

ESTRUCTURA I DINÀMICA EN PÈPTIDS I PROTEÏNES: LA VISIÓ DE LA RESSONÀNCIA MAGNÈTICA NUCLEAR

M. A. CONTRERAS[†], M. GAIRI[§], O. MILLET[†], M. PONS[†]

*Departament de Química Orgànica[†] i Unitat de RMN dels Serveis Científicotècnics[§]
Universitat de Barcelona.*

Martí i Franquès, 1. 08028 Barcelona.

SUMMARY

Nuclear Magnetic Resonance (NMR) spectroscopy is today a well established technique for the determination of three-dimensional structures of macromolecules. With the introduction of labeling techniques, proteins with molecular weights well in excess of 25 kD can be studied. Structural information obtained from NMR is complementary to that obtained from X-ray not only because NMR structures are determined in solution but also because the main source of structural information by NMR, the Nuclear Overhauser Effect (NOE), originates from dynamic phenomena. In this chapter the main steps towards the characterization of structure and dynamics of proteins by NMR are reviewed. These include: i) spectral assignment using multidimensional techniques; ii) extraction of structural constraints from NOEs, coupling constants and chemical shifts; iii) determination of sets of 3D structures fulfilling the experimental constraints; iv) analysis and validation of the structures; and v) characterization of the dynamics of the protein.

INTRODUCCIÓ

L'espectroscòpia de ressonància magnètica nuclear (RMN) es pot considerar actualment una metodologia ben establerta per a la determinació estructural de macromolè-

cules biològiques, com petites proteïnes i fragments de DNA.

Fins ara fa uns anys, l'única tècnica estructural disponible per a l'obtenció d'es-

estructures tridimensionals era la difracció de raigs X de monocristalls. La major part del nostre coneixement estructural sobre les proteïnes està basat en estructures de cristalls. Ara bé, l'obtenció de cristalls de qualitat suficient per a l'estudi per raigs X no és una tasca fàcil i pot resultar inassolible per a algunes proteïnes.

L'espectroscòpia de RMN ha permès l'obtenció d'estructures tridimensionals d'alta resolució en solució, en condicions molt similars a les fisiològiques. Per tant, un primer avantatge d'aquesta metodologia respecte a la difracció de raigs X és que la mesura no està subjecta a la cristallització de la proteïna ni a la possible distorsió produïda per les forces d'empaquetament cristal·lí. D'altra banda, el treball en solució obre les portes a una major comprensió de les regles que governen l'estructura, la funció i la dinàmica de les proteïnes, com també al problema del seu plegament. L'origen de la informació estructural en les dues tècniques, però, és fonamentalment diferent i, de fet, cal veure-les més com a tècniques complementàries que com a competidores per a l'estudi de biomolècules.

Amb la metodologia actual es pot determinar l'estructura tridimensional de proteïnes amb un pes molecular de fins a 30-35 kDa, amb una precisió comparable a una estructura determinada per raigs X amb una resolució de 2-2,5 Å.

La RMN es basa en l'existència d'un nombre discret de nivells d'energia associats a les orientacions possibles del moment magnètic de certs nuclis situats en un camp magnètic intens. La diferència d'energia entre aquests nivells, molt petita, pot mesurar-se per l'absorció o emissió de radiació de la freqüència apropiada (en la zona de les ones de ràdio). Entre els nuclis que es troben habitualment en biomolècules, l'hidrogen és el més fàcilment observable per raons de sensibilitat, seguit del fòsfor. Els isòtops del car-

boni i el nitrogen que poden observar-se per RMN presenten abundàncies naturals molt baixes. La introducció de mètodes de clonatge i expressió que permeten l'enriquiment isotòpic de les mostres en aquests dos nuclis ha obert la possibilitat de la seva utilització de forma generalitzada, especialment per ajudar a l'assignació, per reduir els problemes d'encobriment i per tal de facilitar estudis dinàmics. La informació estructural, però, s'obté majoritàriament a través dels protons (nuclis d'hidrogen).

Les freqüències a les quals absorbeixen els diferents protons (o carbonis, nitrògens, etc.) d'una proteïna són lleugerament diferents segons el seu entorn químic i, per tant, en principi, és possible la seva observació separada. La probabilitat de coincidència accidental entre les freqüències de dos nuclis augmenta amb el nombre total de protons i, per tant, amb la mida de la proteïna. D'altra banda, per poder interpretar les dades de RMN serà necessari dur a terme l'**assignació de l'espectre**, és a dir, identificar dins l'estructura de la proteïna el nucli responsable de cada un dels senyals observats a freqüències diferents.

La informació estructural s'obté principalment a partir de l'anomenat **efecte nuclear Overhauser (NOE)**. Aquest efecte s'origina en les interaccions a través de l'espai entre diferents nuclis, habitualment protons. Les fluctuacions d'aquesta interacció internuclear a certes freqüències determinen la velocitat i el camí de retorn a l'equilibri d'un conjunt de nuclis quan algun d'ells és pertorbat. Aquesta pertorbació consisteix en l'establiment d'un grau d'ocupació dels diferents estats que no correspongui al que marca la seva diferència d'energia i la temperatura de la mostra.

És essencial, per comprendre la naturalesa de la informació estructural proporcionada per la RMN, d'adonar-se que la mesura experimental sobre la qual es fonamenta

és de caràcter cinètic i que depèn tant de l'existència d'interaccions internuclears com del fet que aquestes interaccions fluctuïn amb el temps. L'origen d'aquestes fluctuacions és el moviment de rotació global i la mobilitat interna de certes regions de la molècula.

Hi ha altres paràmetres de RMN mesurables experimentalment que donen informació estructural i que no tenen caràcter cinètic. Es tracta dels **desplaçaments químics** (el valor de la freqüència, que és específic per a un determinat nucli en un entorn particular) i l'**acoblament escalar** (que causa el desdoblament dels senyals per causa de la interacció entre nuclis veïns transmesa a través dels enllaços). Fins i tot en aquests casos, però, el valor tan petit de les diferències d'energia implicades fa que només puguin mesurar-se separadament si l'estructura es manté estable durant un temps prou llarg. En el cas d'equilibris ràpids entre diferents estructures, les dades experimentals reflectiran només una mitjana entre les estructures que caldrà interpretar amb cautela.

EL PROBLEMA DE L'ENCOBRIMENT ACCIDENTAL DE SENYALS. ESPECTROSCÒPIA MULTIDIMENSIONAL

La possibilitat d'encobriment accidental de senyals augmenta amb el nombre de protons presents. El rang de freqüències on es troben tots els protons és aproximadament de 5.000 Hz en un aparell basat en un imant d'11,74 Tesla, però de 8.000 Hz en un imant de 18,79 Tesla. Si suposéssim una resolució d'1 Hz, aquests serien els nombres màxims de freqüències diferents que es poden mesurar. Malauradament la distribució no és uniforme i entorns semblants donen freqüències semblants. En conseqüència, fins i tot per a pèptids petits, la possibilitat d'encobriment accidental d'alguns dels protons és molt gran (figura 1). L'augment de resolució (i de sensibilitat) que proporciona la utilització de camps magnètics més intensos ha marcat l'evolució de la instrumentació. Tot i això, el guany en resolució és només lineal amb el camp i, per si sol, no hauria permès l'estudi de proteïnes.

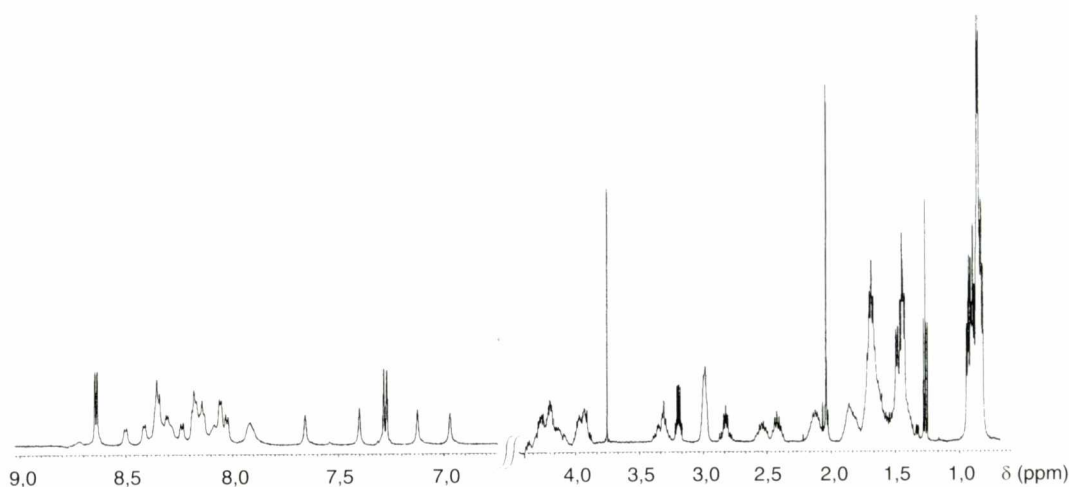


FIGURA 1. Espectre de RMN de ^1H del pèptid MNKE (vint residus) a 20,7 °C en $\text{H}_2\text{O}:\text{D}_2\text{O}$ 9:1 a 500 MHz

La possibilitat d'establir correlacions entre diferents nuclis ha obert la porta a la utilització de tècniques multidimensionals. En els experiments monodimensionals s'analiza cada nucli/freqüència separatament.

En un experiment bidimensional (2D), s'identifiquen parelles de freqüències que corresponen a parells de nuclis relacionats entre si d'alguna forma, que dependrà del tipus d'experiment concret utilitzat (Kessler *et al.*, 1988).

Els experiments homonuclears (que correlacionen nuclis del mateix tipus) 2D més utilitzats són el NOESY (Nuclear Overhauser Effect Spectroscopy) (Jeener *et al.*, 1979), el ROESY (Rotating frame Overhauser Effect Spectroscopy) (Bothner-By *et al.*, 1984; Bax *et al.*, 1985), el COSY (Correlated Spectroscopy) (Aue *et al.*, 1976) i el TOCSY (Total Correlation Spectroscopy) (Braunschweiler *et al.*, 1983). Els dos primers estableixen la

correlació entre protons que es troben a prop a l'espai i es diferencien en la manera en què són afectats per fluctuacions de freqüències diferents. Els dos darrers correlacionen parells de protons que presenten acoblament escalar, ja sigui directament entre ells dos (COSY) o per mitjà d'una cadena d'acoblements que pot implicar altres hidrògens (TOCSY). A la figura 2 es mostra l'espectre TOCSY d'un pèptid de vint residus representat com un mapa de contorn, en el qual cada senyal està caracteritzat per dues freqüències, i la intensitat del pic es representa pel nombre de corbes de nivell.

Si recordem que la informació estructural apareix principalment com a resultat de la interacció entre diferents nuclis (efecte NOE i acoblaments escalars) resulta evident que les tècniques multidimensionals són la forma natural d'obtenir les dades necessàries.

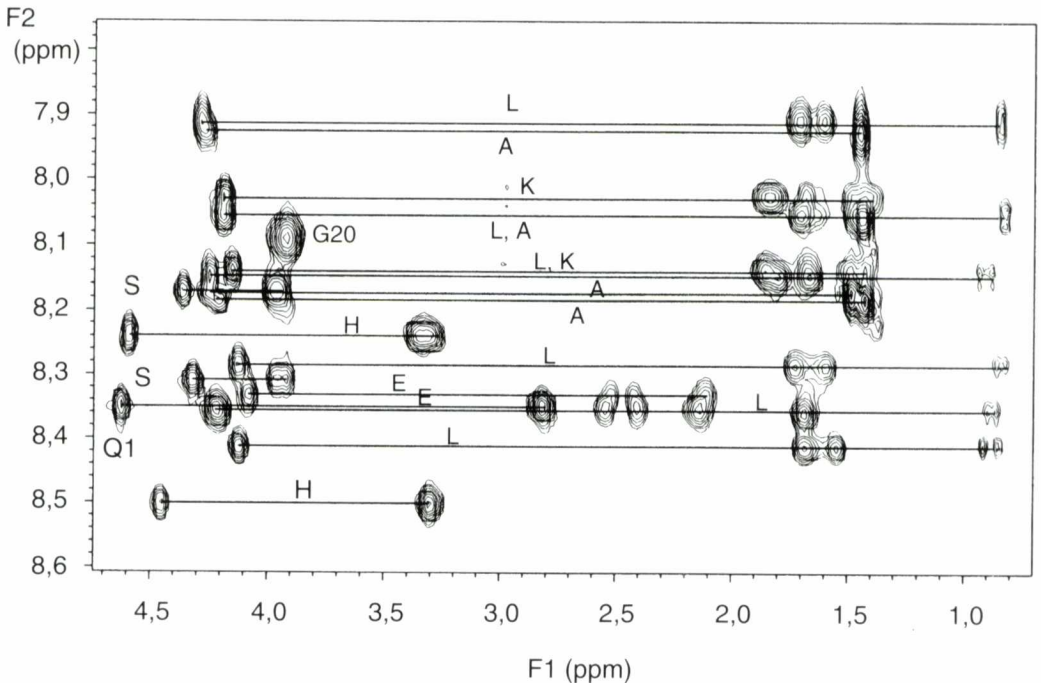
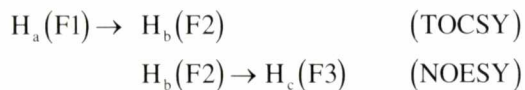


FIGURA 2. Expansió d'un espectre TOCSY del pèptid MNKE a 20,7 °C en H₂O:D₂O 9:1 a 500 MHz

En un experiment tridimensional o tetradimensional (3D, 4D) es consideren grups de tres o quatre freqüències. La possibilitat que es produeixi una coincidència accidental de dos d'aquests grups de n nuclis decreix notablement. Formalment els experiments 3D es poden entendre com la concatenació de dos experiments 2D. Per exemple, a l'experiment 3D TOCSY-NOESY (Oschkinat *et al.*, 1989) el triplet de freqüències (F1, F2, F3) s'originaria de l'establiment de dues correlacions diferents:



De fet, aquesta és la forma com, a la pràctica, s'analitzen aquests tipus d'espectres: com un conjunt de plans (per exemple F2-F3, que correspondria a un experiment NOESY), els quals difereixen pel valor de la freqüència restant (en aquest cas F1). A cada

pla només apareixen senyals d'aquells parells de protons (H_b, H_c), en els quals H_b presenta una correlació TOCSY amb el protó H_a , la freqüència del qual defineix el pla. La figura 3 mostra un dels plans F2-F3 d'un espectre 3D TOCSY-NOESY del depsipèptid deshidrodidemnina a 500 MHz.

Finalment, si alguns dels nuclis d'aquests grups són carboni o nitrogen, es pot aprofitar la major dispersió de freqüències que presenten habitualment aquests nuclis per poder distingir protons encoberts. Els experiments bidimensionals heteronuclears (entre nuclis de diferents menes) més utilitzats són l'HMQC (Heteronuclear Multiple Quantum Coherence) (Müller, 1979; Bax *et al.*, 1983) i l'HSQC (Heteronuclear Single-Quantum Coherence) (Bodenhausen *et al.*, 1980) que correlacionen nitrògens o carbonis amb els protons que tenen units directament. De forma anàloga a com s'ha comentat per als experiments homonuclears, l'extensió a 3D és immediata: l'experiment

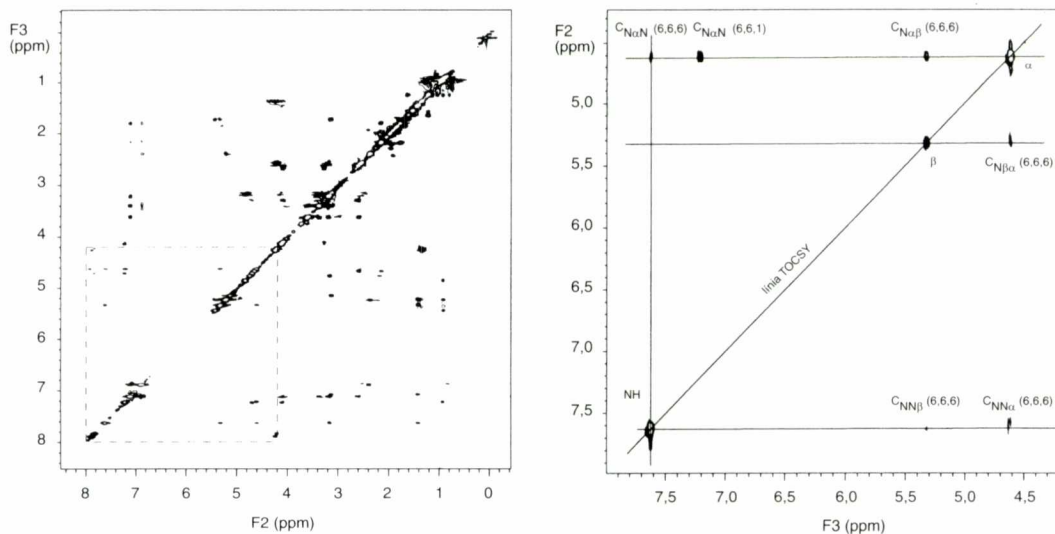


FIGURA 3. Espectre 3D TOCSY-NOESY del depsipèptid deshidrodidemnina en CHCl_3 (15 mM) a 27 °C. L'espectre de l'esquerra s'ha obtingut sumant tots els plans F2-F3 i equival a un experiment NOESY convencional. A la dreta es mostra una expansió d'un pla individual F2-F3 a $F1=7,62$ ppm. La zona no expandida del pla no conté senyals addicionals.

3D NOESY-HMQC (Kay *et al.*, 1989) permet separar els pics NOESY que impliquen un NH caracteritzat per una freqüència de ^{15}N determinada i evitar així les ambigüitats que podrien derivar-se de la coincidència accidental dels desplaçaments químics de dos protons d'amida.

La utilització de mostres doblement marcades amb ^{15}N i ^{13}C permet ampliar el nombre de correlacions possibles a aquelles que impliquen enllaços C-N i C-C (van de Ven, 1995; Clore *et al.*, 1994). Els experiments 3D que en resulten faciliten l'assignació inequívoca d'espectres de gran complexitat. S'han descrit un gran nombre d'aquests experiments per a l'estudi de proteïnes i s'ha establert una manera estàndard d'anomenar-los, que consisteix en la concatenació del símbol que representa els nuclis que es correlacionen i entre parèntesis s'hi afegeixen aquells altres que participen en la propagació de la correlació, però que no són identificats de forma explícita. Per exemple, l'experiment 3D HNCO (Kay *et al.*, 1990) correlaciona HN(i), N(i) i CO (i-1) mentre que l'experiment 3D HN(CO)CA (Bax *et al.*, 1991) correlaciona HN(i), N(i), Ca (i-1) però la informació es transmet per mitjà de CO(i-1), el qual no és detectat de forma directa.

ESTRATÈGIES D'ASSIGNACIÓ

L'estratègia que cal seguir per a l'assignació de proteïnes sense enriquiment amb ^{15}N i ^{13}C va ser desenvolupada pel grup de K. Wüthrich (Wüthrich, 1986) i comprèn dues etapes:

1. Identificació dels protons que formen part del mateix sistema d'espín i que coincideixen amb els d'un mateix residu, ja que no hi ha acoblament apreciable a través de l'enllaç peptídic. El rang de desplaçaments del conjunt de protons de cada sistema

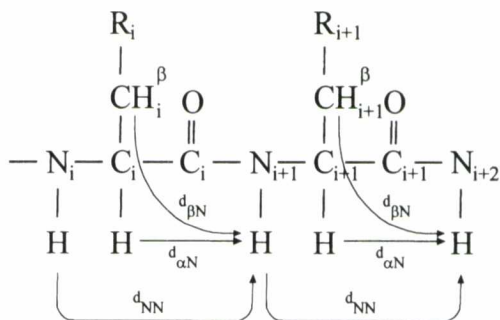


FIGURA 4. Fragment de cadena polipeptídica amb les connectivitats NOE útils per a l'assignació seqüencial

d'espín permet identificar la naturalesa del residu.

2. Identificació dels sistemes d'espín veïns a partir de l'observació de connectivitats NOE entre protons dels dos sistemes (habitualment $d_{\alpha N}$, d_{NN} i $d_{\beta N}$, figura 4). La comparació de la seqüència de residus determinats d'aquesta manera amb la seqüència completa de la proteïna permet identificar l'ordre seqüencial dels sistemes d'espín.

Aquesta metodologia, però, està limitada a proteïnes de menys d'uns 10 kDa. Els principals problemes deriven, d'una banda, de l'encobriment de senyals i, de l'altra, de la utilització d'experiments tipus NOESY per establir la connectivitat entre residus adjacents la qual cosa ja suposa, no sempre correctament, que les distàncies més curtes dins una proteïna impliquen protons de residus adjacents. De fet, les distàncies entre protons adjacents depenen de la conformació local (estructura secundària) i, en una proteïna plegada, residus molt separats a la seqüència poden trobar-se molt propers a l'espai. A la pràctica, els processos d'assignació i determinació de l'estructura es duen a terme de forma iterativa, utilitzant la informació estructural ja obtinguda a partir de dades inequívokes per resoldre ambigüitats en l'assignació.

La disponibilitat de mostres enriquides

TAULA 1. Connectivitats de grups de tres nuclis (pics en experiments 3D) que permeten l'assignació seqüencial de proteïnes marcades amb ^{15}N i ^{13}C

Experiment	Residu i -1					Residu i					Referència
	N	HN	Ca	Ha	CO	N	HN	Ca	Ha	CO	
HCACO								X	X	X	Kay 1990
HN(CA)CO						X	X			X	Clubb 1992
(feble)					X	X	X				
HNCA						X	X	X			Kay 1990
(feble)			X			X	X				
HN(CO)CA			X			X	X				Bax 1991
HNCO					X	X	X				Kay 1990
CBCA(CO)NH			X			X	X				Grzesiek 1992a

amb ^{15}N o ^{13}C permet la utilització d'estratègies inequívokes d'assignació que empren només informació que es transmet per mitjà d'enllaços i que, per tant, no són afectades per la presència de distàncies curtes entre residus no adjacents.

En aquest tipus d'estratègia s'inverteix l'ordre de les dues etapes:

1. Inicialment s'estableixen les connectivitats entre els àtoms que formen la cadena principal.

2. A continuació s'estableixen les connectivitats entre aquests àtoms i les cadenes laterals que permeten identificar el tipus de residu i completar l'assignació.

La taula 1 mostra les connectivitats entre àtoms de la cadena principal proporcionades per diferents experiments 3D. La probabilitat de cometre errors és molt menor si es comparen posicions de pics que tinguin dues de les freqüències en comú.

Les correlacions amb les cadenes laterals poden obtenir-se a partir dels experiments 3D CBCA(CO)NH (Grzesiek *et al.*, 1992a), 3D CBCANH (Grzesiek *et al.*, 1992b), 3D

TOCSY-HMQC (Wijmenga *et al.*, 1989), 3D HCCH-TOCSY (Bax *et al.*, 1990) i altres.

És comunament acceptat que el procés d'assignació d'espectres de RMN és un treball tediós i llarg. Per aquesta raó, molts grups de recerca han investigat la possibilitat d'automatitzar-lo (vegeu, per exemple, Oschkinat *et al.*, 1994). En última instància hom pot esperar una integració completa de l'assignació i el càlcul d'estructures.

Un aspecte particular de l'assignació que té gran importància per a l'obtenció d'estructures correctes és l'anomenada **assignació estereoespecífica** d'alguns protons, com podrien ser els β -metilènics, els metils de Val o els protons α de Gly.

Aquests protons donen sovint senyals diferents i la seva assignació requereix conèixer la conformació local. L'assignació estereoespecífica es duu a terme sovint de forma iterativa durant el procés d'obtenció de l'estructura de la proteïna o bé buscant la compatibilitat entre diferents experiments, tots ells sensibles a aquesta conformació.

INFORMACIÓ ESTRUCTURAL DERIVADA DELS EXPERIMENTS DE RMN

A. Distàncies

La informació estructural extreta dels experiments de RMN consisteix en un gran nombre de restriccions locals que haurà de complir l'estructura proposada per a la molècula en estudi.

Les restriccions de distància entre protons es deriven sobretot a partir de les intensitats dels pics de correlació dels espectres NOESY. La velocitat inicial de creixement dels pics en aquests espectres, mesurada a temps de barreja (temps de bescanvi de magnetització) curts, és proporcional a la velocitat de relaxació creuada σ_{ij} entre els protons i i j . En general, les velocitats de relaxació creuada es poden mesurar a partir d'una sèrie d'experiments NOESY amb diferents temps de barreja.

Per a una molècula rígida que es mou isotròpicament, la velocitat de relaxació creuada entre dos espins i i j està relacionada amb la distància d_{ij} i amb el temps de correlació rotacional τ_c :

$$\sigma_{ij} \propto \tau_c d_{ij}^{-6} \quad [1]$$

Per tant, a partir d'una distància de calibració fixa i coneguda d_{cal} entre dos protons de la molècula (per exemple, distància entre protons geminals o aromàtics), es poden calcular les distàncies d_{ij} :

$$d_{ij} = d_{cal} (\sigma_{cal} / \sigma_{ij})^{1/6} \quad [2]$$

Aquestes equacions suposen que cada parell de nuclis està aïllat de la resta. De fet hi ha camins competitiu per a la relaxació que poden donar lloc a l'aparició d'intensitats incorrectes i, fins i tot, l'observació de pics de correlació que no corresponen a pa-

rells de protons propers a l'espai. Aquests efectes, anomenats globalment **difusió d'espín**, solen ser poc importants quan s'utilitzen temps de barreja curts. Per a temps de barreja més llargs cal incorporar de forma explícita la possible existència de camins alternatius per mitjà de l'anomenada matriu completa de relaxació.

El segon inconvenient associat a la determinació acurada de distàncies és que les proteïnes no són cossos rígids. La mobilitat intramolecular provoca una mitjana no lineal de distàncies i fa que hi hagi diferents temps de correlació efectius per diferents vectors interprotònics en la molècula. Els moviments interns tenen lloc en un rang molt ampli d'escalas de temps. Els moviments més lents són difícils de tractar i constitueixen la font més important d'error en la determinació de distàncies a partir dels NOE. Per tot això, no és usual convertir les intensitats dels pics de NOE directament en distàncies. L'aproximació més estesa és la de convertir les dades de NOE en restriccions de distàncies bastant laxes. Així, és força comú classificar els NOE en forts i febles, o bé en forts, febles i mitjans, a partir d'una anàlisi qualitativa de les seves intensitats. Per exemple, per als NOE forts s'imposa una restricció de distància entre 1,8 Å i 3 Å, per als mitjans entre 1,8 Å i 4 Å i per als febles entre 1,8 Å i 5 Å. El límit inferior de la restricció correspon senzillament a la suma dels radis de van der Waals, és a dir, a la distància mínima a què es poden trobar dos protons sense encobrir-se. En els casos en què no és possible dur a terme l'assignació estereoespecífica, cal recórrer al concepte de pseudoàtom. Es tracta d'un punt de referència de les restriccions i està situat en el centre geomètric de les posicions dels protons considerats (dels metilens, metils, α -H de Gly o aromàtics). La incorporació dels pseudoàtoms implica una correcció en la restricció de les distàncies, donat que augmenta la im-

precisió. Per compensar l'error que es comet en no considerar la posició correcta del protó, se sol augmentar el límit superior de la restricció.

S'ha observat que la precisió i l'exactitud de les estructures resultants depèn més del nombre total de restriccions que s'incorporen als càlculs que de la seva precisió individual. En particular, els NOE a llarga distància, observats entre protons que es troben molt lluny en la seqüència de la proteïna, introdueixen restriccions molt grans a les possibles conformacions de la molècula de forma bastant independent dels límits exactes utilitzats per a la restricció de distància. Òbviament, la utilització de distàncies més precises millorarà la definició, especialment d'aquelles regions que només mostren restriccions locals.

Una altra alternativa a la derivació de distàncies interprotòniques a partir de l'efecte NOE és la utilització d'intensitats dels pics de correlació, directament. L'aproximació de la matriu de relaxació permet aquest tractament, que es coneix com refinament directe del NOE. Molt recentment s'estan desenvolupant diverses estratègies basades en aquesta metodologia. El principal inconvenient fins ara és l'elevat cost computacional, donat que els temps de càlcul són molt llargs.

B. Restriccions d'angles de torsió a partir de constants d'acoblament

Les constants d'acoblament veïnals entre espins poden proporcionar una informació molt útil, que complementa la derivada de les dades de NOE. En particular, els límits dels angles de torsió es poden estimar a partir de la grandària de les constants d'acoblament.

Les constants més fàcils de determinar en proteïnes solen ser les ${}^3J_{\text{HN-}\alpha\text{H}}$. El valor de la constant està relacionat amb l'angle de

torsió Θ definit pels tres enllaços que separen els protons, per mitjà de la relació de Karplus, que per cadenes polipeptídiques adopta l'expressió següent (Karplus *et al.*, 1959):

$${}^3J(\Phi) = A \cos^2\Theta + B \cos \Theta + C,$$

on els paràmetres A, B i C han estat calculats per a polipèptids. L'angle Θ està directament relacionat amb l'angle Φ de la cadena peptídica.

L'equació de Karplus ens permet deduir intervals de valors dels angles de torsió consistents en el valor experimental de les constants d'acoblament. Aquests intervals es poden incorporar en forma de restriccions experimentals en els càlculs de l'estructura, que es descriuran més endavant. Així, valors de ${}^3J_{\text{HN-}\alpha\text{H}} < 6\text{Hz}$ corresponen a intervals de -35° a -85° per als angles Φ de l'esquelet, mentre que per a constants ${}^3J_{\text{HN-}\alpha\text{H}}$ més grans de 8Hz, els valors de l'angle Φ estaran compresos entre -80° i -180° .

Les restriccions dels angles de torsió de les cadenes laterals χ_1 es poden obtenir a partir de les constants d'acoblament ${}^3J_{\alpha\text{H-}\beta\text{H}}$. L'anàlisi conjunta de les constants ${}^3J_{\alpha\text{H-}\beta\text{H}}$ i ${}^3J_{\alpha\text{H-}\beta\text{H}'}$ i les intensitats relatives dels NOE intraresiduals entre els protons NH-C β H, NH-C β H', C α H-C β H, C α H-C β H' permeten l'assignació estereoespecífica dels protons dels metilens.

Per mitjà de la relació de Karplus es poden derivar, anàlogament a l'angle Φ , els intervals de l'angle χ_1 consistents en els valors de les constants ${}^3J_{\alpha\text{H-}\beta\text{H}}$ i ${}^3J_{\alpha\text{H-}\beta\text{H}'}$.

La disponibilitat de mostres de macromolècules biològiques uniformement marcades amb ${}^{13}\text{C}$ i ${}^{15}\text{N}$ permet determinar constants d'acoblament heteronuclears a un, dos i tres enllaços. La constant ${}^3J_{\alpha\text{H-i-N}(i+1)}$ permet obtenir informació sobre l'angle Ψ de l'esquelet de les cadenes polipeptídiques. La relació de Karplus implica que, per a determi-

nats valors de la constant, no hi ha un angle de torsió únic i, per tant, no es pot derivar una única restricció angular. Per solucionar aquest problema s'utilitzen restriccions que penalitzen els angles que donarien constants d'acoblament diferents de l'experimental.

C. Informació que proporcionen els desplaçaments químics

Des de fa temps se sap que els desplaçaments químics de pèptids i proteïnes depenen de l'estructura d'una manera bastant regular. De fet, s'ha observat que els desplaçaments en proteïnes estructuralment homòlogues són homòlegs. Seria de gran interès poder explotar aquesta informació en estudis estructurals.

Fins ara, la utilització dels desplaçaments químics en la determinació estructural ha estat molt limitada, malgrat que en els darrers anys s'han concentrat esforços a trobar relacions empíriques entre l'estructura i els desplaçaments. En particular, s'ha posat una atenció considerable en la manera en què els desplaçaments dels protons α reflecteixen l'estructura secundària. S'ha observat que les regions amb estructura de full β presenten desplaçaments químics cap a camps baixos, mentre que les regions helicoïdals mostren desplaçaments cap a camps alts. També s'ha observat una forta correlació dels desplaçaments químics dels $C\alpha$ i dels C carbonílics amb l'estructura secundària, de manera que es mouen cap a camps baixos en les conformacions helicoïdals i cap a camps alts en les conformacions de cadena estesa o amb fulls β . La identificació de les regions d'estructura secundària a partir d'aquests desplaçaments presenta un grau de fiabilitat molt elevat.

Aquestes mateixes observacions es poden aplicar als pèptids lineals, en els quals la coexistència de diferents conformacions im-

pedeix el càlcul de l'estructura basat en el NOE quantitatiu. D'altra banda, la sensibilitat del desplaçament químic dels protons amida, i de la seva variació amb la temperatura, a diferents factors estructurals, permet caracteritzar petites distorsions locals en elements d'estructura secundària (Contreras *et al.*, 1996).

S'han proposat diferents equacions de caire empíric (Osapay i Case, 1991; Spera i Bax, 1991) i també basades en càlculs mecanicoquàntics (Le *et al.*, 1995) i s'han utilitzat per refinar estructures.

DETERMINACIÓ DE L'ESTRUCTURA TRIDIMENSIONAL D'UNA PROTEÏNA

La construcció d'un model molecular compatible amb la informació geomètrica derivada de les dades de RMN es pot comparar a la reconstrucció d'un trencaclosques. Es tracta d'identificar parells de peces que encaixin, fins a reconstruir tot el model. Ara bé, a diferència dels trencaclosques, les estructures de molècules en solució no són úniques, donat que les dades derivades de la RMN són insuficients per caracteritzar completament l'estructura i, a més a més, la major part de les molècules d'interès són flexibles en solució. Malauradament, les tècniques de recerca sistemàtica són lentes i consumeixen massa temps per ser utilitzades amb macromolècules. A més a més, amb molècules grans, el nombre de conformacions possibles és generalment tan gran que no és factible enumerar-les totes.

En aquestes condicions, el que se sol fer és calcular un conjunt d'estructures que sigui representatiu de totes les conformacions compatibles amb la informació experimental obtinguda de les dades de RMN. El que s'ha de verificar és que aquest conjunt sigui suficientment gran per tal que qualsevol de les propietats geomètriques presents en tots

els membres sigui conseqüència de la informació experimental introduïda en el càlcul. A part, un bon mètode de càlcul ha de ser capaç d'explorar de manera adequada l'espai conformacional.

Hi ha diversos mètodes de càlcul que permeten determinar l'estructura tridimensional d'una proteïna a partir de les dades de RMN. Malgrat que hi ha aproximacions molt simples, com la construcció de models o els mètodes que parteixen de bases de dades derivades de les estructures de raigs X, els principals mètodes computacionals es poden englobar en dos grans grups: els que treballen en l'espai de distàncies i els que treballen en l'espai real. Els primers són uns mètodes que utilitzen la matriu mètrica i que es coneixen amb el nom de geometria de distàncies (DG). Es tracta de projectar un conjunt de distàncies des d'un espai n dimensional a l'espai tridimensional de coordenades cartesianes. D'altra banda, els mètodes que operen a l'espai real treballen directament en l'espai de coordenades cartesianes i impliquen l'aplicació de la minimització o la dinàmica molecular restringida.

Tots aquests mètodes tenen en comú la recerca conformacional del mínim global d'una **funció objectiu** que conté una suma de termes estereoquímics i restriccions experimentals derivades de la RMN. La recerca del mínim global no és un procés simple, donat que la funció objectiu està caracteritzada per molts mínims locals falsos que cal evitar o superar.

A. Geometria de distàncies (DG)

La geometria de distàncies és una branca de les matemàtiques consistent en la construcció d'estructures a partir de distàncies internes. L'adaptació d'aquesta idea al seu ús en els problemes d'estructura química ha estès el nom de geometria de distàncies als

programes computacionals que converteixen restriccions geomètriques en coordenades moleculars.

Des dels inicis de la seva utilització en problemes moleculars, la DG es considerava només com una primera etapa del càlcul. Una segona etapa de dinàmica molecular era necessària per a l'obtenció d'estructures d'energia adequada. La idea era usar la rapidesa de l'algoritme de DG i la seva capacitat d'explorar l'espai conformacional per obtenir estructures de partida, que es podrien refinar amb mètodes de càlcul més sofisticats.

Com ja s'ha comentat, el mètode permet obtenir coordenades a partir d'una matriu de distàncies ideal, on tots els termes es coneixen exactament. Ara bé, la informació que obtenim a partir de les dades de RMN no permet obtenir totes les distàncies, sinó solament les distàncies entre protons separats com a màxim 5 Å. La matriu de distàncies es pot completar, almenys parcialment, amb la informació que prové de l'estructura covalent de la proteïna (longituds i angles d'enllaç), però mai no es poden obtenir totes les distàncies possibles. Per això, en lloc d'una, s'utilitzen dues matrius, una amb límits superiors de distància, U , i una altra amb límits inferiors, L .

L'algoritme de DG es basa en les etapes següents:

a) Construcció de les matrius amb cotes superiors U i inferiors L , per a totes les distàncies entre els àtoms de la molècula. Alguns dels elements u_{ij} i l_{ij} provenen de longituds o angles d'enllaç derivats de l'estructura covalent, i d'interval·ls de distància trobats experimentalment a partir dels NOE.

b) Suavització dels límits (*bound smoothing*) utilitzant la desigualtat triangular (donats tres punts A, B i C de l'espai, la distàn-

cia AC no pot ser més gran que la suma d'AB i BC) i altres relacions geomètriques similars.

c) Generació d'una matriu de distàncies, els termes d_{ij} de la qual són escollits a l'atzar entre els límits superior i inferior $l_{ij} \leq d_{ij} \leq u_{ij}$.

d) Obtenció de les coordenades atòmiques a partir de la matriu mètrica mitjançant un algoritme anomenat *embedding* (que vol dir incrustació) seguit d'un procés d'optimització que minimitza una funció d'error que detecta incompliments de les restriccions o errors en la quiralitat.

Repetint el càlcul de DG diverses vegades, és a dir, escollint diverses vegades a l'atzar la matriu de distàncies entre els límits superior i inferior es pot obtenir un conjunt d'estructures i , per tant, es pot saber com n'està, de ben determinada, l'estructura per a les restriccions imposades i fins a quin punt és «única».

Hi ha diversos programes que fan ús de la geometria de distàncies: EMBED (G. M. Crippen, I. D. Kuntz, T. F. Havel, Universitat de Califòrnia, San Francisco), DISGEO (Havel *et al.*, 1984), DGEOM (J. Blaney, Dupont), DSPACE (D. Hare, Infinity Systems)... i que han estat molt utilitzats per calcular estructures.

El principal inconvenient d'aquests mètodes és la dificultat d'introduir informació experimental derivada de les constants d'acoblament o dels desplaçaments químics. Malgrat que les estructures resultants dels càlculs de DG presenten sovint algunes violacions de les restriccions derivades de les mesures de RMN i també presenten energies elevades que generalment provenen de contactes no enllaçants desfavorables, l'experiència demostra que si hi ha un nombre suficient de restriccions correctes, les estructures obtingudes es poden refi-

nar per obtenir fàcilment mínims de baixa energia.

B. Minimització d'una funció objectiu variable

Aquest mètode sovint s'engloba dins del terme general de geometria de distàncies. Ara bé, ja que opera en l'espai dels angles de torsió, no és estrictament un procediment de DG.

La idea fonamental del mètode es basa a definir una funció positiva, amb valor zero quan es compleixen les restriccions de distància i amb un valor diferent de zero si no es verifiquen. Es tracta, en definitiva, de minimitzar una funció. La definició de la funció objectiu s'ha de fer de manera que es penalitzin tant les restriccions experimentals violades com els encobriments entre àtoms. En general, té la forma:

$$T = \sum_{ij} T_{ij}^{\text{distàncies}} + \sum_{ij} T_{ij}^{\text{encobriment}}$$

Aquesta funció depèn de les coordenades de tots els àtoms de la molècula (3N-6 dimensions). Donat que aquest nombre és molt gran, és convenient reduir el nombre de dimensions, per exemple, treballant en l'espai dels angles de torsió. Així, per a un polipèptid, cada aminoàcid es pot descriure amb les variables dels angles de l'esquelet i de la cadena lateral ($\Phi, \Psi, \chi_1, \chi_2 \dots$).

Per evitar que el procés de minimització s'aturi en mínims locals el que es fa és modificar el pes de diferents tipus de restriccions a la funció objectiu T en etapes progressives de refinament. Al principi només es fan intervenir les restriccions locals (intraresiduals i seqüencials), mentre que en etapes posteriors s'inclouen les restriccions a més llarga distància. Generalment, es comença amb diverses conformacions inicials obtingudes prenent valors a l'atzar dels angles diedres. Una implantació del mètode força

utilitzada la constitueixen els programes DISMAN i DIANA (Güntert *et al.*, 1991), aquest darrer molt flexible per definir estratègies de minimització.

C. Dinàmica molecular restringida

La dinàmica molecular restringida (DMR) és una alternativa simple i molt utilitzada als algorismes de geometria de distàncies. Es basa en una modificació dels programes de simulació de dinàmica molecular.

La dinàmica molecular resol les equacions de Newton del moviment, per a un sistema de partícules:

$$m_i (d^2 r_i / dt^2) = -\nabla V(r_1, r_2, \dots, r_N)$$

Les posicions inicials provenen d'una estructura de partida, que sol ser una estructura obtinguda per DG o un altre mètode més senzill, encara que, en principi, es pot partir de qualsevol estructura. Les velocitats inicials es prenen a partir de la distribució de Maxwell corresponent a la temperatura desitjada.

En la dinàmica molecular restringida (DMR) s'utilitza un potencial que inclou un terme experimental, a més a més del potencial teòric que engloba les contribucions de les longituds i angles d'enllaç, angles de torsió i interaccions no enllaçants. El nou terme de potencial experimental de la DMR no té sentit físic, però representa la informació experimental derivada de les dades de RMN. En general, aquest terme inclourà contribucions de cada tipus de restriccions:

$$V_{\text{experimental}} = V_{\text{NOE}} + V_j + V_\sigma$$

Per exemple, l'expressió del potencial V_{NOE} sol adoptar una forma del tipus:

$$V_{\text{NOE}} = k_{\text{dist}} (d_{ij} - u_{ij})^2 \text{ si } d_{ij} > u_{ij}$$

$$V_{\text{NOE}} = 0 \text{ si } l_{ij} < d_{ij} < u_{ij}$$

$$V_{\text{NOE}} = k_{\text{dist}} (d_{ij} - l_{ij})^2 \text{ si } d_{ij} < l_{ij}$$

on d_{ij} és la distància entre els àtoms i i j i u_{ij} i l_{ij} són els límits superior i inferior d'aquesta distància, respectivament. Per a altres restriccions experimentals, com les constants d'acoblament o els angles de torsió, les expressions dels potencials són anàlogues. El que es verifica en tots els casos és que quan es compleix la restricció experimental el potencial és zero i, a mesura que la restricció no es compleix i s'allunya dels límits imposats, el potencial augmenta.

El gran avantatge de la dinàmica molecular restringida respecte a altres mètodes computacionals es basa en la seva capacitat de superar barreres de potencial i, per tant, de superar mínims locals. Si la simulació es porta a terme a temperatures elevades, augmenta l'eficàcia dels salts de barreres de potencial; per això, moltes vegades s'usen protocols en els quals l'estructura s'escalfa a una temperatura molt elevada (ex.: 1.000 K) i després es refreda lentament. Aquestes estratègies s'anomenen d'*atemperat simulat* (*Simulated Annealing*, SA). En els programes de SA s'utilitza habitualment una funció d'energia potencial simplificada, en la qual les interaccions no enllaçants es representen per un únic terme de repulsió V_{repu}^* :

$$V_{\text{repu}} = 0 \quad \text{si } r \geq r_{\text{min}}$$

$$V_{\text{repu}} = k (s^2 r_{\text{min}}^{-2} - r^{-2})^2 \quad \text{si } r < r_{\text{min}}$$

on r_{min} és habitualment un 80% de la suma dels radis de Van der Waals.

Els programes més aplicats per a la dinàmica molecular restringida o el *Simulated Annealing* són CHARMM (Brooks *et al.*, 1983), DISCOVER (Biosym/Molecular Simulations Ltd.), AMBER (P. K. Weiner *et al.*, 1981; S. J. Weiner *et al.*, 1984), GROMOS (Groningen Molecular Simulation, W. F. van Gunsteren, H. J. C. Berend-

sen, BIOMOS) i X-PLOR (Brunger A. T.).

L'èxit d'un mètode de refinament estructural depèn en bona part del tipus de protocol utilitzat. Malgrat que la dinàmica molecular restringida s'ha utilitzat en alguns casos per generar estructures a partir d'una cadena estesa d'un polipèptid, aquesta aproximació presenta l'inconvenient que la major part del temps computacional s'utilitza en l'avaluació energètica per calcular les forces interatòmiques d'estructures que es troben lluny de satisfer les restriccions experimentals.

Un protocol de determinació d'estructures tridimensionals molt utilitzat és el que implica la utilització combinada dels algorismes de geometria de distàncies i la minimització o la dinàmica molecular restringida, en concret, amb el procediment abans esmentat d'atemperat simulat (SA) (Nilges *et al.*, 1988). Els mètodes de DG s'usen per generar un gran nombre d'estructures consistents en la informació experimental. En una segona etapa, es poden minimitzar les energies internes de les estructures resultants. Aquests càlculs que fan ús de dos tipus de metodologia han estat aplicats amb èxit en moltes determinacions estructurals.

Una observació habitual en l'aplicació del protocol híbrid de geometria de distàncies-dinàmica molecular restringida a una gran varietat d'estructures és que les parts menys ben definides de les molècules adopten, després del refinament per DMR, un rang de conformacions més ampli que el que mostraven després de DG i abans del refinament. Això posa de manifest un dels inconvenients associats al mètode de DG: que no mostra prou eficientment l'espai conformacional. Les estructures derivades de DG i sotmeses a un refinament per DMR milloren el grau d'ajust a les dades experimentals i mostren una energia de Van der Waals inferior.

Molt recentment s'ha proposat la utilit-

zació de SA en l'espai de torsions, és a dir, considerant com a úniques variables els angles de torsió (Güntert *et al.*, 1997). Aparentment, aquest mètode és molt eficaç, ja que utilitza un nombre reduït de graus de llibertat i no ha de tenir en compte moviments de freqüències altes i, per tant, pot utilitzar intervals d'integració més llargs.

ANÀLISI I AVALUACIÓ DE LA QUALITAT DE LES ESTRUCTURES GENERADES AMB ELS MÈTODES COMPUTACIONALS

Una determinació estructural completa requereix una darrera etapa d'avaluació del conjunt d'estructures generades. Per a cada una cal comprovar el compliment de les restriccions experimentals (distàncies interprotòniques, angles de torsió, etc.) dins dels límits dels errors experimentals, l'absència de desviacions significatives respecte a la geometria covalent ideal i de mals contactes no enllaçants. D'altra banda, s'ha d'avaluar la dispersió entre les diferents estructures individuals que ens donarà una idea de la precisió de l'estructura generada.

Donat que la informació experimental més important són les restriccions de distància, un indicador útil del grau d'ajust és la suma de les violacions residuals d'aquestes restriccions en les estructures finals. L'anàlisi de les violacions de les restriccions experimentals es pot dur a terme tant en les primeres etapes del càlcul com després del refinament. La violació sistemàtica d'algunes restriccions en les estructures generades pot indicar errors en l'assignació i, per tant, suggereix una revisió de les restriccions experimentals introduïdes en els protocols de càlcul. Cal tenir en compte, però, que les violacions residuals de les distàncies no solament depenen del tipus d'ajust, sinó que també depenen de l'aproximació usada en

la conversió de les intensitats del NOE en distàncies.

Per tal de mesurar el grau d'ajust de les estructures calculades a les mesures experimentals es pot utilitzar un indicador, anomenat factor-R o factor d'estructura, que, per analogia al terme utilitzat en cristal·lografia, és la desviació mitjana normalitzada entre el factor d'estructura derivat d'un model i l'obtingut a partir de les dades experimentals. Per a les dades derivades de RMN, es tracta d'una mesura directa de l'ajust entre les intensitats de NOE experimentals i les teòriques, quan a partir del model final s'ha recalculat l'espectre NOESY (C. González *et al.*, 1991). En la seva forma més simple i general, el factor-R per RMN es pot definir així:

$$R = \frac{[\sum W_{ij}(\tau_m) | A_{ij}^{\text{calc}}(\tau_m) - A_{ij}^{\text{exp}}(\tau_m) |]}{[\sum W_{ij}(\tau_m) A_{ij}^{\text{exp}}(\tau_m)]}$$

on A_{ij}^{calc} i A_{ij}^{exp} són els elements que connecten els protons i i j de les matrius d'intensitat de NOE teòrica i experimental, respectivament, per un temps de barreja τ_m . Els factors $W_{ij}(\tau_m)$ s'inclouen per tenir en compte els errors de mesura. Hi ha diverses definicions similars del factor R i alguns algoritmes fins i tot calculen diversos tipus diferents de factors R simultàniament.

Quan s'usa la matriu iterativa de relaxació per obtenir les estructures, la convergència en un factor-R tan baix com sigui possible se sol utilitzar com a punt final del procés iteratiu del refinament. Ara bé, el factor-R només mesura la concordància entre les dades experimentals i les calculades i , per tant, un valor baix del factor-R no implica necessàriament que tenim una estructura acurada si la precisió de les mesures és pobre. S'ha definit un altre tipus de factor-R, anomenat «factor-R lliure» (Brünger *et al.*, 1993), que intenta solucionar aquest proble-

ma i que es basa en criteris de validació creuada o independent. Es tracta d'un mètode estadístic que estima la qualitat de l'ajust a les dades observades sense fer cap hipòtesi sobre la distribució dels errors en les dades de RMN.

La qualitat dels contactes no enllaçants en les estructures calculades es pot avaluar calculant l'energia de Van der Waals amb el potencial de Lennard-Jones (E_{L-J}) per a totes les estructures. Aquesta energia és molt sensible als mals contactes. Per a una bona estructura, E_{L-J} sol ser negativa, mentre que en presència d'algun mal contacte, és positiva.

En la determinació d'estructures per RMN no se n'obté una de sola, sinó tot un conjunt. Això deriva del fet que el que es busca són estructures compatibles amb un conjunt de dades experimentals de relativa baixa precisió i la solució no és única. La desviació quadràtica mitjana (RMSD) entre els diferents membres d'una família de solucions o respecte a una estructura mitjana dóna una mesura de la precisió amb la qual està definida l'estructura. S'observa, però, que sovint la precisió és molt diferent entre regions. En general, les regions d'estructura secundària o aquelles que presenten un nombre elevat de contactes queden més ben definides, mentre que altres zones mostren una menor precisió. El nombre total de restriccions per residu determina en gran mesura la precisió de l'estructura, tot i que la utilització de restriccions més acurades millora també la qualitat. La presència de mobilitat local té com a efecte general una disminució en la intensitat dels NOE i , per tant, en dificulta l'observació. El resultat és un nombre menor de restriccions i una estructura menys precisa. Aquesta interpretació s'ha pogut confirmar en molts casos fent mesures de temps de relaxació.

L'existència d'equilibris conformacionals ràpids és una font de complicació, ja que la presumpció implícita en tot el procés

de determinació de l'estructura d'una proteïna és que aquesta és única. La presència de contradiccions internes en les dades (especialment entre NOE i constants d'acoblament) és el criteri a partir del qual és detecten aquests equilibris. En aquests casos les estructures s'han de recalculer utilitzant mètodes d'introducció de les restriccions experimentals que siguin compatibles amb la presència d'equilibris conformacionals. Els principals són els mètodes de dinàmica molecular sobre conjunts d'estructures o bé la introducció de restriccions que només s'apliquen si, durant la dinàmica, el conjunt de les darreres estructures no compleixen aquesta restricció. Els conjunts d'estructures resultants òbviament presentaran un grau considerable de variabilitat, però que reflecteix la naturalesa dinàmica de la informació estructural que buscàvem.

EXEMPLES D'ESTRUCTURES 3D DE PROTEÏNES DETERMINADES PER RMN

La figura 5 mostra la superposició del conjunt de les divuit estructures finals calculades a partir de les dades de RMN per a la kaliotoxina, un potent inhibidor de canals de potassi present en un verí d'escorpió. A la



FIGURA 5. Superposició dels àtoms de l'esquelet peptídic de les divuit estructures finals calculades per a la kaliotoxina

figura 6 (annex 2) es mostra una representació de la mateixa estructura feta amb el programa MolScript (Kraulis, 1991), que permet visualitzar fàcilment els elements d'estructura secundària de les proteïnes. A la figura 7 es presenten les seqüències d'altres toxines homòlogues. L'estructura tridimensional de les diferents toxines és molt similar: conté una hèlix α i un full β format per dues làmines disposades en sentit antiparal·lel. L'hèlix α i el full β estan connectats per dos dels tres ponts disulfur que estableixen el plegament de tota la família de toxines. A la taula 2 es detallen les restriccions experimentals utilitzades i la desviació quadràtica mitjana del conjunt d'estructures de les diferents toxines.

Comparant les diferents estructures pot veure's com la utilització de camps més alts, que permet obtenir un nombre més elevat de restriccions, condueix a estructures amb menor dispersió. Igualment pot veure's que la inclusió d'assignacions estereoespecífiques té una gran importància en la precisió de les estructures resultants. La utilització de marcatge uniforme amb ^{13}C i ^{15}N permet la mesura de nombres més alts de restriccions, especialment restriccions que impliquen protons situats a les cadenes laterals on l'encobriment sovint és més important. Per a l'estudi de proteïnes més grans, la utilització de marcatge és imprescindible.

Un exemple de determinació estructural d'una proteïna totalment marcada és el de la interleucina 1β , (G. M. Clore *et al.*, 1991) de 153 residus i 17,4 kDa de pes molecular, duta a terme a partir de 3.146 restriccions experimentals, de les quals 2.630 eren restriccions de distància interprotoniques. Les estructures finals presentaven una RMSD atòmica de 0,41 Å per als àtoms de l'esquelet peptídic i de 0,82 Å per a tots els àtoms.

Actualment, el límit per a la mida de les proteïnes que poden estudiar-se per RMN

TAULA 2. Comparació de la informació estructural derivada de la RMN per a la determinació de l'estructura tridimensional de diverses toxines d'escorpi

TOXINA (nombre aa)	Freqüència 1H	Marcatge	Restriccions experimentals derivades de la RMN				RMSD Tots àtoms	Estructures finals Esquelet peptidic	Referència
			Distàncies (NOE inter)	Distàncies (enllaços H)	Angles diedres (°)	Assignacions estereosp.			
ChTX(37)	500 MHz	No	110	4	29		1,84 Å	Bontems <i>et al.</i> , 1991	
IbTX (37)	500 MHz	No	148	5	15		1,49 Å	Johnson <i>et al.</i> , 1992	
MgTX(39)	600 MHz	¹⁵ N i ¹³ C	253	10	60		1,32 Å	0,68 Å Johnson <i>et al.</i> , 1994	
NTX (39)	600 MHz	No	275	7	42		0,75 Å	Dauplais <i>et al.</i> , 1995	
AgTX2 (38)	600 MHz	¹⁵ N	309	9	43	21	0,86 Å	0,38 Å Krezel <i>et al.</i> , 1995	
KTX (38)	600 MHz	No	218	6	29	12	1,09 Å	0,46 Å Aiyar <i>et al.</i> , 1995	
KTX (38)	500 MHz	No	189	10	23	7	1,42 Å	0,68 Å Gairi <i>et al.</i> , 1997	

no l'imposa ja l'encobriment de senyals sinó la pèrdua de sensibilitat causada per l'eixamplament de l'amplada de banda associat a l'augment del temps de correlació i

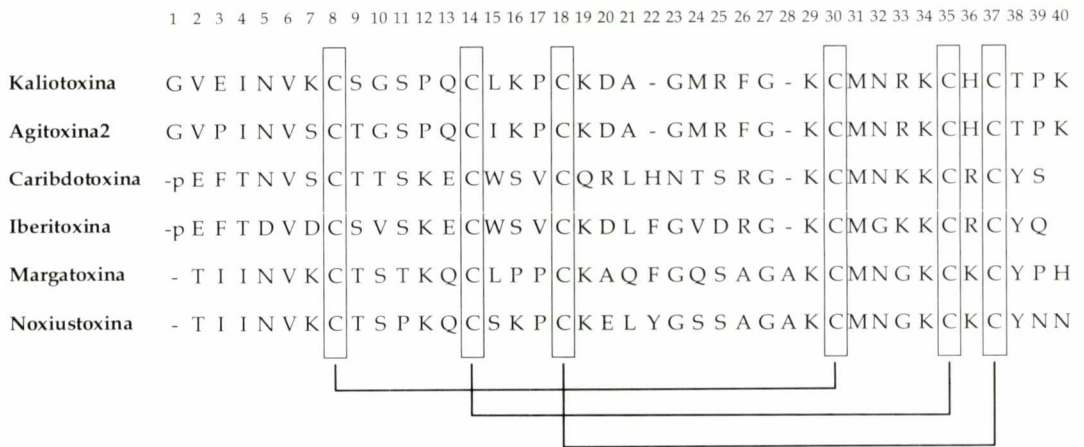


FIGURA 7. Comparació de les seqüències de diferents toxines d'escorpi. Els ponts disulfur que connecten els residus de *cis* estan indicats a la part inferior de la figura

la velocitat de relaxació espín-espín. De fet, la transferència de magnetització sobre la qual es basen els diferents experiments 3D i 4D requereix que l'amplada de línia sigui més petita que la constant d'acoblament utilitzada per establir la correlació. L'estratègia utilitzada actualment per disminuir l'amplada de banda consisteix en la substitució de molts dels protons de la molècula per deuteris (per expressió de la proteïna en un medi deuterat) de manera que la relaxació dipolar dels protons restants –que permetran determinar l'estructura– sigui menys efectiva.

La combinació del marcatge fraccional de les biomolècules amb deuteri, el marcatge uniforme amb ^{13}C i ^{15}N i els experiments de RMN multidimensionals han permès recentment determinar estructures tridimensionals de proteïnes de 30 kDa de pes molecular. Una de les proteïnes monomèriques més grans per a la qual s'ha determinat l'estructura tridimensional per RMN és el domini N-terminal de l'enzim I del sistema fosfoenolpiruvat, fosfotransferasa de sucres, que consta de 259 residus i presenta un pes molecular de 30 kDa. La determinació estructural (D. S. Garrett *et al.*, 1997) està basada en un conjunt de 4.251 restriccions experimentals derivades de la RMN i la precisió de les coordenades per a les cinquanta estructures calculades finals és de $0,79 \pm 0,18$ Å per als àtoms de l'esquelet peptídic i $d'1,06 \pm 0,15$ Å per a tots els àtoms de la proteïna. L'estructura derivada per RMN és molt similar a la determinada també recentment per raigs X.

Cal destacar, com una frontera on encara han de passar moltes coses, els resultats recents en l'assignació de proteïnes de membrana utilitzant tècniques de RMN d'estat sòlid (Marassi *et al.*, 1997) que obren la via per a estudis tridimensionals d'aquesta important, i inconeguda, gran classe de proteïnes.

DINÀMICA DE PROTEÏNES PER RMN

La funció de moltes proteïnes està associada a processos que evolucionen amb el temps. La RMN és una de les eines més versàtils per a l'estudi de processos dinàmics a diferents escales de temps. En tots els casos, però, caldrà inicialment completar l'assignació de l'espectre i obtenir una estructura –amb totes les salvetats derivades de l'existència d'equilibris conformacionals– per poder analitzar els resultats dinàmics.

Un exemple de procés lent (habitualment en l'escala de temps dels segons als mil·lsegons) és el plegament de la proteïna a partir d'una forma desnaturalitzada. En aquest tipus d'estudi s'han utilitzat amb èxit els mètodes de bescanvi d'hidrogen per deuteri que, un cop assignat l'espectre, permeten la identificació de les regions on s'inicia el plegament o de possibles intermediaris.

Processos més ràpids (en l'escala de temps dels mil·lsegons als microsegons) es troben, per exemple, en equilibris entre formes que difereixen en la conformació de ponts disulfur o bé en l'orientació de cadenes laterals. Les mesures de temps de relaxació en el sistema de coordenades giratori ofereixen informació quantitativa sobre la velocitat d'aquests processos.

Finalment, cal esmentar els moviments que tenen lloc en l'escala dels picosegons i que afecten la relaxació dels diferents nuclis que poden donar senyal per RMN. Aquests processos, entre els quals es troba la difusió rotacional, són els causants de la modulació de les interaccions que donen lloc als NOE que permeten l'obtenció de les estructures tridimensionals. La utilització d'aquest mateix tipus de mesures –NOE, temps de relaxació– per caracteritzar la dinàmica interna de les proteïnes només és possible si es coneix *a priori* la distància entre els nuclis implicats. En el cas de nuclis de ^{15}N amb un protó directament unit, la distància és simplement

la distància d'enllaç, i les mesures de relaxació proporcionen informació dinàmica sobre fluctuacions que tenen lloc a freqüències determinades. Les dades experimentals poden contrastar-se amb les prediccions de diferents models de moviment (Palmer *et al.*, 1996) o bé interpretar-se de forma genèrica –sense recórrer a un model explícit– tenint en compte uns paràmetres que proporcionin una idea sobre l'amplitud de les fluctuacions i la seva escala de temps (Lipari i Szabo, 1982a, 1982b). Un estudi recent sobre la dihidrofolat reductasa (Epstein *et al.*, 1995) posa de manifest les diferents escales de temps implicades en la mobilitat de proteïnes i la seva relació amb els processos de reconeixement molecular. La imatge dinàmica de les proteïnes a la qual tenim accés a través de la RMN ha de completar la nostra visió estructural i és possible que ens ofereixi noves claus sobre els seus aspectes funcionals.

AGRAÏMENTS

Els autors volen agrair les llargues discussions i l'encoratjament d'Ernest Giralt. L'espectre 3D de la deshidrodidemnina ha estat gentilment cedit per Francisco Cárdenas i Miguel Feliz. Imma Fernández ha participat també en el projecte de determinació d'estructures de toxines. La recerca del grup descrita en aquest treball ha estat finançada per la CICYT (PB94-0924) i la Generalitat de Catalunya (Centre de Referència en Biotecnologia) i ha utilitzat els instruments dels Serveis Científicotècnics de la Universitat de Barcelona.

BIBLIOGRAFIA GENERAL

CAVANAGH, J.; FAIRBROTHER, W. J.; PALMER, A. G.; SKELTON, N. J. (1996). «Protein NMR Spectroscopy. Principles and Practice». Academic Press, San Diego.

CLORE, G. M.; GRONENBORN, A. M. (1991). «Two- three- and four-dimensional NMR methods for obtaining larger and more precise three-dimensional structures of proteins in solution». *Annu. Rev. Biophys. Biophys. Chem.*, núm. 20, pàg. 29-63.

EVANS, J. N. S. (1995). «Biomolecular NMR Spectroscopy». Oxford University Press, Oxford.

JAMES, T. L.; OPPENHEIMER, N. J. Eds. (1994). «Methods in Enzymology Vol. 239: Nuclear Magnetic Resonance». Academic Press.

NEUHAUS, D.; WILLIAMSON, M. (1989). «The Nuclear Overhauser effect in Structural and Conformational Analysis». VCH Publishers, Inc, Nova York.

NILGES, M. (1996). «Structure calculation from NMR data». *Current Opinion in Structural Biology*, núm. 6 (5), pàg. 617-623.

WÜTHRICH, K. (1986). «NMR of Proteins and Nucleic Acids». Wiley, Nova York.

BIBLIOGRAFIA CITADA

AIYAR, J.; WITHKA, J. M.; RIZZI, J. P.; SINGLETON, D. H.; ANDREWS, G. C.; LIN, W.; BOYD, J.; HANSON, D. C.; SIMON, M.; DETHLEFS, B.; LEE, C.; HALL, J. E.; GUTMAN, G. A.; CHANDY, K. G. (1995). «Topology of the pore-region of a K⁺ channel revealed by the NMR-derived structures of scorpion toxins». *Neuron*, núm. 15, pàg. 1169-1181.

AUE, W. P.; BARTHOLDI, E.; ERNST, R. R. (1976). «Two-dimensional spectroscopy. Application to nuclear magnetic resonance». *J. Chem. Phys.*, núm. 64, pàg. 2229-2246.

BAX, A.; GRIFFEY, R. H.; HAWKINS, B. L. (1983). «Correlation of proton and nitrogen-15 chemical shifts by multiple quantum NMR». *J. Magn. Reson.*, núm. 55, pàg. 301-315.

BAX, A.; DAVIES, D. G. (1985). «Practical aspects of two-dimensional transverse NOE spectroscopy». *J. Magn. Reson.*, núm. 63, pàg. 207-213.

BAX, A.; CLORE, G. M.; GRONENBORN, A. M. (1990). «¹H-¹H correlation via isotropic mixing of ¹³C magnetization. A new three-dimensional approach for assigning ¹H and ¹³C spectra of ¹³C-enriched proteins». *J. Magn. Reson.*, núm. 88, pàg. 425-431.

BAX, A.; IKURA, M. (1991). «An efficient 3D NMR technique for correlating the proton and ¹⁵N backbone amide resonances with the alpha carbon of the preceding residue in uniformly ¹⁵N/¹³C enriched proteins». *J. Biomol. NMR*, núm. 1, pàg. 99-104.

BODENHAUSEN, G.; RUBEN, D. J. (1980). «Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy». *Chem. Phys. Lett.*, núm. 69, pàg. 185-189.

- BONTEMS, F.; ROUMESTAND, C.; BOYOT, P.; GILQUIN, B.; Y VON DOLJANSKY; MENEZ, A.; TOMA, F. (1991). «Three-dimensional structure of natural charybdotoxin in aqueous solution by $^1\text{H-NMR}$ ». *Eur. J. Biochem.*, núm. 196, pàg. 19-28.
- BOTHNER-BY, A. A.; STEPHENS, R. L.; LEE, J.; WARREN, C. D.; JEANLOZ, R. W. (1984). «Structure determination of a tetrasaccharide: transient nuclear overhauser effects in the rotating frame». *J. Am. Chem. Soc.*, núm. 106, pàg. 811-813.
- BRAUNSCHWEILER, L.; ERNST, R. R. (1983). «Coherence transfer by isotropic mixing: application to proton correlation spectroscopy». *J. Magn. Reson.*, núm. 53, pàg. 521-528.
- BROOKS, B. R.; BRUCCOLERI, R. E.; OLAFSON, B. O.; STATES, D. J.; SWANINATHAN, S.; KARPLUS, M. (1983). «CHARMM: a program for macromolecular energy, minimization and dynamics calculation». *J. Comput. Chem.*, núm. 4, pàg. 187.
- BRÜNGER, A.; CLORE, G.; GRONENBORN, A.; SAFFRICH, R.; NILGES, M. (1993). «Assessing the quality of solution nuclear magnetic resonance structures by complete cross-validation». *Science*, núm. 261, pàg. 328-331.
- CONTRERAS, M. A.; HAACK, T.; ROYO, M.; GIRALT, E.; PONS, M. (1997). «Temperature coefficients of peptides dissolved in hexafluoroisopropanol (HFIP) monitor distortions of helices». *Lett. Pept. Sci.*, núm. 4, pàg. 29-39.
- CLORE, G. M.; WINGFIELD, P. T.; GRONENBORN, A. M. (1991). «High-resolution three-dimensional structure of Interleukin 1 β in solution by three- and four-dimensional nuclear magnetic resonance spectroscopy». *Biochemistry*, núm. 30, pàg. 2315-2323.
- CLORE, G. M.; GRONENBORN, A. M. (1994). «Multidimensional Heteronuclear Nuclear Magnetic Resonance of Proteins». *Methods Enzymol.*, núm. 239, pàg. 349-363.
- CLUBB, R. T.; THANABAL, V.; WAGNER, G. (1992). «A constant-time three dimensional triple-resonance pulse scheme to correlate intrasidue ^1H , ^{15}N and ^{13}C chemical shifts in ^{15}N - ^{13}C labeled proteins». *J. Magn. Reson.*, núm. 97, pàg. 213-217.
- DAUPLAIS, M.; GILQUIN, B.; POSSANI, L. D.; GURROLA-BRIONES, G.; ROUMESTAND, C.; MENEZ, A. (1995). «Determination of the three-dimensional solution structure of Noxiustoxin: analysis of structural differences with related short-chain scorpion toxins». *Biochemistry*, núm. 34, pàg. 16563-16573.
- EPSTEIN, D. M.; BENKOVIC, S. J.; WRIGHT, P. E. (1995). «Dynamics of dihydrofolate reductase-folate complex: catalytic sites and regions known to undergo conformational change exhibit diverse dynamic features». *Biochemistry*, núm. 34, pàg. 11037-11048.
- GAIRÍ, M.; ROMI, R.; FERNÁNDEZ, I.; ROCHAT, H.; MARTIN-EAUCLAIRE, M-F.; VAN RIETSCHOTEN, J.; PONS, M.; GIRALT, E. (1997). «3D-Structure of kaliotoxin: is residue 34 a key for channel selectivity?». *J. Pept. Sci.*, núm. 3, pàg. 314-319.
- GARRETT, D. S.; SEOK, Y-J.; LIAO, D. I.; PETERKOFKY, A.; GRONENBORN, A. M.; CLORE, G. M. (1997). «Solution structure of the 30KDa N-terminal domain of enzyme I of the *Escherichia coli* phosphoenolpyruvate:sugar phosphotransferase system by multidimensional NMR». *Biochemistry*, núm. 36, pàg. 2517-2530.
- GONZÁLEZ, C.; RULLMAN, J. A. C.; BONVIN, A. M. J. J.; BOELEN, R.; KAPTEIN, R. (1991). «Toward and NMR R-factor». *J. Mag. Reson.*, núm. 91, pàg. 659-664.
- GRZESIEK, S.; BAX, A. (1992a). «Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR». *J. Am. Chem. Soc.*, núm. 114, pàg. 6291-6293.
- GRZESIEK, S.; BAX, A. (1992z). «An efficient experiment for sequential backbone assignments of medium-sized isotopically enriched proteins». *J. Magn. Reson.*, núm. 99, pàg. 201-207.
- GÜNTERT, P.; BRAUN, W.; WÜTHRICH, K. (1991). «Efficient computation of three-dimensional protein structures in solution from magnetic resonance data using the program DIANA and the supporting program CALIBA, HABAS and GLOMSA». *J. Mol. Biol.*, núm. 217, pàg. 517-530.
- GÜNTERT, P.; MUMENTHALER, C.; WÜTHRICH, K. (1997). «Torsion angle dynamics for NMR structure calculation with the new program DYANA». *J. Mol. Biol.*, núm. 273, pàg. 283-298.
- HAVEL, T. F.; WÜTHRICH, K. (1984). *Bull. Math. Biol.*, núm. 46, pàg. 673-698.
- JEENER, J.; MEIER, B. H.; BACHMANN, P.; ERNST, R. R. (1979). «Investigation of exchange processes by two-dimensional NMR spectroscopy». *J. Chem. Phys.*, núm. 71, pàg. 4546-4553.
- JOHNSON, B. A.; SUGG, E. E. (1992). «Determination of the three-dimensional structure of Iberitoxin in solution by ^1H nuclear magnetic resonance spectroscopy». *Biochemistry*, núm. 31, pàg. 8151-8159.
- JOHNSON, B. A.; STEVENS, S. P.; WILLIAMSON, J. M. (1994). «Determination of the three-dimensional structure of Margatotoxin by ^1H , ^{13}C , ^{15}N triple-resonance nuclear magnetic resonance spectroscopy». *Biochemistry*, núm. 33, pàg. 15061-15070.
- KARPLUS, M. (1959). «Contact electron-spin coupling of nuclear magnetic moments». *J. Chem. Phys.*, núm. 30, pàg. 11-15.
- KAY, L. E.; MARION, D.; BAX, A. (1989). «Practical aspects of 3D heteronuclear NMR of proteins». *J. Magn. Reson.*, núm. 84, pàg. 72-84.
- KAY, L. E.; IKURA, M.; TSCHUDIN, R.; BAX, A. (1990). «Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins». *J. Magn. Reson.*, núm. 89, pàg. 496-514.
- KREZEL, A. M.; KASIBHATLA, C.; HIDALGO, P.; MACKINNON,

- R.; WAGNER, G. (1995). «Solution structure of the potassium channel inhibitor agitoxin 2: Caliper for probing channel geometry». *Protein Sci.*, núm. 4, pàg. 1478-1489.
- KESSLER, H.; GEHRKE, M.; GRIESINGER, C. (1988). «Two-dimensional NMR spectroscopy: background and overview of experiments». *Angew. Chem. Int. Ed. Engl.*, núm. 27, pàg. 490-536.
- KRAULIS, P. J. (1991). «MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures». *J. Appl. Crystallogr.*, núm. 24, pàg. 946-950.
- LE, H.; PEARSON, J. G.; DE DIOS, A. C.; OLDFIELD, E. (1995). «Protein structure refinement and prediction via NMR chemical shifts and quantum chemistry». *J. Am. Chem. Soc.*, núm. 117, pàg. 3800-3807.
- LIPARI, G.; SZABO, A. (1982a). «Model free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity». *J. Am. Chem. Soc.*, núm. 104, pàg. 4546-4559.
- LIPARI, G.; SZABO, A. (1982b). «Model free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Analysis of experimental results». *J. Am. Chem. Soc.*, núm. 104, pàg. 4559-4570.
- MARASSI, F. M.; RAMAMOORTHY, A.; OPELLA, S. J. (1997). «Complete resolution of the solid state NMR spectrum of a uniformly N15 labeled membrane protein in phospholipid bilayers». *Proc. Nat. Acad. Sci. USA*, núm. 94, pàg. 8551-8556.
- MÜLLER, L. (1979). «Sensitivity enhanced detection of weak nuclei using heteronuclear multiple quantum coherence». *J. Am. Chem. Soc.*, núm. 101, pàg. 4481-4484.
- NILGES, M.; CLORE, G. M.; GRONENBORN, A. M. (1988). «Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations». *FEBS Lett.*, núm. 229, pàg. 317-324.
- OSAPAY, K.; CASE, D. A. (1991). «A new analysis of proton chemical shifts in proteins». *J. Am. Chem. Soc.*, núm. 113, pàg. 9436-9444.
- OSCHKINAT, H.; CIESLAR, C.; GRONENBORN, A. M.; CLORE, G. M. (1989). «Three-dimensional homonuclear Hartmann-Hahn nuclear Overhauser enhancement spectroscopy in H₂O and its application to proteins». *J. Magn. Reson.*, núm. 81, pàg. 212-216.
- OSCHKINAT, H.; CROFT, D. (1994). «Automated assignment of multidimensional nuclear magnetic resonance spectra». *Methods Enzymol.*, núm. 239, pàg. 308-318.
- PALMER III, A. G.; WILLIAMS, J.; MCDERMOTT, A. (1996). «Nuclear magnetic resonance studies of biopolymer dynamics». *J. Phys. Chem.*, núm. 100, pàg. 13293-13310.
- SPERA, S.; BAX, A. (1991). «Empirical correlation between protein backbone conformation and C α and C β ¹³C nuclear magnetic resonance chemical shifts». *J. Am. Chem. Soc.*, núm. 113, pàg. 5490-5492.
- VAN DE VEN, F. J. M. (1995). «Multidimensional NMR in Liquids. Basic Principles and Experimental Methods». VCH.
- WEINER, P. K.; KOLLMAN, P. A. (1981). «AMBER: assisted model building with energy refinement. A general program for modeling molecules and their interaction». *J. Comput. Chem.*, núm. 2, pàg. 287-303.
- WEINER, S. J.; KOLLMAN, P. A.; CASE, D. A.; SINGH, U. C.; GHIO, C.; ALAGONA, G.; PROFETA, S.; WEINER, P. (1984). «A new force field for molecular mechanical simulation of nucleic acids and proteins». *J. Am. Chem. Soc.*, núm. 106, pàg. 765-784.
- WIJMEGA, S. S.; HALLENGA, K.; HILBERS, C. W. (1989). «A three-dimensional multiple-quantum coherence homonuclear Hartmann-Hahn experiment». *J. Magn. Reson.*, núm. 84, pàg. 634-642.
- WÜTHRICH, K. (1986). «NMR of proteins and nucleic acids». Wiley, Nova York.