA, qui va fer el primer DNA i/o la primera proteïna? Intrigats per la ubiqüitat de l'RNA, Woese, Crick i Orgel van publicar les primeres es-

peculacions sobre el seu possible rol funcional durant l'evolució de la vida (Crick, 1968; Orgel, 1968; Woese, 1967). Es va començar a perfilar la hipòtesi d'un món prebiòtic, on les funcions d'emmagatzematge del material genètic i de catalisi de reaccions metabòliques havien estat dutes a terme per molècules de RNA. Aquests autors, però, no van anticipar la possibilitat que algunes d'aquestes molècules de RNA catalític haguessin sobreviscut fins als sistemes actuals.

Els primers estudis estructurals sobre l'RNA van demostrar que l'estructura tridimensional d'aquesta molècula és molt diferent a la de la doble hèlix del DNA (Kim *et al.*, 1973). L'estructura tridimensional de l'RNA de transferència (tRNA) va revelar una arquitectura globular, amb un plegament complicat, i amb una distribució de grups funcionals més semblant a les proteïnes que al DNA (Kim *et al.*, 1973).

El primer indici experimental de l'existència d'excepcions al dogma general, i de la capacitat de l'RNA per dur a terme altres funcions, va arribar el 1970 (Baltimore, 1970). El descobriment de la retrotranscriptasa inversa, l'enzim responsable de la síntesi de DNA a partir del genoma de RNA dels retrovirus, va demostrar la capacitat de l'RNA per dur a terme funcions de codificador del genoma, normalment associades al DNA.

L'RNA, doncs, podia generar estructures tridimensionals globulars, semblants a les dels enzims, i podia ser utilitzat en lloc del DNA com a material genètic. La peça que faltava per acabar el trencaclosques va arribar el 1982 amb el descobriment pels grups de Cech i Altman de les ribozimes, o molècules de RNA que catalitzen reaccions metabòliques *in vivo* (Guerrier-Takada *et al.*, 1983; Kruger *et al.*, 1982). La capacitat de l'RNA per dur a terme les tres funcions contingudes en el dogma de la biologia va res-

suscitar les hipòtesis sobre el paper de l'RNA en l'evolució inicial de la vida, i el terme *món de RNA* (Gilbert, 1986) es va crear per referir-se a aquesta fase hipotètica de l'evolució.

La dificultat de provar l'existència d'un metabolisme que hauria existit fa més de 3.500 milions d'anys és immensa, i és possible que la hipòtesi del món de RNA no pugui ser mai provada inequívocament. Però el volum de les dades circumstancials continua creixent. A finals dels vuitanta van començar a aparèixer els primers treballs dedicats a la selecció *in vitro* de molècules de RNA amb noves activitats catalítiques. Des d'aleshores els laboratoris de Joyce, Szostak i altres han estat capaços de seleccionar molècules de RNA que catalitzen diverses de les funcions essencials per a la replicació i l'expressió d'un genoma de RNA, inclosa la catalisi de l'enllaç peptídic entre dos aminoàcids (Ellington i Szostak, 1990; Joyce i Orgel, 1986; Zhang i Cech, 1997).

L'estudi dels components de l'aparell de traducció genètica ha aportat dades directes de la relativa antiguitat de les molècules de RNA respecte a les proteïnes en aquest procés metabòlic. El pas central en la transmissió del codi genètic a informació en forma d'aminoàcid és la reacció d'aminoacilació del tRNA per l'enzim aminoacil-tRNA sintetasa (Schimmel i Soll, 1979). Aquesta reacció és l'únic punt del procés de síntesi de proteïnes on la informació continguda en el sistema de codons és transformada a seqüència proteica. La relació de reconeixement estructural entre el tRNA i les aminoacil-tRNA sintetases lliga l'evolució de les dues molècules i permet comparar-ne les antiguitats respectives utilitzant mètodes filogenètics (Brown i Doolittle, 1995; Schimmel i Ribas de Pouplana, 1995). D'aquests estudis es desprèn que l'aparició de les molècules de tRNA, i la seva especificitat per aminoàcid, precedeix l'aparició de les

aminoacil-tRNA sintetases (Ribas de Pouplana *et al.*, 1998). Per tant, en un escenari biòtic anterior, les molècules de tRNA eren probablement aminoacilades per algun altre catalitzador, possiblement altres molècules de RNA.

Però l'evidència més convincent del paper ancestral de la catalisi per RNA ha vingut dels estudis estructurals sobre el ribosoma (Green i Noller, 1997; Noller, 1991a; Noller, 1991b). L'obtenció de l'estructura tridimensional del ribosoma és un dels objectius més ambiciosos de la biologia estructural. La complexitat del problema ve donada per la massa molecular del ribosoma ($\sim 3 \times 10^6$ daltons) i pel fet que és el resultat de la interacció de tres molècules de RNA i més de cinquanta proteïnes diferents. Els ribosomes es componen de dues subunitats, de diferent massa molecular. La subunitat gran del ribosoma bacterià (anomenada 50S, en relació amb el seu coeficient de sedimentació) conté dues molècules de RNA (23S i 5S) i trenta-una proteïnes diferents. Aquesta subunitat forma el centre catalític del ribosoma i interactua amb l'extrem aminoacilat del tRNA.

Diversos estudis estructurals han localit-

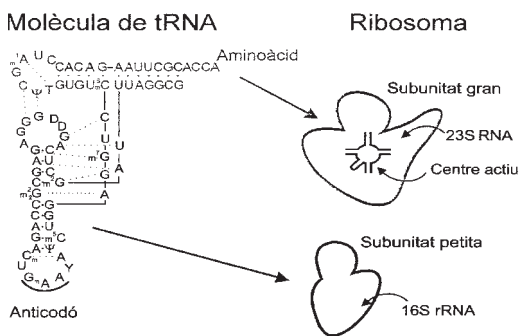


FIGURA 2. Esquema de les interaccions entre el ribosoma bacterià i la molècula de tRNA. La regió de l'RNA 23S que conté la zona activa del ribosoma està detallada dins del model de la subunitat gran. (Modificat a partir de Schimmel i Alexander, 1998).

zat el centre actiu del ribosoma bacterià en la proximitat de la molècula de RNA 23S, (una cadena de RNA de 12.900 bases; vegeu la figura 2). Experiments duts a terme el 1992 pel grup de Noller havien ja fet intuir que el centre catalític del ribosoma podia estar format únicament per les molècules de RNA de la subunitat gran (Noller *et al.*, 1992). Però la resposta definitiva a aquesta qüestió arribarà només amb la imminent obtenció de l'estructura tridimensional del ribosoma a escala atòmica (Moore, 1998). A partir d'aquesta fita la recerca sobre les bases estructurals de l'origen de la vida continuarà, tant esbrinant les bases de les transicions moleculars que varen donar lloc al nom de DNA, com provant de reproduir les reaccions centrals de metabolisme cel·lular amb hipotètics catalitzadors primitius. Potser no haurem d'esperar trenta anys abans de poder veure les primeres protocèl·lules creades *in vitro* exclusivament sobre la base de molècules de RNA.

CONCLUSIÓ

Els principis teòrics de la cristal·lografia i la RMN ja havien estat definits fa trenta anys. Però la seva aplicació a la biologia molecular i, sobretot, la disponibilitat dels seus resultats per a la majoria dels investigadors és un fet molt més recent. El creixement dels bancs de dades i la proliferació de sistemes per estudiar-los, han posat a les mans de tots els biòlegs molècules eines d'anàlisi estructural fins fa molt poc inexistents. La paraula clau és ara «educació». Les futures generacions d'investigadors han de ser formades en la utilització d'aquestes eines, i estimulades a considerar els problemes biològics com a xarxes de qüestions interdisciplinàries, amb bases moleculars que, gràcies als estudis estructurals, comencem a entendre amb detall.

AGRAÏMENTS

Gràcies a la Dra. Àngels Almenar, pels comentaris i discussions, i a la Dra. Rebecca Alexander, per permetre'm adaptar la figura dos.

BIBLIOGRAFIA

- BAJORATH, J.; R. STENKAMP; A. ARUFFO (1993). «Knowledge-based model building of proteins: concepts and examples». *Protein Sci.*, núm. 2, pàg. 1798-810.
- BALTIMORE, D. (1970). «RNA-dependent DNA polymerase in virions of RNA tumour viruses». *Nature*, núm. 226, pàg. 1209-11.
- BENSON, D. A.; M. S. BOGUSKI; D. J. LIPMAN; J. OSTELL; B. F. OUELLETTE (1998). «GenBank». *Nucleic Acids Res.*, núm. 26, pàg. 1-7.
- BOHR, N. (1933). «Light and Life». *Nature*, núm. 131, pàg. 421.
- BROWN, J. R.; W. F. DOOLITTLE (1995). «Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications». *Proc. Natl. Acad. Sci. USA*, núm. 92, pàg. 2441-5.
- CLORE, G. M.; A. M. GRONENBORN (1998). «Determining the structures of large proteins and protein complexes by NMR». *Trends Biotechnol.*, núm. 16, pàg. 22-34.
- CRICK, F. H. C. (1968). «The origin of the genetic code». *Journal of Molecular Biology*, núm. 38, pàg. 367-379.
- DELBRUCK, M. (1935). «Ueber die Natur der Genmutation und der Genstruktur». *Nachr. Akad. Wiss. Goettingen Math. Physik Kl.*, núm. 6, pàg. 223.
- DORSET, D. L. (1996). «Electron crystallography». *Acta Crystallogr. B.*, núm. 52, pàg. 753-69.
- ELLINGTON, A. D.; J. W. SZOSTAK (1990). «In vitro selection of RNA molecules that bind specific ligands». *Nature*, núm. 346, pàg. 818-22.
- FERREIRA, S. T.; T. COELHO-SAMPAIO (1996). «Intrinsic fluorescence as a probe of structure-function relationships in Ca(2+)-transport ATPases». *Bioscience Reports*, núm. 16, pàg. 87-106.
- FRANKLIN, R. E.; R. G. GOSLING (1953). «Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate». *Nature*, núm. 172, pàg. 156.
- GILBERT, W. (1986). «The RNA world». *Nature*, núm. 319, pàg. 618.
- GREEN, R.; H. F. NOLLER (1997). «Ribosomes and translation». *Annu. Rev. Biochem.*, núm. 66, pàg. 679-716.
- GUERRIER-TAKADA, C.; K. GARDINER; T. MARSH; N. PACE; S. ALTMAN (1983). «The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme». *Cell*, núm. 35, pàg. 849-57.
- HANSEN, J. C.; J. LEBOWITZ; B. DEMELER (1994). «Analytical ultracentrifugation of complex macromolecular systems». *Biochemistry*, núm. 33, pàg. 13155-63.
- HENDRICKSON, W. A.; J. L. SMITH; S. SHERIFF (1985). «Direct phase determination based on anomalous scattering». *Methods Enzymol.*, núm. 115, pàg. 41-55.
- HOOG, S. S.; B. ZHAO; E. WINBORNE; S. FISHER; D. W. GREEN; R. L. DESJARLAI; K. A. NEWLANDER; J. F. CALLAHAN; M. L. MOORE; W. F. HUFFMAN; [et al.] (1995). «A check on rational drug design: crystal structure of a complex of human immunodeficiency virus type 1 protease with a novel gamma-turn mimetic inhibitor». *J. Med. Chem.*, núm. 38, pàg. 3246-52.
- HUNTER, W. N. (1995). «Rational drug design: a multidisciplinary approach». *Mol. Med. Today*, núm. 1, pàg. 31, 34.
- JOYCE, G. F.; L. E. ORGEL (1986). «Non-enzymic template-directed synthesis on RNA random copolymers. Poly(C, G) templates». *J. Mol. Biol.*, núm. 188, pàg. 433-41.
- KAINOSHO, M. (1997). «Isotope labelling of macromolecules for structural determinations». *Nat. Struct. Biol.*, núm. 4, pàg. 858-61.
- KELLY, S. M.; N. C. PRICE (1997). «The application of circular dichroism to studies of protein folding and unfolding». *Biochim. Biophys. Acta*, núm. 1338, pàg. 161-85.
- KENDREW, J. C. (1950). «The crystal structure of horse met-myoglobin. I. General features: the arrangement of the polypeptide chains». *Proceedings of the Royal Society of London*, núm. A201, pàg. 62-89.
- KENDREW, J. C.; G. BODO; H. M. DINTZIS; R. G. PARRISH; H. W. WYCKOFF; D. C. PHILLIPS (1958). «A three dimensional model of the myoglobin molecule obtained by x-ray analysis». *Nature*, núm. 181, pàg. 662-666.
- KIM, S. H.; G. J. QUIGLEY; F. L. SUDDATH; A. MCPHERSON; D. SNEDEN; J. J. KIM; J. WEINZIERL; A. RICH (1973). «Three-dimensional structure of yeast phenylalanine transfer RNA: folding of the polynucleotide chain». *Science*, núm. 179, pàg. 285-8.
- KRUGER, K.; P. J. GRABOWSKI; A. J. ZAUG; J. SANDS; D. E. GOTTSCHLING; T. R. CECH (1982). «Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena». *Cell*, núm. 31, pàg. 147-57.
- LEWIN, B. (1970). «Second golden age of molecular biology». *Nature*, núm. 227, pàg. 1009-1013.
- LURIA, S. E.; M. DELBRÜCK (1943). «Mutations from virus sensitivity to virus resistance». *Genetics*, núm. 28, pàg. 491-511.
- MARGALEF, R.; R. PARÉS; A. PREVOSTI; J. VINAS; J. BAGUNYA; I. ARAGÓ (1971) MARGALEF, R. [ed.] «Cap on va la biologia moderna». *Col·loquis*. Barcelona: Societat Catalana de Biologia, Vol. 6.
- MCPHERSON, A., JR. (1976). «The growth and preliminary investigation of protein and nucleic acid crys-

- tals for X-ray diffraction analysis». *Methods. Biochem. Anal.*, núm. 23, pàg. 249-345.
- MEYER, E.F. (1997). «The first years of the Protein Data Bank». *Protein Science*, núm. 6, pàg. 1591-7.
- MOFFAT, K.; Z. REN (1997). «Synchrotron radiation applications to macromolecular crystallography». *Curr. Opin. Struct. Biol.*, núm. 7, pàg. 689-96.
- MOORE, P. B. (1998) «The three-dimensional structure of the ribosome and its components» *Annu. Rev. Biophys. Biomol. Struct.*, núm. 27, pàg. 35-58.
- NAGASE, H.; K. KAWAI; J. HAYAKAWA; H. WAKITA; A. MIZUSUNA; H. MATSUURA; C. TAJIMA; Y. TAKEZAWA; T. ENDOH (1998). «Rational drug design and synthesis of a highly selective nonpeptide delta-opioid agonist, (4aS*, 12aR*) -4a -(3-hydroxyphenyl) -2 -methyl- 1,2,3,4,4a,5,12,12a -octahydro-pyrido [3,4-b]acridine (TAN-67) [In Process Citation]». *Chem. Pharm. Bull. (Tokyo)*, núm. 46, pàg. 1695-702.
- NOLLER, H. F. (1991a). «Ribosomal RNA and translation». *Annu. Rev. Biochem.*, núm. 60, pàg. 191-227.
- NOLLER, H. F. (1991b). «Ribosomes. Drugs and the RNA world». *Nature*, núm. 353, pàg. 302-3.
- NOLLER, H. F.; V. HOFFARTH; L. ZIMNIAK. (1992). «Unusual resistance of peptidyl transferase to protein extraction procedures». *Science*, núm. 256, pàg. 1416-9.
- ORGEL, L. E. (1968). «Evolution of the genetic apparatus». *Journal of Molecular Biology*, núm. 38, pàg. 381-393.
- PABO, C. O.; R. T. SAUER (1984). «Protein-DNA recognition». *Annu. Rev. Biochem.*, núm. 53, pàg. 293-321.
- PABO, C. O.; R. T. SAUER (1992). «Transcription factors: structural families and principles of DNA recognition». *Annu. Rev. Biochem.*, núm. 61, pàg. 1053-95.
- PARDEE, A. B.; F. JACOB; J. MONOD (1959). «The genetic control and cytoplasmic expression of 'inducibility' in the synthesis of B-galactosidase by E. coli.». *Journal of molecular biology*, núm. 1, pàg. 165-178.
- RIBAS DE POUPLANA, L.; R. J. TURNER; B. A. STEER; P. SCHIMMEL (1998). «Genetic code origins: tRNAs older than their synthetases?». *Proc. Natl. Acad. Sci. USA*, núm. 95, pàg. 11295-300.
- ROBERTS, G. C. K. (1993). «NMR of macromolecules. A practical approach.». A: RICKWOOD, D., B. D. HAMES [eds.], *The practical approach series*. Oxford: IRL press.
- ROST, B., C. SANDER (1994). «Structure prediction of proteins – where are we now?». *Curr. Opin. Biotechnol.*, núm. 5, pàg. 372-80.
- ROST, B.; R. SCHNEIDER; C. SANDER (1993). «Progress in protein structure prediction?». *Trends Biochem. Sci.*, núm. 18, pàg. 120-3.
- ROULD, M. A.; J. J. PERONA; D. SÖLL; T. A. STEITZ (1989). «Structure of E. coli Glutaminyl-tRNA synthetase complexed with tRNA_Gln and ATP at 2.8 Å resolution». *Science*, núm 246, pàg. 1135-1142.
- SAGDOPAL, A. (1968). «The genetic code after the excitement». *Advances in Genetics*, núm. 14, pàg. 325-404.
- SCHIMMEL, P.; R. ALEXANDER (1998). «Perspectives: protein synthesis. All you need is RNA». *Science*, núm. 281, pàg. 658-9.
- SCHIMMEL, P.; L. RIBAS DE POUPLANA (1995). «Transfer RNA: from minihelix to genetic code». *Cell*, núm. 81, pàg. 983-6.
- SCHIMMEL, P. R.; D. SOLL (1979). «Aminoacyl-tRNA synthetases: general features and recognition of transfer RNAs». *Annu. Rev. Biochem.*, núm. 48, pàg. 601-48.
- SCHRÖDINGER, E. (1945). *What is life?* Nova York: Cambridge University press.
- STENT, G. S. (1968). «That was the molecular biology that was». *Science*, núm. 160, pàg. 390-395.
- TIVOL, W. F. (1995). «Solution of the phase problem in crystallography and application to dynamical electron diffraction». *Acta Crystallogr. A.*, núm. 51, pàg. 708-16.
- WAGNER, G. (1997). «An account of NMR in structural biology». *Nat. Struct. Biol.*, núm. 4 Suppl, pàg. 841-4.
- WATSON, J. D.; F. H. C. CRICK (1953). «A structure for deoxyribose nucleic acid». *Nature*, núm. 171, pàg. 737-738.
- WILLIAMSON, M. P.; T. F. HAVEL; K. WUTHRICH (1985). «Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H nuclear magnetic resonance and distance geometry». *J. Mol. Biol.*, núm. 182, pàg. 295-315.
- WOESE, C. (1967) «The evolution of the genetic code». *The genetic code*. Nova York: Harper & Row, pàg. 179-195.
- ZHANG, B.; T. R. CECH (1997). «Peptide bond formation by in vitro selected ribozymes». *Nature*, núm. 390, pàg. 96-100.