

00577



UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO

FACULTAD DE QUÍMICA

4

PROGRAMA DE MAESTRIA Y DOCTORADO EN  
CIENCIAS QUÍMICAS

DETERMINACION DE LA ESTRUCTURA  
SECUNDARIA DEL FRAGMENTO 1 - 108 DE LA  
PROTEINA SV40 POR RMN

29/1976

**T E S I S**  
QUE PARA OBTENER EL GRADO DE  
**MAESTRO EN CIENCIAS QUÍMICAS**  
**(QUÍMICA ANALÍTICA)**  
P R E S E N T A  
**Q. MARCO ANTONIO VERA RAMÍREZ**

DIRECTOR DE LA TESIS: DR. J. FEDERICO DEL RÍO PORTILLA



MEXICO, D. F.

2001



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Jurado asignado

Presidente Dr. Lorenzo Segovia Forcella  
Primer vocal Dr. Carlos Martín Cerda García-Rojas  
Secretario M. en C. Ricardo Jorge Cárdenas Pérez

Primer suplente Dra. Araceli Patricia Peña Álvarez

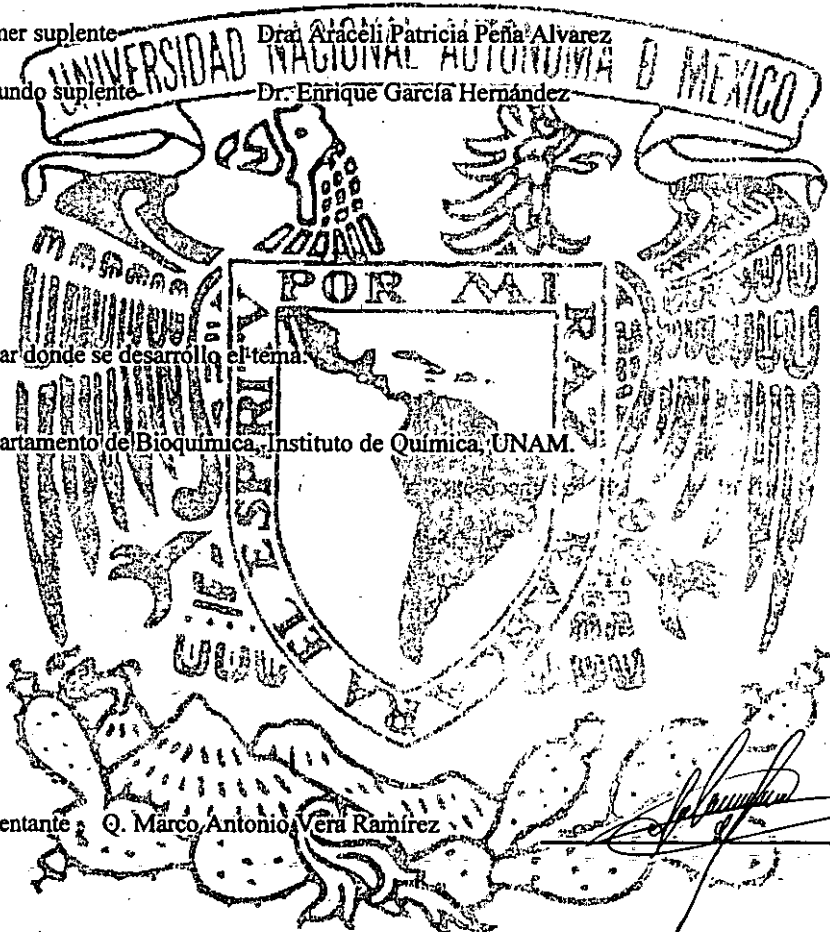
Segundo suplente Dr. Enrique García Hernández

Lugar donde se desarrolló el tema.

Departamento de Bioquímica, Instituto de Química, UNAM.

Sustentante Q. Marco Antonio Vera Ramírez

Asesor Dr. José Federico del Río Portilla



Two handwritten signatures are present. The first signature is written in black ink and appears to be 'Marco Antonio Vera Ramírez'. The second signature is written in blue ink and appears to be 'José Federico del Río Portilla'.

*A mis amigos:*


Abraham, Acely, Adela, Alba A., Alfonso, Ana, Blanquita N., Carmelita, Cristina (Diabólica), David, David, Eliseo, Ernesto, Ethel, Georgito, Guillermo, Héctor, Isaias, Javier (Pirrol), Jorge, Juanito, Juanita, Laguna, Larissa, Leticia, Martita, Mauricio (Perrostro), Miguel (Huerta), Miguel, Paty, Perfecto, Rafita (P3), Roberto, Rosalba, Sandra, Sinoe, Viviana, Williams, Itzcóatl Alvarez M., José A. Bautista M., Guillermo Bolaños, Graciela Canseco M., Roberto Castro, Carlos Cosío, Marco A. Cruz S., Rafael De León C., T. Jesús Espinosa F., F. Javier, Esquivel F., Ricardo Figueroa G., Enrique G. Jardón, Ely Hernández M., Dulce M. Juárez M., F. Humberto, León A., Gustavo Lozano, Saúl Luis J., Miguel A. Martínez S., María L. Medina G., Raquel Meneses E., Blanca Estela Mojica García, Sofia H. Montalvo J., R. Hilda Navarro H., M Paz Orta V., Juan C. Piliado V., Claudia R. Lugo, Rodrigo Ramírez H., Carlos Reyna T., Francisco Rivera, Roberto C. Rodríguez P., Aldo Sánchez G., Olga Laura Santamaría Rangel, Anabel Sierra R., Fabián J. Urbina G., Edgar Vargas M., Katia V. Vázquez G.

Les agradezco a todos su apoyo, su confianza, su valiosa amistad, los momentos que hemos pasamos juntos (buenos y malos), sus consejos, el haberme escuchado.

Sobre todo, quiero agradecerles el respeto con el que hemos llevado nuestra amistad, parte básica en nuestra vida y que todos merecemos, a pesar de nuestras diferencias, y que ha hecho que nuestra convivencia sea mejor y sobre el cual cimentaremos una mejor amistad muy duradera.

Le agradezco a los amigos de la Huelga que me hayan mostrado la realidad del país... las grandes diferencias que subsisten; por ustedes aprendí que (entre muchas otras cosas) las ideas nos separan... espero que la inteligencia nos una...

Por todo, a todos, gracias.



Quiero agradecerle de forma muy especial a Federico del Río Portilla el haberme brindado la oportunidad de trabajar con él, su paciencia para repetirme millones de veces las cosas que nunca entendí, por explicarme muchas cosas más allá de lo puramente académico y por su apoyo.

Pero sobretodo, quiero agradecerte tu amistad, la confianza y el apoyo que me has brindado... pocos saben ser amigos como tú.

AGRADECIMIENTOS NACIONAL AUTÓNOMO DE B...  
Agradezco a los profesores:

Dr. Barbarín

M. en C. María del Rosario

Dr. Federico

Dr. Enrique

M en C. Fernando

Dr. Asdrubal Almicar

Dra. Luz Elena

Arreguín

Covarrubias Herrera

del Río Portilla

García

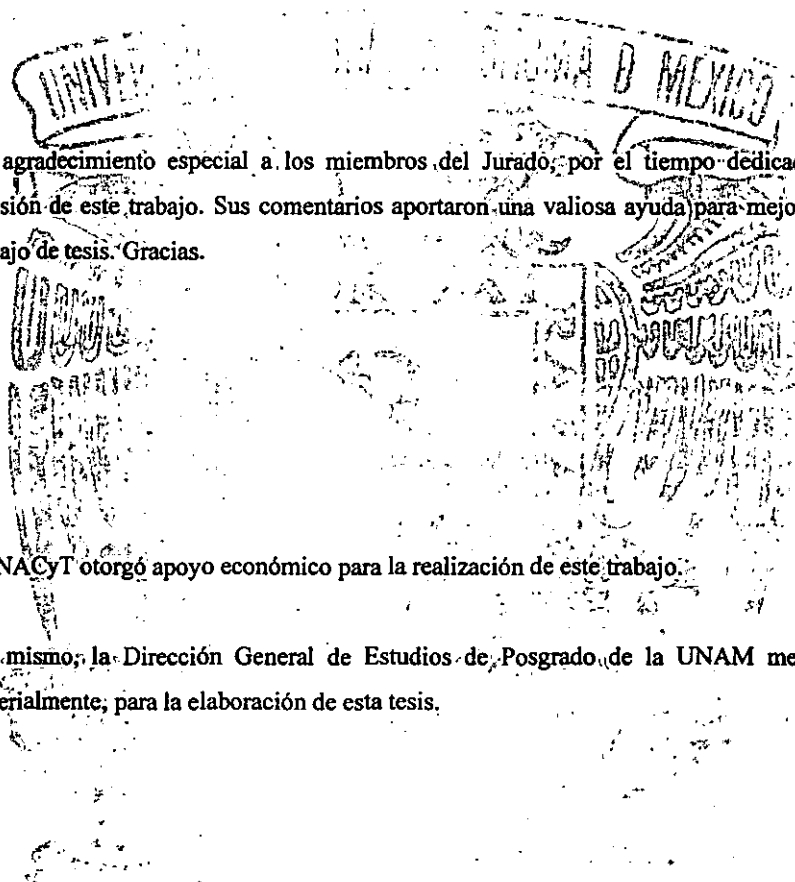
León Cedeño

Paz Larralde

Vera Avila

Por sus enseñanzas, su confianza, el apoyo que siempre me brindaron, por sus las palabras de aliento, por enseñarme la disciplina del trabajo; por hacerme ver que sólo con el esfuerzo se pueden lograr los objetivos trazados en la vida.

Especialmente les agradezco el que me orientaran en el camino del conocimiento y a librarme de mis prejuicios, para acercarme un poco a la verdad...



Un agradecimiento especial a los miembros del Jurado, por el tiempo dedicado a la revisión de este trabajo. Sus comentarios aportaron una valiosa ayuda para mejorar este trabajo de tesis. Gracias.

CONACyT otorgó apoyo económico para la realización de este trabajo.

Así mismo, la Dirección General de Estudios de Posgrado de la UNAM me apoyó materialmente, para la elaboración de esta tesis.

Capítulo	Página
<i>Introducción y objetivos</i> .....	1
<i>Proteínas, generalidades</i> .....	3
Proteína SV40[1-108].....	7
<i>Resonancia Magnética Nuclear, generalidades</i> .....	11
Obtención de las restricciones	
Señales por el Efecto Nuclear Overhauser (NOE).....	21
Constantes de acoplamiento escalar ( $^3J$ ).....	24
Desplazamiento químico.....	25
Cálculo de las estructuras terciarias a partir de los datos de RMN	
Geometría de distancias.....	26
Minimización de la energía y dinámica molecular restringida.....	29
<i>Desarrollo experimental</i> .....	32
Identificación de las señales del espectro 3D TOCSY-HSQC, ayudado por el experimento 2D HSQC, para una asignación probable a determinado grupo de aminoácidos.....	33
Secuenciación de las franjas asignadas a cada residuo, para su ordenación dentro de la estructura primaria, utilizando el espectro 3D NOESY-HSQC	35
Identificación y asignación de las señales de correlación a larga distancia.	37
Identificación de los elementos regulares de la estructura secundaria de la proteína SV40[1-108], a través del índice de protón.....	40
Cálculo de las estructuras.....	43
<i>Análisis de resultados y discusión</i> .....	44
<i>Conclusiones</i> .....	63
<i>Bibliografía</i> .....	64
<i>Anexo 1</i> .....	66
<i>Anexo 2</i> .....	68
<i>Anexo 3</i> .....	82



## Introducción y objetivos

El estudio estructural de proteínas ha permitido en años recientes encontrar las relaciones estructura-actividad. Su aplicación más importante, aparte de la comprensión de los procesos de biología molecular, es en ingeniería genética para el desarrollo de nuevos fármacos analizando el sitio activo de la proteína, determinando su estructura.

La Resonancia Magnética Nuclear (RMN) se ha establecido como una de las herramientas más importantes en la elucidación estructural de compuestos químicos. Actualmente dicha técnica ha evolucionado a un nivel tal, que pareciera que ningún problema es lo suficientemente complejo como para que esta no lo pueda resolver. Hoy en día se puede elucidar prácticamente cualquier estructura química, ya sea de compuestos orgánicos o inorgánicos, siempre que dicha estructura esté constituida por núcleos magnéticamente activos.

Desde hace unos 13 años que la RMN incursiona en el área bioquímica para elucidar la estructura tridimensional de las proteínas y en años recientes el número de estructuras proteicas que se han resuelto mediante RMN, ha ido en aumento. La complejidad y el tamaño de estos compuestos, esenciales para la vida, complica mucho la interpretación de los espectros de RMN debido a la superposición de las señales. Muchos han sido los esfuerzos para incrementar el tamaño límite de las proteínas que se pueden estudiar y agilizar el procesamiento de los datos mediante el marcaje isotópico selectivo con  $^{15}\text{N}$ ,  $^{13}\text{C}$  y  $^2\text{H}$ , el desarrollo de nuevas secuencias de pulsos para obtener espectros de 2, 3 y 4 dimensiones y el procesamiento automático. Sin embargo, estos trabajos permiten la obtención de la estructura después de un trabajo continuo de dos a tres años.

Conocer la estructura tridimensional de las proteínas es útil en algunas aplicaciones de la biología molecular, ingeniería genética o en el desarrollo de nuevos fármacos, pero la técnica para resolver estructura de proteínas es un área que no ha sido suficientemente explotada en México, esto como consecuencia de que no existe la infraestructura adecuada ni el personal capacitado para la realización de dichas tareas.

Los objetivos del presente trabajo son:

- ✦ Cimentar las bases para utilizar la Resonancia Magnética Nuclear en la elucidación de la estructura de proteínas en México.
- ✦ Determinar la estructura secundaria de la proteína SV40 por RMN.
- ✦ Obtener la estructura terciaria de la SV40, a baja resolución, utilizando espectros 3D de Resonancia Magnética Nuclear.



## Proteínas, generalidades<sup>1</sup>

Las proteínas se encuentran presentes en todos los seres vivos, son biopolímeros formados por diferentes combinaciones de los 20 aminoácidos más comúnmente encontrados en la naturaleza. El número de proteínas presentes en un organismo puede ser tan pequeño como media docena, como en el caso del virus bacteriófago M13, o tan grande como 50000, como en el caso del ser humano. Las proteínas están codificadas en el ácido desoxirribonucleico (ADN) presente en todas las células.

Las funciones de las proteínas son muy variadas. Una gran cantidad de reacciones bioquímicas dentro de las células son catalizadas por proteínas enzimáticas. Por ejemplo, en procesos fermentativos de ciertos alimentos como en la panificación, en la fermentación alcohólica (vinos y cervezas) o manufactura de quesos, se encuentra involucrada esta acción enzimática. Otras proteínas están involucradas en el transporte de electrones, iones y moléculas pequeñas. Las proteínas son componentes clave del sistema inmunológico y en el control de la expresión genética de otras proteínas. Así como hay una gran diversidad en las funciones de las proteínas, hay un amplio espectro de pesos moleculares que van desde los 4 kdaltons hasta 2,700 kDa.

La biología molecular está centrada en el hecho de que el ADN es transcrito en ARN y éste se traduce en una proteína. El ADN, y ciertos tipos de ARN, son transcritos en cadenas de ARN mensajero (mARN, proceso que generalmente ocurre dentro del núcleo de los eucariotes) y la traducción en proteínas ocurre en los ribosomas, que son los organelos especializados en esta tarea. Generalmente las proteínas son sintetizadas con un fragmento de 15 - 26 aminoácidos extra, los cuales sirven como señal para que sean translocadas a través de las membranas celulares; una vez que las proteínas han llegado al sitio en el que se necesitan, estos fragmentos son eliminados. Ciertas enzimas y hormonas proteicas pueden sufrir eliminaciones posteriores hasta lograr su estructura activa. Además, las proteínas pueden sufrir modificaciones covalentes posteriores a su síntesis. Estas pueden ser la *acetilación* del amino terminal o la *amidación* del carbonilo terminal; adición de ácidos grasos o lípidos, por ejemplo el farnesilo que mantiene a la proteína unida a la membrana celular; *glicosilación*, una de las modificaciones más comunes, que tiene diversas finalidades, siendo la más común el reconocimiento celular mediante su interacción con otras moléculas; *fosforilación*, que sirve para modificar la afinidad por activadores o

inhibidores, principalmente en la transducción de señales; o la formación de *enlaces disulfuro*, en el que cisteínas separadas en la secuencia, quedan unidas mediante un enlace covalente, manteniendo juntas dos regiones separadas.

Las diferentes combinaciones de las cadenas laterales de los aminoácidos, proporcionan una estructura y función única a cada proteína. A pesar de que el tamaño y las propiedades químicas de las cadenas laterales, de los aminoácidos, son muy variadas, éstas pueden ser clasificadas en tres categorías generales: hidrófobas, polares/cargadas y polares neutras (figura 1.1).

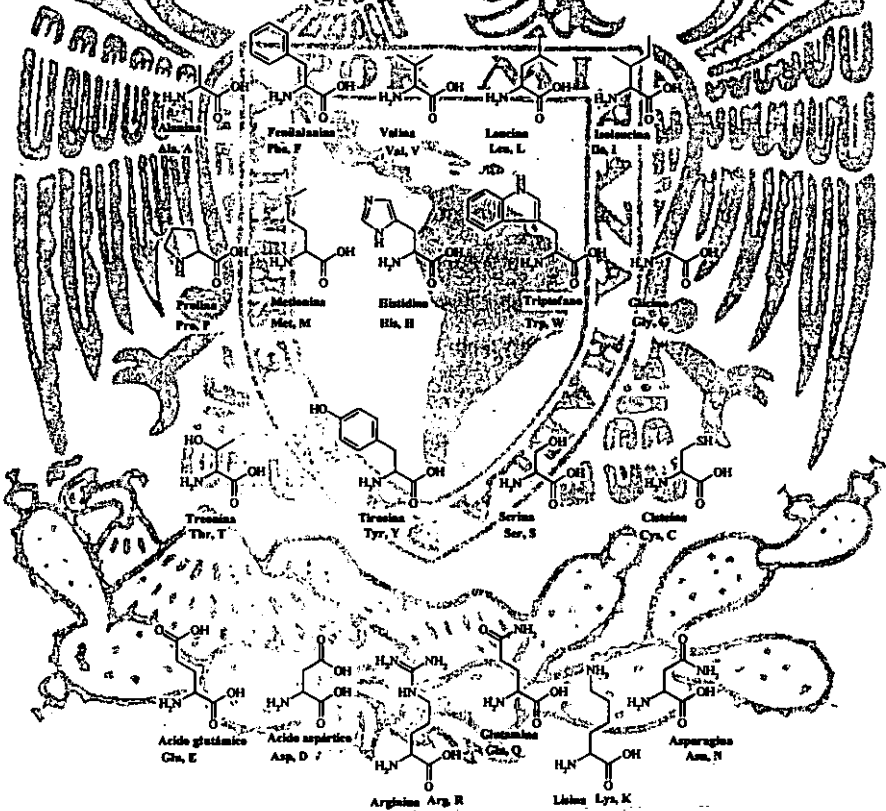


Figura 1.1.- Estructura química de los 20 aminoácidos más comunes presentes en la mayoría de las proteínas. Aunque existen algunos otros aminoácidos particulares de ciertos seres vivos, estos 20 aminoácidos son comunes a todos; estereoquímicamente casi siempre pertenecen a la serie L. Su estructura química general es:  $H_2N-HCR-COOH$ , donde R es la cadena lateral que caracteriza a cada uno de ellos. Existen varias formas de agrupar los aminoácidos, todas ellas dependen de las características fisicoquímicas de la cadena lateral (R); la más general los agrupa en base en su polaridad y afinidad por el agua. De esta manera, tenemos los hidrófobos, cuya cadena lateral está compuesta por metilos y metilenos; los polares neutros como la Thr, que posee un OH en su cadena lateral y los polares cargados, los cuales por las condiciones del pH fisiológico se encuentran protonados ( $NH_3^+$ , Gln, Asn, etc.) o desprotonados ( $COO^-$ , Glu, Asp, etc.)

Los aminoácidos poseen la estructura general  $\text{H}_3\text{N}-\text{CHR}-\text{COO}^-$ , donde R, la cadena lateral, determina la identidad única y, por lo tanto, su estructura y su reactividad. La formación del polipéptido o proteína involucra una reacción de condensación de los aminoácidos donde el nitrógeno de la amina de un residuo queda unido al carbonilo del ácido de otro residuo para formar un enlace de amida, también llamado enlace peptídico.

El orden lineal en el que los aminoácidos se encuentran unidos en la proteína es conocida como la **estructura primaria** de la proteína, llamada comúnmente como **secuencia de aminoácidos**. Segmentos discretos de la cadena polipeptídica pueden plegarse en segmentos regulares con motivos repetidos, conocido como **estructura secundaria** (figura 1.2). Es la identidad y la secuencia de los aminoácidos dentro de la cadena polipeptídica la que determina la estructura secundaria. Sin embargo, hasta ahora aún no están suficientemente comprendidos los diversos factores que influyen en la formación de ésta, por lo que las predicciones a partir de la estructura primaria aún no son exactas. Las dos principales estructuras regulares dentro de las proteínas son las  **$\alpha$  hélices** y las **hebras plegadas  $\beta$** . Otro de los elementos regulares de gran importancia son los pequeños segmentos que permiten a la cadena cambiar de dirección conocidos como **vueltas en U**. Regiones de la proteína que no contienen una estructura regular se conocen como **regiones no repetitivas**.

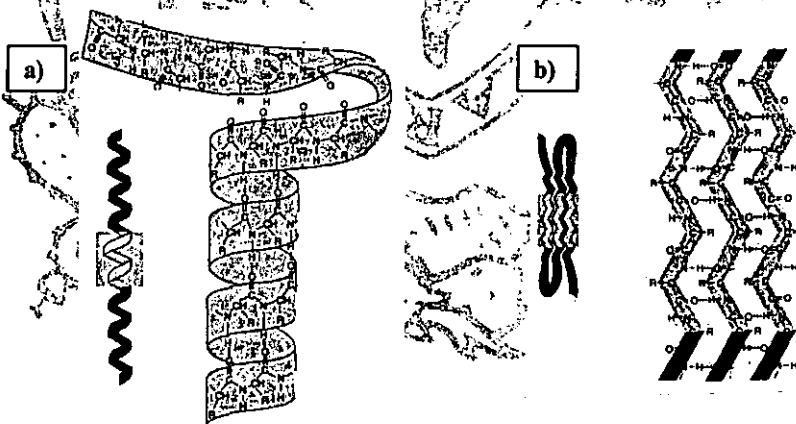


Figura 1.2.- Elementos regulares más comunes de la estructura secundaria: a) las  $\alpha$  hélices, b) las hebras plegadas  $\beta$  y c) las vueltas en U. Las proteínas son macromoléculas altamente estructuradas con regiones de conformación local particular. Estas conformaciones locales dependen de muchos factores, entre los que se encuentran: la polaridad de las cadenas laterales y los puentes de hidrógeno. En la figura se muestra una  $\alpha$  hélice (una de varias conformaciones existentes de este tipo; los diferentes tipos de hélices dependen del número de aminoácidos por cada vuelta y del diámetro de la misma); también, se muestra una hoja plegada  $\beta$  antiparalela (en el caso de las hojas sólo hay 2 tipos: las paralelas y antiparalelas); las vueltas en U también adquieren varias conformaciones dependiendo de la dirección y de la disposición de los aminoácidos en esta estructura local.

La forma en la que los elementos de la estructura secundaria se pliegan entre sí para formar una estructura globular compacta es conocida como la *estructura terciaria*. Fuerzas no covalentes como puentes de hidrógeno, interacciones electrostáticas e interacciones hidrofóbicas, juegan un papel relevante en la determinación de la estructura terciaria. En algunas proteínas, los enlaces covalentes son formados por la oxidación de dos grupos sulfhidrilo, de cisteínas para dar lugar a un enlace disulfuro, el cual proporciona una gran estabilidad adicional a la estructura. Es la estructura terciaria la que le confiere a la proteína su actividad, su forma y dimensiones, además de que es el medio por el cual dos aminoácidos, distantes dentro de la secuencia, puedan encontrarse próximos en el espacio. En muchas ocasiones, la forma activa de algunas proteínas está constituida por un complejo de dos o más cadenas polipeptídicas que se mantienen unidas por medio de enlaces no covalentes o, en algunos casos, mediante enlaces disulfuro. Este arreglo funcional de subunidades polipeptídicas define la *estructura cuaternaria* de las proteínas.

Para entender la función de las proteínas a nivel molecular es importante conocer su estructura tridimensional. La diversidad en la estructura de las proteínas es resultado de la flexibilidad de rotación alrededor de enlaces sencillos entre los átomos. El par de electrones libre en el átomo de nitrógeno de cada residuo está deslocalizado entre éste y el átomo de carbono, confiriendo un carácter de doble enlace a la unión entre el nitrógeno y el carbono, por lo que cada unidad peptídica es plana ( $\omega = 180^\circ$ ), poseyendo dos grados de libertad rotacional, especificados por los ángulos de torsión  $\phi$  y  $\psi$  (figura 1.3). El número de ángulos de torsión asociados con las cadenas laterales, R, varía en cada uno de los aminoácidos. La conformación permitida de la proteína será aquella que evite el contacto entre átomos no enlazados.

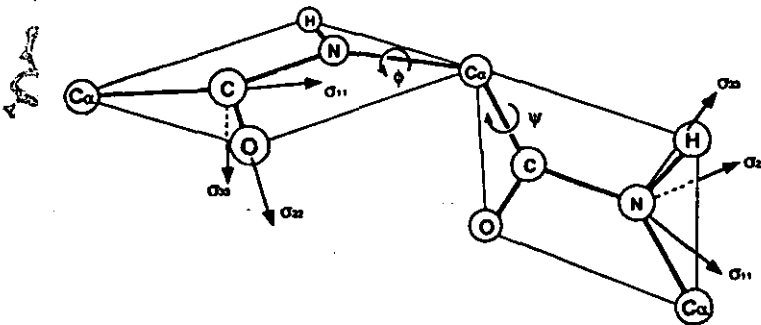


Figura 1.3.- Planos de las unidades peptídicas y nomenclatura de los ángulos de torsión interplanar. El enlace peptídico (C<sup>carbonilo</sup>-N) posee un carácter parcial de doble enlace, debido a la deslocalización del par electrónico del nitrógeno; la deslocalización del par electrónico restringe el giro del enlace C<sup>carbonilo</sup>-N, quedando restringido el libre giro a los enlaces N-C <sup>$\alpha$</sup>  (ángulo  $\phi$ ) y C <sup>$\alpha$</sup> -C<sup>carbonilo</sup> (ángulo  $\psi$ ).

Para cada proteína, el número de conformaciones que puede adoptar es astronómico, debido a la gran cantidad de enlaces y a la rotación que presentan algunos de ellos. Sin embargo, cada proteína se pliega en una estructura activa única. Una de las principales fuerzas que determina y mantiene la estructura de la proteína es el aislamiento del interior de la proteína al contacto con el disolvente. La mayor contribución a esta fuerza se encuentra en el efecto hidrofóbico de las cadenas laterales de los aminoácidos apolares que tienden a concentrarse en el interior de la proteína. Otra contribución muy importante para el mantenimiento de la estructura terciaria de las proteínas es el efecto de puentes de hidrógeno que se presenta entre el protón de amida y el oxígeno de carbonilo en el esqueleto de la cadena y en las cadenas laterales (puente salino), con lo cual se elimina el agua de esos sitios y contribuye al efecto solvofóbico. Este efecto es vital en el mantenimiento de la estructura, ya que los puentes de hidrógeno dan lugar a la formación de las dos principales estructuras regulares de la proteína, las  $\alpha$  hélices que son el resultado de girar las cadenas peptídicas ( $\phi = -57^\circ$ ), en las que el carbonilo y el hidrógeno de la amida siempre quedan orientados de manera paralela al eje de torsión de la hélice, posibilitando el puente de hidrógeno; y las hojas plegadas  $\beta$ , las cuales se forman al extender la cadena e interactuar con otros residuos igualmente extendidos.

Basados en las dos estructuras regulares principales y utilizando el hecho de que los aminoácidos pueden ser clasificados de manera general como hidrofílicos (polares) e hidrofóbicos (apolares), podemos entender la estructura básica de la mayoría de las proteínas. El plegamiento de las proteínas busca maximizar el número de puentes de hidrógeno y minimizar el número de cadenas laterales de aminoácidos apolares expuestos al disolvente (agua).

### Proteína SV40 [1-108]

El SV40 es un virus que infecta a diversas especies de primates. El virus atrajo la atención mundial cuando en 1960, algunos investigadores encontraron que era uno de los muchos virus que contaminaban las células del riñón de monos macacos rhesus, que fueron utilizados en la producción inicial de la vacuna contra la poliomielitis de Sabin y Salk.

Preocupados por los posibles efectos adversos del SV40, los investigadores pronto encontraron que producía anomalías en células animales y cáncer en hámsters. Basados en

este y otros estudios, en 1961 el gobierno federal de los E.E.U.U. estableció que la vacuna contra la poliomiélitis estuviera libre del virus, aunque al parecer algo de la vacuna contaminada con SV40 fue administrada en algunos lugares de los E.E.U.U. hasta 1963.

En las dos décadas siguientes, se realizaron intensas investigaciones para determinar si el SV40 podía causar cáncer en el ser humano. Finalmente, los estudios más generales no mostraron un incremento de cáncer o muertes relacionadas con cáncer, en personas a las que se les habían administrado la vacuna contra la poliomiélitis contaminada con SV40. Sin embargo, algunos estudios sugerían un incremento en el riesgo de padecer ciertos tipos raros de cáncer en la gente que había sido expuesta a las primeras vacunas.

El problema de las vacunas contaminadas contra la poliomiélitis resurgió recientemente, cuando algunos investigadores al utilizar técnicas de biología molecular, especialmente la reacción en cadena de la polimerasa (PCR), encontraron trazas del virus en ciertos tumores raros en seres humanos como en el mesotelioma (un cáncer de ciertos tejidos pulmonares), osteosarcoma (cáncer de hueso), ependioma (cáncer en el cerebro). Sin embargo, la relación entre estos tipos de cáncer y el SV40 era incierta, ya que algunos investigadores no encontraron trazas del SV40 en personas con estas afecciones.

Actualmente el problema no está solucionado. Existe un enorme y creciente banco de datos que no muestran una relación directa entre el incremento de estos tipos de cáncer y la exposición al SV40. En enero de 1997, investigadores del Instituto Nacional de Cáncer (NCI) reportaron que después de más de 30 años de seguimiento, un grupo de norteamericanos expuestos a las vacunas iniciales contra la polio contaminadas con SV40, no han mostrado un incremento en la frecuencia de aparición de estos tipos de cáncer.

El antígeno T del virus siamés (SV40Tag) es una proteína reguladora multifuncional que estimula la transcripción de los genes y forma complejos con las proteínas reguladoras de los ciclos celulares como la Trp53 y Rb1 que están involucradas en el cáncer humano de pecho.

La proteína del antígeno T del virus siamés (SV40Tag) está compuesto por 708 aminoácidos. En el presente trabajo sólo se utilizaron los aminoácidos 1-108 (SV40[1-108]) de la región amino terminal, debido a que esta región ha mostrado ser capaz de inducir la



transformación celular por sí sola. Además, ha mostrado que posee la capacidad de inhibir la actividad de las proteínas Trp53 y Rb1, que son las responsables del control celular; esta proteína promueve la transformación celular de células normales a cancerosas.

Por estas y otras características adicionales el sistema proteico del virus SV40 es ideal para:

- ✧ examinar la fidelidad y la bioquímica de la replicación del ADN en los humanos
- ✧ determinar el riesgo en seres humanos de promover un proceso carcinogénico
- ✧ identificar el modo de acción y el mecanismo que llevan a la transformación celular *in vitro* y la formación tumoral *in vivo*
- ✧ estudiar la mutagénesis en células de mamíferos
- ✧ regulación de la expresión genética en células eucariontes
- ✧ investigación de la apoptosis celular
- ✧ estudio de los efectos citotóxicos de los medicamentos
- ✧ comparar la replicación de las células normales y las células cancerígenas
- ✧ el sistema del virus SV40 es una herramienta valiosa para el estudio funcional de la maquinaria proteica durante la replicación del ADN celular en eucariontes; debe tenerse en cuenta el papel multifuncional del antígeno T mayor.
- ✧ iniciador de las funciones proteicas que promueven la replicación del ADN por mutagénesis
- ✧ funciona como helicasa para desenrollar el ADN, previo al funcionamiento de la maquinaria de polimerización de ADN

La estructura primaria del fragmento 1-108 de la proteína SV40 es:

1		5		10		15								
Met	Asp	Lys	Val	Leu	Asn	Arg	Glu	Glu	Ser	Leu	Gln	Leu	Met	Asp
				20		25								30
Leu	Leu	Gly	Leu	Glu	Arg	Ser	Ala	Trp	Gly	Asp	Ile	Pro	Leu	Met
				35		40								45
Arg	Lys	Ala	Tyr	Leu	Lys	Lys	Cys	Lys	Glu	Phe	His	Pro	Asp	Lys
				50		55								60
Gly	Gly	Asp	Glu	Glu	Lys	Met	Lys	Lys	Met	Asn	Thr	Leu	Tyr	Lys
				65		70								75
Lys	Met	Glu	Asp	Gly	Val	Lys	Tyr	Ala	His	Gln	Pro	Asn	Phe	Gly
				80		85								90
Gly	Phe	Trp	Asp	Ala	Thr	Glu	Ile	Pro	Thr	Tyr	Gly	Thr	Asp	Glu
				95		100								105
Trp	Glu	Gln	Trp	Trp	Asn	Ala	Phe	Asn	Glu	Glu	Asn	Leu	Phe	Cys
				108										
Ser	Glu	Glu												

El fragmento de la proteína SV40[1-108] tiene un peso molecular de 12384 g/mol (12.4

kDa), con una fórmula mínima: C<sub>573</sub>H<sub>874</sub>O<sub>172</sub>N<sub>147</sub>S<sub>8</sub>.



## Resonancia Magnética Nuclear, generalidades<sup>5-12</sup>

La resonancia magnética nuclear (RMN) es la técnica espectroscópica más eficaz en la determinación de la estructura, conformación y propiedades dinámicas de moléculas orgánicas e inorgánicas en disolución, sólidos e incluso en gases, aunque esto último no es muy usual. Este método mide las interacciones entre la materia, dentro de un campo magnético externo, y la radiofrecuencia. A finales de la década de los 60 e inicio de los 70, comenzó a aplicarse la RMN en la determinación de la estructura y conformación de pequeñas moléculas de gran importancia biológica. Inicialmente, estos estudios se hicieron en espectrómetros de 60 y hasta 100 MHz, donde la información conformacional era derivada del desplazamiento químico, constantes de acoplamiento y medidas de efecto nuclear Overhauser, en espectros de una dimensión.

Cuando se incrementó el campo magnético de los espectrómetros (200 - 500 MHz) a finales de los 70 y comienzos de los 80, se incrementó la complejidad de las moléculas que se han estudiado con esta técnica. Adicionalmente, se inició el análisis de macromoléculas biológicas con un mejor detalle, aunque en esos momentos no se obtenía con gran definición la estructura tridimensional (3D) de las macromoléculas, la RMN se utilizó para obtener información cualitativa acerca de la estructura de pequeños biopolímeros.

El número de aplicaciones de la RMN a las ciencias biológicas se incrementó sustancialmente desde mediados de los 80, debido a múltiples mejoras instrumentales y a los nuevos métodos que se han desarrollado desde entonces. Particularmente, el desarrollo de la metodología para la asignación y determinación de las estructuras de péptidos y proteínas mediante espectros de RMN 2D. Inicialmente, la nueva metodología suministró la capacidad de determinar la estructura tridimensional completa de pequeñas proteínas (< 10 kDa), basados en la determinación de una gran cantidad de restricciones de distancias interprotónicas a partir de espectros NOE 2D (NOESY). Con el reciente desarrollo de los espectros 3D y 4D y los espectrómetros de campo alto (750 y 800 MHz) el límite de peso molecular se ha incrementado a más de 30 kDa (un poco más de 350 aminoácidos).

En general, muchos de los avances metodológicos recientes en RMN, se han basado en el desarrollo de una gran cantidad de secuencias de pulsos para la obtención de diversos tipos de información de una muestra, así como de los diferentes núcleos que se pueden utilizar ahora.

Marcando uniformemente las proteínas con  $^{13}\text{C}$ ,  $^{15}\text{N}$  y/o  $^2\text{H}$ , muchos experimentos nuevos pueden aplicarse para obtener mejor información de los espectros de RMN. Esto incluye el uso de los amplios intervalos de las constantes de acoplamiento heteronucleares, para dispersar las señales en espectros multidimensionales, reduciendo con ello el traslape de las mismas; de las constantes de acoplamiento heteronuclear se extrae información sobre restricciones de ángulos diedros, utilizados en el refinamiento de la estructura, además del uso de espectros editados selectivamente por heteronúcleos. Utilizando las secuencias de pulsos apropiadas, es posible crear subespectros en los que solamente protones unidos a  $^{13}\text{C}$  o  $^{15}\text{N}$  sean selectivamente incluidos o excluidos, simplificando con ello considerablemente los espectros y facilitando el trabajo de asignación de las señales. Además, las técnicas de edición isotópica han encontrado una amplia aplicación en años recientes.

Muchas de las secuencias de pulsos desarrolladas recientemente, se han beneficiado por la incorporación de la tecnología de pulsos de gradiente de campo. Estos gradientes se utilizan en la selección de las vías de transferencia de coherencia (que determina la naturaleza multidimensional de los experimentos de RMN) y para la supresión de señales no deseadas del disolvente (generalmente agua) se han desarrollado significativas mejoras metodológicas.

Los avances instrumentales que han acompañado a estos nuevos desarrollos metodológicos, son el incremento en el campo magnético, mejoras en la precisión y flexibilidad en la electrónica que genera los pulsos de radiofrecuencia, el desarrollo del *hardware* para los pulsos de gradiente de campo y mejoras en la tecnología de la sonda.

La estrategia general (figura 2.1) para la determinación de la estructura tridimensional de una proteína (estructura terciaria de la proteína), utilizando espectroscopia de RMN involucra varios pasos: (1) asignación de las señales de RMN ( $^1\text{H}$ ,  $^{13}\text{C}$  y  $^{15}\text{N}$ ) a átomos o grupos de átomos de la proteína. (2) extracción de las restricciones experimentales de los datos de RMN, las cuales proveen de información fundamental acerca de la posición espacial de estos átomos dentro de la proteína y (3) el uso de las restricciones para que un programa de cómputo obtenga conjuntos de estructuras de la proteína, cada una de las cuales satisface las restricciones experimentales.

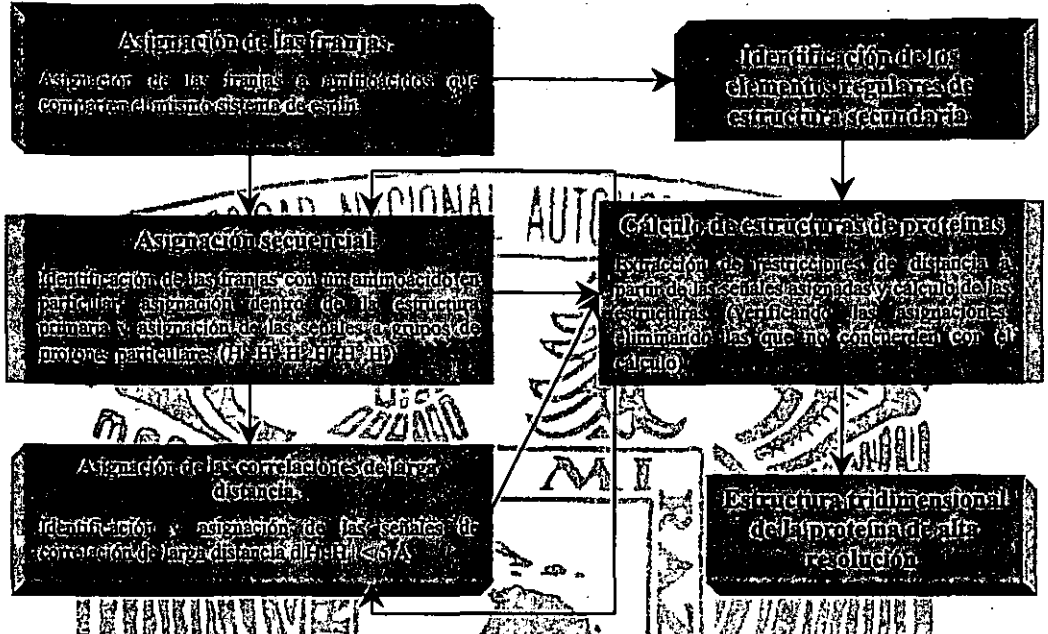


Figura 2.1. Esquema general para la determinación de la estructura tridimensional de las proteínas utilizando RMN. La parte fundamental en la determinación de la estructura terciaria de las proteínas mediante RMN es la asignación de las señales de los espectros (2D, 3D o 4D) a una correlación entre núcleos. Por ello la identificación precisa de las señales se encuentra en las 3 primeras fases del proceso. Durante la asignación de las franjas, provisionalmente, se asigna cada franja a un aminoácido (o grupo de ellos) que poseen señales similares al sistema de espines observado (Lys, Val, Phe, Gly, etc.). En la asignación secuencial se identifica el aminoácido que genera la franja, así como su posición dentro de la estructura primaria (a diferencia del paso anterior, ahora se asigna cada franja a un aminoácido en particular, p. Ej. Ala-15, Thr-20, etc.). Con la asignación de las correlaciones de larga distancia obtenemos las restricciones que proporcionan la estructura terciaria de las proteínas.

Hay dos aproximaciones generales a través de las cuales se efectúa la asignación de las señales de resonancia (paso 1); y la elección entre ellas estará determinada, esencialmente, por la masa molecular de la proteína. La primera aproximación involucra exclusivamente el uso de experimentos de RMN bidimensionales (2D) de correlación homonuclear  $^1\text{H}$ - $^1\text{H}$  (como el *COSY*, *TOCSY* y *NOESY*). Actualmente, esta aproximación aún es ampliamente utilizada para la determinación de proteínas cuyos pesos moleculares estén por debajo de los 5 KDa. Este método se ha utilizado con un éxito muy limitado para proteínas cuyos pesos moleculares superan los 10 kDa; como una ruta más eficiente se ha extendido la estrategia de asignar experimentos heteronucleares de tres dimensiones, aprovechando el marcaje isotópico de  $^{13}\text{C}$  y/o  $^{15}\text{N}$ . La presencia de ese marcaje se utiliza simplemente para dispersar los espectros 2D homonucleares de acuerdo al desplazamiento químico del núcleo de  $^{13}\text{C}/^{15}\text{N}$  al que se encuentran unidos, incrementando de manera significativa la dispersión espectral. Los experimentos de triple resonancia, que correlacionan los desplazamientos químicos de  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$  en subconjuntos de

señales relacionados a través de correlación escalar, proveen una gran alternativa para lograr la asignación de señales a una secuencia de aminoácidos específica.

El desarrollo de la metodología de RMN-2D, a principios de los 80, fue el punto de partida para el estudio estructural detallado de las proteínas en disolución. Si la información acerca de las interacciones entre espines se extrae de experimentos unidimensionales (1D), se deben aplicar selectivamente pulsos a una señal de resonancia particular y medir los efectos por los cambios que sufre otra señal en el espectro. Este método es sencillo cuando no hay un gran traslape de las señales, pero rápidamente se vuelve poco práctico aun con moléculas pequeñas. En contraste, si se extienden dichas mediciones a una segunda dimensión, esta evitará muchos problemas en la sobreposición de las señales y en un solo experimento se obtendrán todas las correlaciones posibles. Los experimentos bidimensionales están compuestos por los siguientes elementos (figura 2.2): periodo de preparación; periodo de evolución ( $t_1$ ), durante el cual los núcleos al precesar en el plano  $xy$  son "marcados" con la información de su propio desplazamiento químico; periodo de mezclado, durante el cual los espines adquieren información de otros núcleos con los cuales se está acoplado y el periodo de detección, en el que se registra el decaimiento libre de la inducción (FID,  $t_2$ ). Estos elementos básicos al ser combinados crean experimentos más complejos, que son la base de los experimentos en 3 y 4 dimensiones. En la obtención de un experimento 2D (figura 2.3), este esquema se repite en múltiples ocasiones, incrementando periódicamente el tiempo de evolución ( $t_1$ ). Una vez concluida la adquisición de datos, se lleva a cabo una primera transformada de Fourier (ésta se efectúa sobre  $t_2$ ); de ella se obtienen señales cuya amplitud, en los diversos incrementos efectuados en  $t_1$ , estará *modulada* por la frecuencia del núcleo con el que correlacionó durante el periodo de mezclado en la frecuencia propia del núcleo que origina dicha señal. Esta modulación se traduce en el aumento y la disminución de la intensidad de las señales, de una forma muy similar al libre decaimiento de la inducción (FID, en la segunda dimensión  $t_1$ ) y si se realiza una segunda transformada de Fourier, obtendremos un espectro cuadrado con dos dimensiones de frecuencia ( $\omega_1$ ,  $\omega_2$ ). En los experimentos homonucleares 2D, las señales de la diagonal son iguales a los experimentos 1D y los elementos de correlación entre los espines son las señales fuera de la diagonal.

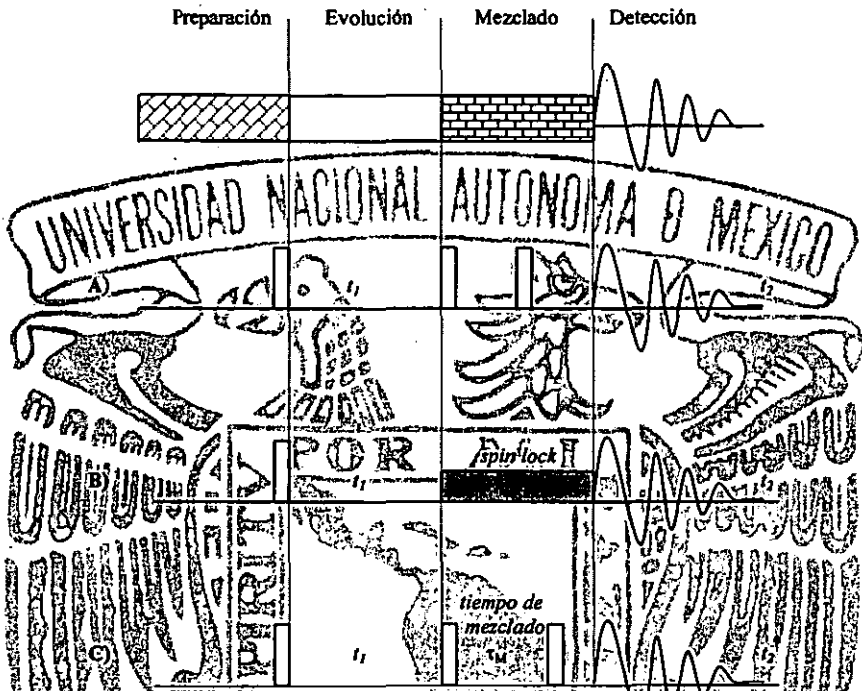


Figura 2.2. Elementos básicos de los experimentos de RMN-2D. A) secuencia de pulsos del experimento DQFCOSY, B) del experimento TOCSY, y C) del experimento NOESY. Se observa que la naturaleza del periodo de mezclado determina el tipo de correlaciones observadas en el experimento. Los rectángulos vacíos son pulsos de  $90^\circ$  y la región sombreada en el experimento B (TOCSY) son los pulsos de mezclado isotrópico (spin-lock), que transfiere la magnetización entre los núcleos de un sistema de espines;  $t_1$  es el tiempo de mezclado en el que se crea el NOE.

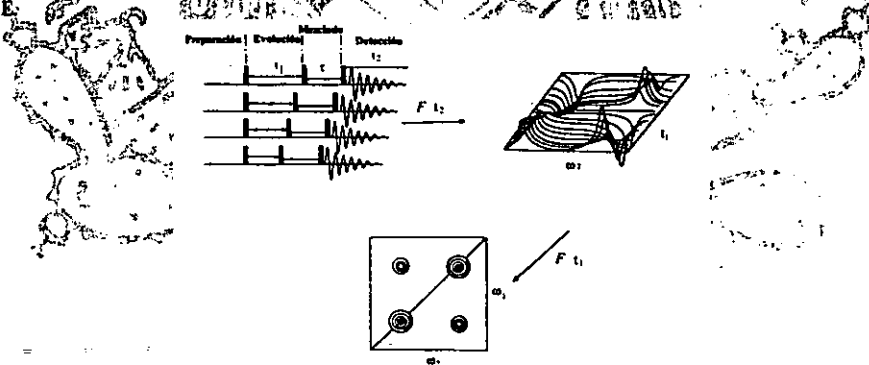


Figura 2.3.- Esquema para obtener experimentos 2D. Con una secuencia de pulsos se obtienen  $n$  experimentos en los que se incrementa periódicamente el tiempo de evolución ( $t_1$ ). La primera transformada de Fourier (que opera sobre el tiempo de adquisición,  $t_2$ ) genera  $n$  espectros 1D cuya intensidad está modulada por el acoplamiento con otros, durante  $t_1$ . Al tomar una frecuencia a lo largo de  $t_1$  se observa una variación de la intensidad de las señales, este comportamiento armónico es idéntico al FID que observamos durante el periodo de detección; al realizar una nueva transformada de Fourier sobre  $t_1$ , obtenemos el espectro 2D; las señales de la diagonal son similares a las que se obtendrían en un espectro 1D y las señales fuera de la diagonal nos indican la correlación entre dos núcleos; el significado de las correlaciones depende de la secuencia de pulsos utilizada en cada experimento.

Los experimentos se distinguen por la naturaleza de la correlación obtenida durante el periodo de mezclado. El acoplamiento escalar entre protones separados hasta por tres enlaces de distancia es utilizado en el experimento de espectroscopia de **CORrelación (COSY)** para generar las señales de correlación (señales fuera-de-la-diagonal). El experimento de **CORrelación TOTAL (TOCSY)**, también conocido como **HOHAHA**, experimento **HOmonuclear HArtmann-HAHn** es de un valor muy especial, ya que potencialmente pueden observarse todas las correlaciones entre los protones que conforman el **sistema de espines** de una molécula; esto es, el grupo de protones que pertenecen a un aminoácido ( $H^N, H^{\alpha}, H^{\beta}, H^{\gamma}, H^{\delta}$ ) produce un patrón de acoplamiento característico por su desplazamiento químico, el número de señales y la multiplicidad de éstas (figura 2.4). Los espectros **NOESY** (donde se observa el Efecto Nuclear Overhauser) correlaciona los protones cercanos en el espacio ( $<5 \text{ \AA}$ ). Como puede verse en la figura 2.2 la diferencia entre estos dos experimentos es la naturaleza del periodo de mezclado.

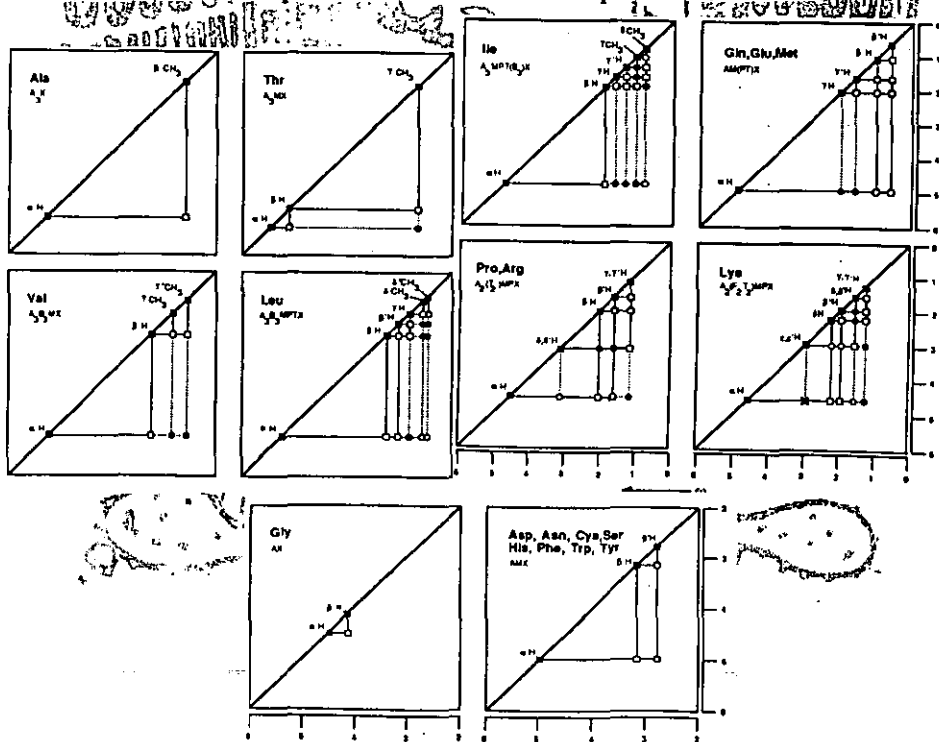


Figura 2.4.- Sistema de espines de los 20 aminoácidos más abundantes en la naturaleza. En este esquema observamos la similitud entre los sistemas de espín de varios aminoácidos. Gln, Glu y Met presentan 6 señales en el espectro ( $H^N$  [no se muestra en estos esquemas],  $H^{\alpha}, H^{\beta}, H^{\gamma}, H^{\delta}, H^{\epsilon}$ ), ocasionalmente los  $H^{\beta}$  y  $H^{\gamma}$  aparecen como una sola señal (son equivalentes). Algo similar sucede con los aminoácidos Asp, Asn, Cys, Ser, His, Phe, Trp y Tyr donde se observan 4 señales y eventualmente los protones  $H^{\beta}$  son equivalentes. Los aminoácidos con sistemas de espín único son Ala, Thr, Val, Gly e Ile.



En la primera etapa de la asignación de las señales de RMN, se asigna el sistema de espines a un grupo aminoácidos cuyo patrón de señales es similar en los experimentos 2D de acoplamiento escalar (*TOCSY*, *COSY*).

Una vez que se han identificado todos los sistemas de espines (o la mayor parte de ellos), es necesario asignarlos a residuos específicos dentro de la secuencia (estructura primaria). Ello se realiza utilizando conectividades a través del espacio, obtenidos de espectros de efecto nuclear Overhauser que correlaciona pares de protones que se encuentran a menos de 5 Å de distancia.

El procedimiento de asignación secuencial depende de la observación de conectividades entre los protones  $H^N$ ,  $H^\alpha$  y  $H^\beta$  de residuos adyacentes en la secuencia. Se ha demostrado que en todas las proteínas, independientemente de la región de estructura secundaria, al menos uno (a veces más) de estos protones se encuentra a una distancia aproximada de 3.5 Å, y consecuentemente darán señal de correlación de mediana intensidad en los espectros NOE. Las distancias más útiles son las  $d_{N-N}$  (esto es, la distancia del protón  $H^N$  del residuo  $i$  al protón  $H^N$  del residuo  $i+1$ ),  $d_{N-\alpha}$  (del protón  $H^N$   $i$  al protón  $H^\alpha$   $i+1$ ),  $d_{N-\beta}$  (del protón  $H^N$   $i$  al protón  $H^\beta$   $i+1$ ) y  $d_{N-N+2}$  (del protón  $H^N$   $i$  al protón  $H^N$   $i+2$ ).

Pueden ocurrir interrupciones en la asignación secuencial, como resultado de la sobreposición de las señales de dos protones de amidas con desplazamientos químicos idénticos o por causa de la propia estructura, generalmente debido a la presencia de prolinas que carecen de  $H^N$ . Una vez que una pequeña secuencia de residuos de aminoácidos (3 ó 4) ha sido identificada por sus sistemas de espines, se coloca dentro de la estructura primaria del péptido. Estadísticamente, si de un grupo de 3 aminoácidos (de los 20 más comunes) se hacen todas las combinaciones, se obtienen 6840 posibles arreglos; en el caso de que sea un conjunto de 4 aminoácidos se obtienen 116280 posibles arreglos. Por esta razón, cuando se ha identificado una secuencia de 3 ó 4 aminoácidos se puede asignar a una secuencia específica con una certeza del 99.98 %; una vez localizado el fragmento se puede ampliar fácilmente en ambas direcciones. Este proceso se repite hasta que todas las asignaciones se hayan realizado.

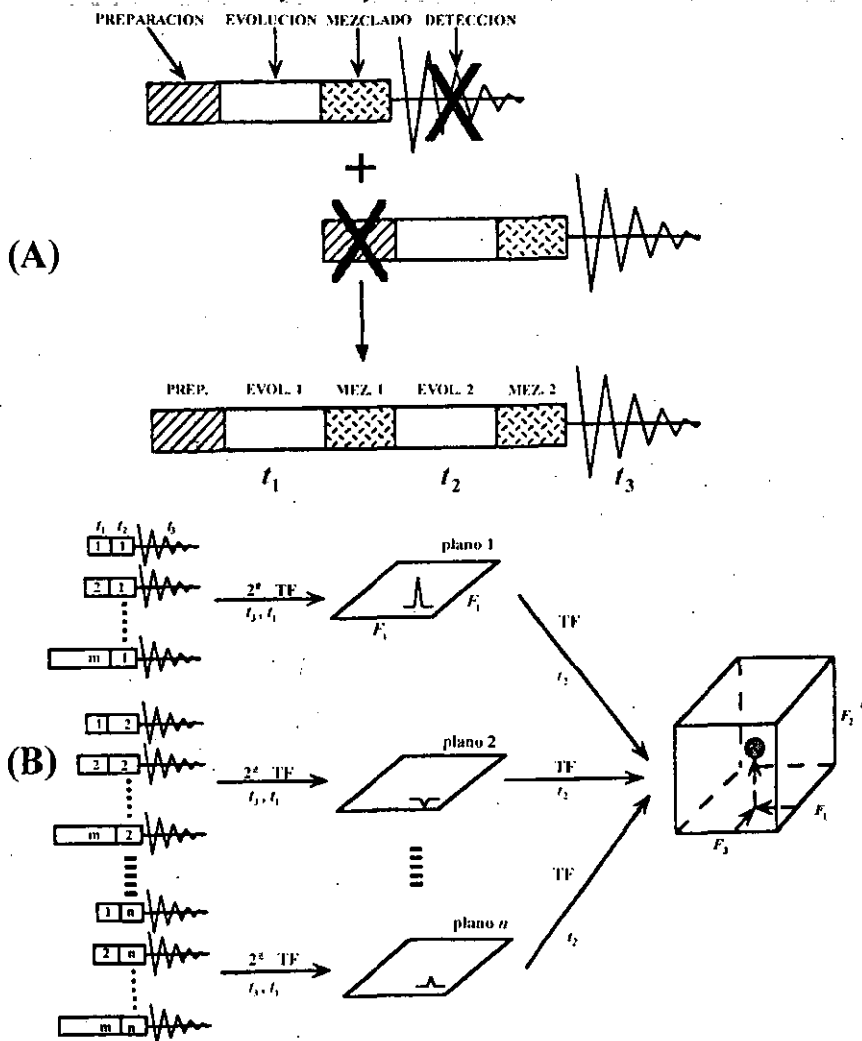
La asignación completa e inequívoca de las señales utilizando solamente experimentos homonucleares, generalmente es inadecuado para proteínas con peso molecular mayor a 5 kDa,

debido al bajo tiempo de correlación que incrementa la eficiencia de la relajación transversa, ensancha las señales y disminuye la transferencia de coherencia a través del acoplamiento escalar homonuclear; adicionalmente, al incrementar el número de protones disminuye aún más la resolución del espectro. Los experimentos 3D-homonucleares proveen algo de ayuda, aunque la ganancia de adicionar otra dimensión se pierde por la ineficacia en la transferencia de coherencia debido a la débil correlación escalar H-H.

A finales de los '80 se dio un nuevo avance al incrementar las dimensiones de los espectros de RMN, pasando de 2 a 3 ó 4 dimensiones. Un experimento de 3 dimensiones puede considerarse como la combinación de varios experimentos bidimensionales, cuando el periodo de detección ( $t_3$ ) de un experimento es substituido por otro experimento 2D al que se le ha eliminado el periodo de preparación. De esta manera, el experimento 3D contiene dos periodos de evolución ( $t_1$  y  $t_2$ ), dos periodos de mezclado y un periodo de detección ( $t_3$ ). Los dos tiempos de evolución se incrementan de manera independiente y al realizar tres transformadas de Fourier, de forma similar a como se hizo con los experimentos 2D, se obtiene un espectro con tres ejes de frecuencia ortogonalmente arreglados como un cubo (figura 2.5).

Los progresos en la tecnología de ADN recombinante, permiten obtener proteínas marcadas isotópicamente con relativa facilidad. Generalmente se utilizan proteínas marcadas uniformemente con  $^{13}\text{C}$  y  $^{15}\text{N}$ , aunque se obtienen proteínas con marcas selectivas (es decir, sólo con  $^{13}\text{C}$  o  $^{15}\text{N}$ , marcadas sólo las leucinas o algún otro aminoácido, etc.). Las propiedades magnéticas de estos núcleos (ambos poseen número cuántico de espín  $|I| = 1/2$ ) les permiten ser utilizados en los experimentos de RMN, al explotar sus constantes de acoplamiento escalar a uno y/o dos enlaces de distancia que son características. Las constantes de acoplamiento escalar mayores son la gran ventaja de los espectros heteronucleares sobre los homonucleares, ya que es mucho más eficiente la transferencia de polarización en el primer caso. El experimento de Correlación Heteronuclear con transferencia de polarización Cuántica Simple (HSQC, figura 2.6) nos proporciona los desplazamientos químicos de los  $^{15}\text{N}$  ó  $^{13}\text{C}$  (del esqueleto de la proteína y de las cadenas laterales, en una proteína uniformemente marcada) y de los protones ( $^1\text{H}$ ) que están directamente unidos, con una gran sensibilidad. La secuencia de pulsos involucra la transferencia inicial de magnetización del  $^1\text{H}$  al  $^{15}\text{N}$  a través del acoplamiento a un enlace de distancia, un periodo de evolución ( $t_1$ ) donde la magnetización es marcada con el desplazamiento químico del  $^{15}\text{N}$  y el regreso de la magnetización del  $^{15}\text{N}$  al  $^1\text{H}$  para finalmente detectar la señal

en protón ( $^1\text{H}$ ) durante  $t_2$ . Dos transformaciones de Fourier nos proporcionan el espectro 2D, en los que se obtiene directamente los desplazamientos químicos de los protones y el desplazamiento químico del nitrógeno al que se encuentra unido.



**Figura 2.5.-** (A) Construcción de un experimento de RMN-TF-3D a partir de dos experimentos 2D. El periodo de detección de la primera secuencia de pulsos es eliminado junto con el periodo de preparación del segundo experimento. Los tiempos de evolución se transforman ahora en  $t_1$  y  $t_2$  y el periodo de detección es  $t_3$ . (B) Procesando el experimento 3D. Los tiempos de evolución  $t_1$  y  $t_2$  se incrementan  $m$  y  $n$  veces, respectivamente; primero, se obtiene una serie de  $m$  espectros en los que se incrementó  $t_1$ , mientras  $t_2$  se mantiene constante. A continuación, se incrementa  $t_2$  y con este nuevo valor fijo, se repite la serie de experimentos incrementado  $t_1$ . Este proceso se repite  $n$  veces. Al efectuar las dos primeras transformadas de Fourier (sobre  $t_3$  y  $t_1$ ), se obtiene un conjunto de  $n$  espectros en 2D. La intensidad de las señales en este espectro estarán moduladas por el incremento en  $t_2$ , por lo que al aplicar una tercera transformada de Fourier creará el tercer eje en el dominio de las frecuencias.



Figura 2.6. Secuencia de pulsos utilizado en el experimento 2D-HSQC en el que la señal de correlación proporciona el desplazamiento químico de los átomos de hidrógeno y de los átomos de nitrógeno ( $^1\text{H}$ - $^{15}\text{N}$ ) que se encuentran directamente unidos en una proteína marcada con  $^{15}\text{N}$ . Esta secuencia de pulsos es exactamente la misma si se utiliza  $^{13}\text{C}$  y su interpretación es equivalente, sólo cambia la frecuencia a la que se aplican los pulsos de excitación del heteronúcleo ( $^{13}\text{C}$  ó  $^{15}\text{N}$ ).

Es fácil pasar de experimentos 2D a experimentos 3D, utilizando la secuencia de pulsos del experimento 2D-TOCSY y la secuencia 2D-HSQC, en el experimento resultante el sistema de espines de los aminoácidos (que corresponde a la porción TOCSY del experimento) está dispersado en el desplazamiento químico del nitrógeno (esta es la parte del experimento que corresponde al HSQC), disminuyendo el traslape de las señales. Este experimento, llamado 3D-TOCSY-HSQC, es interpretado a través de pequeñas franjas bidimensionales ( $^1\text{H}$ - $^1\text{H}$ ), que son tomadas del desplazamiento químico del  $^{15}\text{N}$ ; se facilita la búsqueda de este último desplazamiento químico utilizando el espectro 2D-HSQC y posteriormente, en el experimento 3D, trasladándonos al plano que corresponda con dicho desplazamiento químico. Ubicado el desplazamiento químico antes mencionado, se corta esta franja y se asigna al grupo de aminoácidos que compartan el sistema de espines que coincida con el que presenta el espectro. Para asignar a un aminoácido particular es necesario utilizar el experimento 3D-NOESY-HSQC (que se construye de la misma manera en la que se hizo el 3D-TOCSY-HSQC), para lo cual primero se obtienen las conectividades secuenciales, utilizando las correlaciones entre los HN del esqueleto. Una vez encontradas las secuencias se comienza a identificar las posibles combinaciones de las franjas con la estructura primaria de la proteína. Como se indicó anteriormente, cuatro aminoácidos son suficientes para hacer una asignación a dicha secuencia

con un 99 % de certeza. Cuando se han asignado todos, o la mayoría, de los sistemas de espines y se han colocado en su posición dentro de la secuencia correspondiente a la estructura primaria, se procede a obtener las restricciones de larga distancia de este mismo espectro.

**OBTENCIÓN DE LAS RESTRICCIONES**  
**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
*Señales por el efecto nuclear Overhauser (NOE).*

Una vez que se ha concluido el proceso de la asignación secuencial, el siguiente paso en el proceso de elucidación estructural es la extracción de la información espacial de estos datos de RMN. Cada una de estas piezas de información son utilizadas como restricción en los algoritmos adecuados, que de alguna manera generaran una familia de estructuras que satisfacen todas las restricciones experimentales. Cuando la desviación promedio estándar (RMS) entre todas ellas es menor a 2 Å, se dice que se obtiene la estructura tridimensional. Como ya se ha mencionado, la fuente más importante de información es el efecto nuclear Overhauser homonuclear (NOE  $^1\text{H}$ - $^1\text{H}$ ), que se observa entre protones que se encuentran a una distancia menor a 5 Å, comúnmente llamados *próximos en el espacio*. Las constantes de acoplamiento escalar (acoplamiento  $J$ , a tres enlaces de distancia) proporcionan información sobre los ángulos de torsión que también forman parte de las restricciones con las que se calculan las estructuras. Los desplazamientos químicos de  $^1\text{H}$  y  $^{13}\text{C}$  en la mayoría de los casos, son indicadores muy confiables de la estructura secundaria de la proteína, ya que estos dependen del ambiente químico que los rodea y éste se modifica significativamente si se trata de una  $\alpha$  hélice o una hoja plegada  $\beta$ .

Dos protones *próximos en el espacio* se acoplan *dipolarmente* (a diferencia del acoplamiento escalar, que es un mecanismo de acoplamiento a través de los enlaces). La *modulación* de este acoplamiento dipolar es resultado del movimiento molecular que permite la relajación de los protones; el efecto nuclear Overhauser es una manifestación de esta relajación, que puede observarse como una señal fuera de la diagonal en los experimentos 2D o 3D *NOESY*. Para moléculas enormes como las proteínas, esta relajación ocurre principalmente a través de la modulación del acoplamiento dipolar entre dos protones de una proteína que se está moviendo en la disolución. Al incrementarse el tamaño de la molécula, su movimiento caótico en el medio (y por lo tanto su modulación dipolar) ocurre a menor frecuencia:

$$\tau_c = \frac{4\pi\eta r_h^3}{3kT},$$

donde  $\eta$  es la viscosidad del disolvente,  $r_h$  es el radio hidrodinámico de la molécula (considerada como una esfera),  $k$  es la constante del Boltzmann;  $\tau_c$  (tiempo de correlación) indica cuanto tiempo tarda la molécula en girar 1 radian. Al imponerse el proceso de baja frecuencia, el acoplamiento de relajación cruzada (cross-relaxation) protón-protón se hace más eficiente. Cuando se incrementa la eficiencia de la relajación, se incrementa la eficiencia con la que se manifiesta el efecto nuclear Overhauser durante el periodo de mezclado. La eficiencia en la formación del NOE ( $\rho$ ) depende de la fuerza de la interacción dipolar entre dos protones

$$\rho = \frac{1}{r^6} \tau_c$$

donde  $r$  es la distancia entre los protones, y la dependencia  $\rho \propto 1/r^6$  indica que esta eficiencia con la que se manifiesta el NOE desaparece rápidamente con la distancia internuclear. El NOE generalmente no se observa entre protones que están alejados más de 5 Å. Por ello, la observación de señales fuera de la diagonal en los espectros NOESY de las proteínas proporciona información de gran utilidad sobre su estructura tridimensional (estructura terciaria). Como se observan varios cientos de señales NOE, incluso en las proteínas pequeñas, éstas son una de las fuentes de información más importantes en el proceso de elucidación estructural. Para ello es necesario asignar las señales a protones específicos, determinar su área o volumen (dependiendo del tipo de espectro utilizado, 2D o 3D) y de esto determinar la distancia relativa entre estos protones.

La asignación de todas, o casi todas, las señales se lleva a cabo en los experimentos anteriormente mencionados. Cualquier ambigüedad en la asignación se puede resolver mediante un proceso de cálculo iterativo, en el que se calculan estructuras de proteínas con los datos obtenidos. Estos resultados de estructuras preliminares se utilizan para resolver las asignaciones ambiguas, ya que estas señales serían inconsistentes con las estructuras calculadas. Se hacen nuevos cálculos incluyendo nuevas asignaciones y eliminando o corrigiendo las erróneas del cálculo anterior; las nuevas estructuras, deberán ser mejores que las anteriores, utilizándose para

hacer nuevas asignaciones. Este procedimiento se repite cuantas veces sea necesario, hasta que ya no haya ambigüedades que resolver o señales que asignar.

Las señales en el espectro *NOE* poseen diversas intensidades, por lo que al asignarlas se debe evaluar la distancia a la que están los protones. Como se mencionó anteriormente, la eficiencia con la que se crea el efecto nuclear Overhauser es proporcional a la distancia entre los protones. Sin embargo, debido a que los demás protones en la molécula generan campos oscilantes de frecuencias similares, (como resultado del movimiento molecular); éstos pueden contribuir a la relajación cruzada del protón en observación. Este fenómeno es llamado *difusión de espín* (DS), ya que la transferencia de magnetización tiene lugar en etapas sucesivas a través de los protones vecinos. Los resultados observables de la difusión de espín son: 1) cambio en la velocidad en la que se crea el *NOE* y 2) la aparición de falsas señales fuera de la diagonal que correlaciona a dos núcleos a través de un tercero, dando la apariencia de que están directamente relacionados. Esto último crea problemas de inconsistencia entre la estructura y la información experimental, por lo que debe de evitarse el fenómeno de difusión de espín. Como la DS es un fenómeno indirecto, necesita de un mayor tiempo para manifestarse, por lo que al utilizarse tiempos de mezclado cortos se favorecen las señales *NOE* y se suprimen las debidas a la difusión de espín.

La manifestación del *NOE* es aproximadamente lineal cuando se utilizan tiempos de mezclado cortos. En estas circunstancias se asume que la intensidad de la señal fuera de la diagonal es directamente proporcional a  $r^{-6}$ . La proporcionalidad puede estimarse de dos formas: a) midiendo la intensidad ( $I_{ref}$ ) de una señal entre dos protones separados por una distancia conocida ( $d_{ref}$ ), por ejemplo, la distancia  $^2H-^1H$  que es de 1.8 Å, o b) asignándole a la señal más intensa del espectro la distancia de 1.8 Å, esto último es lo que con mayor frecuencia se hace. De esta manera se puede estimar la distancia entre dos protones, utilizando la siguiente relación:

$$d_{ij} = d_{ref} \left( \frac{I_{ref}}{I_{ij}} \right)^{\frac{1}{6}},$$

donde  $I_{ij}$  es la intensidad de la señal fuera de la diagonal de interés. Debido a que estas distancias son estimaciones, es muy conveniente dividir las en tres categorías generales: 1.8-2.8 Å, 1.8-3.5 Å

y 1.8-5.0 Å, donde 1.8 Å es el radio de van der Waals que corresponde a la distancia de la nube electrónica de dos protones que están en contacto (cabe mencionar que la medición exacta de estas distancia no es fundamental para los cálculos de dinámica molecular, es más importante contar con una gran cantidad de datos, aunque éstas se clasifiquen en los tres grupos mencionados). Pareciera que dividir las distancias en categorías, descarta información, pero en realidad es mejor utilizar estas categorías que sobrestimar la información. De cualquier forma, se ha demostrado que el factor determinante en la calidad final de las estructuras obtenidas por RMN es el número de restricciones utilizadas, más que su precisión.

### Constantes de acoplamiento escalar ( $^3J$ )<sup>11</sup>

Desde hace mucho tiempo se conoce que la magnitud de las constantes de acoplamiento escalar depende de varios factores. En particular, las constantes de acoplamiento a tres enlaces de distancia tienen una gran dependencia con el ángulo diedro ( $\theta$ ) formado entre los dos átomos acoplados. Esta dependencia está descrita en la ecuación de Karplus:

$$J(\theta) = A \cos^2 \theta - B \cos \theta + C$$

donde las constantes  $A$ ,  $B$  y  $C$  son determinadas empíricamente para diferentes sistemas químicos, como el ángulo entre los  $H^N-H^\alpha$  en las proteínas (figura 1.3). Como es una función oscilante (figura 2.7), hay múltiples valores que satisfacen la ecuación, aunque en la práctica los ángulos diedros  $H^N-H^\alpha$  en proteínas están restringidos al intervalo  $\phi = -30^\circ$  hasta  $\phi = 180^\circ$ . Los elementos regulares de la estructura secundaria tienen valores característicos de  $^3J_{HN\alpha}$ : para las  $\alpha$ -hélices  $^3J \approx 3 - 5$  Hz ( $^3J \approx 3.9$  Hz,  $\phi \approx -57^\circ$ ) y para las hojas plegadas  $\beta$   $^3J \approx 8.5 - 12$  Hz (antiparalela  $^3J \approx 8.9$  Hz,  $\phi \approx -139^\circ$ , paralela  $^3J \approx 9.7$  Hz,  $\phi \approx -119^\circ$ ). Valores intermedios entre estos dos intervalos generalmente son el producto del promedio de los ángulos de torsión debido a la gran movilidad interna de las regiones terminales o de regiones internas sin estructura secundaria definida. Aunque los valores de  $^3J_{HN\alpha}$  se pueden medir para todos los residuos, se utilizan sólo aquellas de los aminoácidos que pertenecen a las regiones con estructura secundaria definida ( $\alpha$  hélices o las hojas plegadas  $\beta$ ), convirtiéndolas en ángulos para utilizarse como restricciones en el cálculo de las estructuras.

Aunque existen numerosos métodos para medir la constante de acoplamiento, se debe tener cuidado con la resolución de los espectros que se utilizarán con este fin. Se han desarrollado



métodos muy precisos para medir  ${}^3J_{HN\alpha}$  que utilizan la constante de acoplamiento  ${}^{15}\text{N}-{}^1\text{H}$ . Por ejemplo, el experimento HSQC-J modulado, que básicamente consiste en un experimento HSQC normal con un periodo extra de espera  $\tau_2$  previo a la adquisición de la señal. Durante esta espera, la magnetización del protón de la amida se ha marcado con la frecuencia de  ${}^{15}\text{N}$  durante  $t_1$ , para evolucionar de acuerdo con el acoplamiento al  ${}^1\text{H}$ . La intensidad de las señales fuera de la diagonal están moduladas por  ${}^3J_{HN\alpha}$  y  $\tau_2$ , de acuerdo con la siguiente ecuación:

$$I(\tau_2) = A[\cos(\pi J \tau_2) \cos(\pi J \tau_1) - 0.5 \sin(\pi J \tau_1) \sin(\pi J \tau_2)] e^{-\tau_2/T_2}$$

donde  $I(\tau_2)$  es el volumen de la señal fuera de la diagonal en función del tiempo de espera  $\tau_2$ ,  $A$  es el volumen de la señal fuera de la diagonal cuando  $\tau_2 \approx 0$ ,  $J$  es la constante de acoplamiento  ${}^3J_{HN\alpha}$  y  $T_2$  es el tiempo de relajación transverso aparente. Se obtienen varios espectros en los que se incrementan los valores de  $\tau_2$  y los cambios en la intensidad de las señales fuera de la diagonal se ajustan a la ecuación anterior a través de una regresión no lineal.

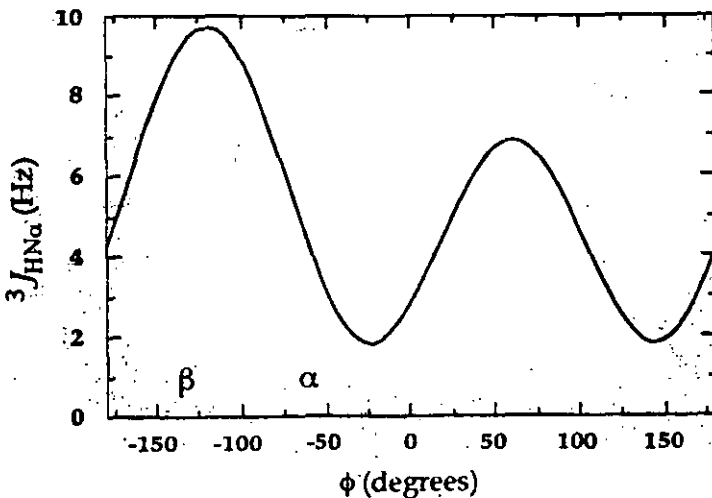


Figura 2.7.- Gráfico que muestra la dependencia de la constante de acoplamiento ( ${}^3J_{HN\alpha}$ ) en función del ángulo de torsión  $\phi$ .

### Desplazamiento químico.

El desplazamiento químico de los núcleos es muy sensible al ambiente químico que lo rodea. Se ha observado que el desplazamiento químico de los núcleos de protón ( ${}^1\text{H}$ ) y del

carbono ( $^{13}C^\alpha$ ,  $^{13}C^\beta$  y  $^{13}C^{\text{carbonilo}}$ ) están fuertemente influidos por el tipo de estructura secundaria a la que pertenecen. Por ejemplo, los protones  $H^\alpha$  pertenecientes a una  $\alpha$ -hélice muestran un corrimiento de su desplazamiento químico hacia campo alto, relativo a los valores que se presentan en los péptidos de cadena aleatoria; ocurre lo contrario (es decir, un corrimiento del desplazamiento químico hacia campo más bajo) si los  $H^\alpha$  se encuentran en una hoja plegada- $\beta$ . Estos patrones se han interpretado en términos de los índices de desplazamiento químico para los núcleos anteriormente mencionados, lo que permite una predicción muy precisa de los elementos de la estructura secundaria de la proteína.

### CÁLCULO DE LAS ESTRUCTURAS TERCIARIAS A PARTIR DE LOS DATOS DE RMN.

Un paso esencial en la aplicación de los experimentos de RMN-TF en la determinación de la estructura de las proteínas, ha consistido en el desarrollo de los métodos computacionales que utilizan las restricciones estructurales derivadas de RMN para generar las estructuras terciarias. A pesar de que se han propuesto múltiples métodos de cálculo, los más utilizados en la actualidad son: geometría de las distancias y dinámica molecular restringida.

#### Geometría de distancias.

Geometría de distancias es una rama de las matemáticas aplicadas a la construcción de estructuras a partir de distancias internas, siendo muy útil para la construcción de la estructura terciaria de las proteínas, partiendo de las restricciones impuestas por las distancias intramoleculares obtenidas por los experimentos. Existen dos algoritmos para el cálculo de las estructuras de proteínas a partir de los datos de RMN: uno que opera con distancias espaciales (llamado aproximación de matriz métrica) y el otro en el espacio de ángulos de torsión.

El primer programa de geometría de distancias desarrollado para calcular estructuras de datos de RMN estuvo basado en la aproximación de matriz métrica (DISGEO). Esta involucra cinco fases distintas: 1) generación de la matriz de las distancias de los enlaces, 2) refinamiento de las distancias de los enlaces, 3) generación de distancias de prueba, 4) fijación (embedding) y 5) optimización. En el primer paso, la matriz que se construye contiene los límites superior e inferior de las distancias entre los enlaces de dos átomos directamente unidos; una matriz simétrica de este tipo contiene  $(N^2 - 1)/2$  distancias para proteínas con  $N$  átomos. La distancia mínima inicial es la suma de los radios de van der Waals de los dos átomos, mientras la distancia

máxima es la derivada de los datos de los experimentos NOE, o en caso de no contar con ellos (como sucede cuando se utiliza el algoritmo de espacio de ángulos de torsión), se utiliza cualquier valor arbitrario, tan largo como el propio tamaño de la proteína. En las restricciones iniciales también se utilizan las distancias típicas de los enlaces covalentes y los ángulos predeterminados por la estructura primaria de la proteína.

A continuación se reduce el tamaño de la molécula, mediante un método llamado refinamiento del enlace. El método más simple de refinamiento utiliza la desigualdad del triángulo que incrementa el tamaño de la distancia mínima y disminuye simultáneamente la distancia máxima, hasta que ambos son geoméricamente consistentes. Considere tres átomos  $i$ ,  $j$  y  $k$  que se encuentran a distancias  $u_{ij}$ ,  $u_{ik}$  y  $u_{jk}$ . El átomo  $k$  se encuentra entre los átomos  $i$  y  $j$ , por lo que la distancia entre

$$u_{ij} \leq u_{ik} + u_{kj}$$

si en la matriz las distancias de los enlaces no satisfacen la ecuación anterior (es decir,  $u_{ij} > u_{ik} + u_{kj}$ ), entonces se reduce la distancia  $u_{ij}$  hasta que se iguala con la suma de las distancias  $u_{ik}$  y  $u_{kj}$ . Este procedimiento se aplica a todos los conjuntos de tres átomos, hasta que ya no existan modificaciones posibles.

Se aplica una desigualdad similar para los enlaces menores, permitiéndoles ajustarse hasta que alcancen el límite menor. Posterior al ajuste de la distancia de los enlaces, se eligen las distancias de prueba que se utilizarán en la generación de las coordenadas atómicas, fijándolas entre los límites superior e inferior previamente establecido. Este proceso de fijación genera una matriz métrica utilizando la regla de los cosenos. Las coordenadas tridimensionales son calculadas utilizando los tres eigenvalores máximos y los eigenvectores asociados. Eigenvectores negativos implican distancias negativas, por lo que se descartan las estructuras que den estos valores. Las estructuras que surgen de este proceso podrán no satisfacer todas las distancias, por lo que es necesario un proceso de optimización que satisfaga las distancias máxima y mínima de los enlaces.

La optimización se lleva a cabo con una función de error que mide dicho error entre las distancias de la estructura calculada y las condiciones límite. La forma típica de la función de error es (E):

$$E = \begin{cases} 0 & \text{si } d_{ij} \geq u_{ij} \\ k_{dist} (v_{ij} - d_{ij})^2 & \text{si } v_{ij} \leq d_{ij} \leq u_{ij} \\ k_{dist} (d_{ij} - v_{ij})^2 & \text{si } d_{ij} \geq v_{ij} \end{cases}$$

donde  $k_{dist}$  es un factor de ponderación y  $d_{ij}$  es el valor de la distancia al evaluarse el error (E). Se aplica entonces un algoritmo que disminuya E hasta un valor predeterminado.

Los programas desarrollados para funcionar en el espacio de ángulos de torsión, inician con una estructura aleatoria generada con base en la estructura primaria, la longitud de los enlaces y los ángulos estándares. La estructura inicial es refinada mediante la variación de los ángulos de torsión para minimizar una función objetivo. Como este método sólo varía los ángulos durante los cálculos, utiliza una tercera parte de los recursos de cómputo de los que utiliza el método basado en la matriz métrica.

La función objetivo utilizada en la minimización, es similar a la que se utiliza en el algoritmo anterior. El problema al tratar de minimizar la función objetivo introduciendo de manera simultánea todas las variables, es que la función posee múltiples mínimos locales que impiden que se alcance de manera directa el mínimo global. En lugar de ello, primero deberán de optimizar las restricciones locales (las restricciones secuenciales e intrarresiduo) y posteriormente se introducen las restricciones de mayor distancia, hasta que se hayan considerado todas las restricciones en el cálculo.

La geometría de distancias se considera el método óptimo (rápido y confiable) para calcular las estructuras iniciales de los datos de RMN. Sin embargo, cada una de las aproximaciones anteriores tienen ciertas dificultades.

El éxito del algoritmo de matriz métrica depende estrechamente de la adecuada elección de las distancias de prueba. En la mayoría de los programas que utilizan esta aproximación, dicha

distancia es un valor elegido de manera aleatoria entre los límites máximo y mínimo del enlace. Con ello se asume que el conjunto de las distancias posee una distribución uniforme entre los límites de enlace, mientras que en la teoría de polímeros se predice que la distribución estará sesgada hacia los límites inferiores, en las cadenas con libre giro. El resultado es que esta aproximación tiende a producir conformaciones extendidas con pequeña desviación media cuadrada (RMSD), contrario a las que se obtienen en los algoritmos de espacio de ángulos de torsión, especialmente en aquellas regiones en las que se poseen pocas restricciones NOE.

Los programas que calculan las estructuras mediante geometría de distancias en el espacio de ángulos de torsión, tienen una mayor probabilidad de proporcionar un plegado incorrecto del polipéptido (imágenes especulares). Sin embargo, estas estructuras son fácilmente detectadas y pueden descartarse inmediatamente.

#### Minimización de la energía y dinámica molecular restringida

Las estructuras calculadas con el método de geometría de las distancias poseen una elevada energía molecular irreal, debido a que no se toman en cuenta las interacciones entre los átomos no enlazados como las fuerzas de van der Waals y las interacciones Coulómbicas, o las consideran de una manera muy simplificada como términos de bajo nivel estérico. Consecuentemente, las estructuras obtenidas por este método necesitan ser refinadas mediante una simulación dinámica o una simple minimización de la energía. Un típico campo de fuerza físico o potencial efectivo para una proteína tiene la forma:

$$V_{\text{física}} = \sum_{\text{enlace}} \frac{1}{2} K_b (b - b_0)^2 + \sum_{\text{ángulo}} \frac{1}{2} K_\theta (\theta - \theta_0)^2 + \sum_{\text{diédros}} \frac{1}{2} K_\phi (\phi - \phi_0)^2 + \dots$$

$$\sum_{\text{diédros}} \frac{1}{2} K_\phi [1 + \cos(n\phi - \delta)] + \sum_{\text{pares}(i,j)} \left[ \frac{C_{12}(i,j)}{r_{ij}^{12}} - \frac{C_6(i,j)}{r_{ij}^6} + \frac{q_i q_j}{4\pi\epsilon_0 \epsilon_r r_{ij}} \right]$$

donde los términos  $K$  son constantes de fuerza. El primer término es un armónico de potencial que representa el estiramiento del enlace covalente a lo largo de  $b$  y la longitud del enlace con la menor energía  $b_0$ , que varía con el tipo de enlace covalente. Términos similares son utilizados para describir la energía de *arqueo* de los ángulos de enlace  $\theta$ . Se utilizan dos términos en la

descripción de la distorsión de los ángulos diedros: un término armónico utilizado en ángulos diedros  $\zeta$  en los que no son posibles las transiciones (por ejemplo, los ángulos diedros que forman los anillos aromáticos), mientras que el término senoidal es utilizado para ángulos diedros  $\phi$  que pueden rotar  $360^\circ$ . El término final es la suma de todas las interacciones interatómicas de los pares de protones: la primera parte es la suma de las interacciones de van der Waals (un potencial típico de Lennard-Jones 6-12) y la segunda parte es la suma de todas las interacciones Coulómbicas.

Aunque existen numerosas variantes de este campo de fuerza físico, que incluyen explícitamente términos de enlace por puentes de hidrógeno, las aproximaciones generales al refinamiento estructural mediante RMN es el mismo. La estrategia general consiste en adicionar un potencial de restricción al campo de fuerza, tal que la estructura sea refinada contra los términos físicos dentro del campo de fuerza (es decir, la geometría covalente y las interacciones de no-enlace) mientras cumpla con las restricciones experimentales. Así:

$$V_{total} = V_{físico} + V_{res.exp}$$

Los términos de restricciones experimentales generalmente son potenciales elevados al cuadrado, de manera similar a los que se utilizan en los algoritmos de geometría de distancias:

$$V_{res.exp} \cong V_{NOE} + V_{diedro}$$

donde

$$V_{NOE} = \begin{cases} k_{NOE} (d_{ij} - u_{ij})^2 & \text{si } d_{ij} > u_{ij} \\ 0 & \text{si } l_{ij} \leq d_{ij} \leq u_{ij} \\ k_{NOE} (l_{ij} - d_{ij})^2 & \text{si } l_{ij} > d_{ij} \end{cases}$$

y

$$V_{\text{diebro}} = \begin{cases} k_{\text{diebro}} (\varphi_{ij} - \varphi_{ij}^*)^2 & \text{si } \varphi_{ij} > \varphi_{ij}^* \\ 0 & \text{si } \varphi_{ij}' \leq \varphi_{ij} \leq \varphi_{ij}^* \\ k_{\text{diebro}} (\varphi_{ij}' - \varphi_{ij})^2 & \text{si } \varphi_{ij}' > \varphi_{ij} \end{cases}$$

$k_{NOE}$  y  $k_{\text{diebro}}$  son constantes de fuerza, que son utilizados para valorar estos términos relativos respecto a otros términos físicos en un campo de fuerza.

En otros estudios, las estructuras obtenidas mediante geometría de distancias son refinadas mediante la simple minimización de energía en un campo de fuerza de dinámica molecular expandido. Las constantes de fuerza  $k_{NOE}$  y  $k_{\text{diebro}}$  usualmente se eligen de entre cualquier valor arbitrariamente bajo de las restricciones experimentales, que corresponden a una energía  $kT/2$ , el promedio de la energía cinética de un átomo por grado de libertad. En otras palabras, se toleran pequeñas violaciones, ya que la molécula posee un número considerable de grados de libertad. El problema con esta aproximación simplista es que numerosas estructuras caerán en los mínimos locales cercanos y pueden nunca alcanzar el mínimo de energía global.

### Desarrollo experimental

La proteína enriquecida en  $^{15}\text{N}$  fue obtenida en los laboratorios de la Harvard Medical School por los investigadores del grupo del Doctor G. Wagner mediante técnicas de ingeniería genética, incorporando el plásmido a microorganismos adecuados para la expresión de la proteína de interés, desarrollándolos en un medio cuya única fuente de nitrógeno era cloruro de amonio enriquecido ( $^{15}\text{NH}_4\text{Cl}$ ). Obtenida la muestra, se separó mediante técnica de cromatografía de partición. Purificada la proteína, se disolvió en un buffer de tris-etilendiamina a pH 7-7.5, en una mezcla de  $\text{H}_2\text{O}/\text{D}_2\text{O}$  90:10. Los espectros 2D-HSQC, 3D-TOCSY-HSQC y 3D-NOESY-HSQC se obtuvieron con esta muestra en equipos de RMN a 750 MHz y 500 MHz para la resonancia del  $^1\text{H}$ ; Varian Unity Plus.

El experimento 3D-TOCSY-HSQC fue obtenido con un tiempo de "spin lock" de 20 ms, mientras que en el experimento 3D-NOESY-HSQC se utilizó un tiempo de mezclado de 0.100 s.

Los experimentos se procesaron con una función seno cuadrado a  $90^\circ$ , con 256 puntos en la dimensión de  $^{15}\text{N}$  y 512 puntos en cada dimensión de  $^1\text{H}$ . Con el fin de incrementar al doble el número de puntos en la dimensión indirecta, se realizó una predicción lineal de 128 puntos en la dimensión de  $^{15}\text{N}$ , corrigiendo la línea base de acuerdo con el algoritmo de Shmider-Wagner. Así se incrementó el número de datos con información, en lugar de utilizar el llenado con ceros (zero filling), típicamente utilizado en RMN-TF, para incrementar la resolución de los espectros. El proceso se llevó a cabo con el programa Felix (MSI). Los espectros en tres dimensiones (una vez procesados mediante tres transformaciones de Fourier), son analizados y se asignan las señales con ayuda del programa Xeasy.<sup>13</sup> (ver. 1.2). La integración de los picos de las señales del espectro 3D-NOESY-HSQC se llevó a cabo con la subrutina *peakint* que está integrada al programa Xeasy; los datos de volumen obtenidos en la integración son transformados en restricciones de distancias utilizando el programa Caliba.<sup>13</sup> Los cálculos de las estructuras se realizaron en la super computadora Berenice8 (8 procesadores R10000 a 195 MHz con 2048 Mb de memoria RAM), Bonampak (procesador R4400 a 250 MHz con 256 Mb de memoria RAM) y Uxmal (procesador Pentium III a 500 MHz con 256 Mb de memoria RAM), en todas ellas se utiliza el programa CNS<sup>14</sup> (Crystallographics and Nuclear Magnetic Resonance System, ver. 1.0a). El procesamiento de los datos de energía se efectuaron mediante programas desarrollados por el



propio grupo de trabajo. La visualización de las estructuras se efectuó en Bonampak utilizando el programa MOLMOL<sup>15</sup> (MOLEculas display & MOLEculas procesing, ver. 2K.1).

### *Identificación de las señales del espectro 3D-TOCSY-HSQC, ayudado por el experimento*

*HSQC, para una asignación probable a determinado grupo de aminoácidos.*

El primer paso en el proceso de asignación de señales, consiste en determinar el desplazamiento químico del protón de amida del esqueleto de la proteína ( $^1\text{HN}$ ) en el espectro 3D  $^1\text{H}$ - $^{15}\text{N}$ -TOCSY-HSQC, en el cual dos dimensiones son de protón (y son las que forman los planos de la parte correspondiente al experimento TOCSY) y la tercera (dimensión (dimensión fuera del plano) es de nitrógeno-15 (parte heteronuclear que dispersa las señales evitando el traslape, experimento HSQC). El programa Xeasy, utilizado en el proceso de asignación de las señales, sólo puede mostrar espectros en dos dimensiones, que correspondería a uno sólo de los 128 planos que integran el espectro en tres dimensiones; sería un proceso muy laborioso ver cada uno de los planos para determinar en cual se encuentra la señal del  $^1\text{HN}$ . Para facilitar este trabajo se utiliza el experimento 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC.

El experimento 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC revela de manera inmediata los desplazamientos químicos de los protones de amida tanto en el desplazamiento químico de protón ( $^1\text{H}$ ) como en el desplazamiento químico de nitrógeno ( $^{15}\text{N}$ ). Ya que hemos obtenido los desplazamientos químicos de los 104 aminoácidos (las prolinas carecen de protón de amida, por lo que no aparecen en estos experimentos) utilizamos el experimento 3D  $^1\text{H}$ - $^{15}\text{N}$ -TOCSY-HSQC.

Usando los desplazamientos químicos de nitrógeno obtenidos del experimento 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC, por medio del programa Xeasy, nos movemos al plano que corresponda a dicho desplazamiento químico (de  $^{15}\text{N}$ ); en este plano asignaremos la señal de  $^1\text{HN}$  a la que posea los dos desplazamientos químicos de protón idénticos al que se determinó en el experimento 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC; esta señal corresponde a la diagonal en los experimentos típicos de dos dimensiones. Una vez que hemos encontrado las señales de los protones  $^1\text{HN}$ , cortamos franjas (de aproximadamente 0.3 ppm de ancho por 13 ppm de largo), utilizadas en la identificación de los sistemas de espín y la asignación a un grupo de aminoácidos.

El experimento TOCSY puede mostrar (en teoría) todas las señales de la molécula; en la práctica sólo se observan señales de correlación entre protones que no rebasen los 7 enlaces de distancia. Afortunadamente, el aminoácido más largo es la lisina, en la que hay 7 enlaces entre el HN y  $^6\text{H}$ ; en los aminoácidos restantes el HN se encuentra separado 4 ó 5 enlaces en promedio de los demás protones, por lo que en la gran mayoría de las ocasiones se observan todas las señales de los protones que contienen los aminoácidos, en los desplazamientos químicos acorde al ambiente electrónico que los rodea (estas dos características, el número de señales observadas y la posición de cada una de ellas, son las que definen el sistema de espín de cada uno de los aminoácidos).

En cada una de las franjas que hemos "cortado" del espectro analizaremos el sistema de espín de los aminoácidos con el fin de identificarlo y asignarlo dentro de la estructura primaria de la proteína. Desafortunadamente el sistema de espín de ciertos aminoácidos es muy similar entre sí (presentan el mismo número de señales y desplazamiento químico de estas es muy similar), por lo que solamente se puede asignar cada franja a un grupo de aminoácidos que comparten el sistema de espines (figura 3.1). Así, podemos englobar a 8 aminoácidos en un solo grupo en el que se observan tres señales (que corresponden al HN,  $^{\alpha}\text{H}$  y  $^{\beta}\text{H}$ ): Asp, Asn, Cys, His, Phe, Ser, Trp y Tyr. En el grupo conformado por: Gln, Glu y Met se observan cuatro señales. El grupo que está compuesto por Ile, Leu y Val presenta mayores dificultades, ya que pueden presentar 4 ó 5 señales, debido a la similitud de los desplazamientos químicos de la región alifática. Los aminoácidos restantes no se engloban en ningún grupo, por que sus desplazamientos químicos los hacen característicos y su identificación es menos ardua: la alanina (Ala) presenta tres señales con desplazamientos químicos perfectamente distinguibles; la treonina (Thr) muestra 4 señales; la lisina (Lys) muestra 5 ó 6 señales en el espectro ( $^{\beta}\text{H}$  y  $^{\delta}\text{H}$  tienen desplazamientos similares), mientras que la arginina (Arg) muestra 4 ó 5 ( $^{\beta}\text{H}$  y  $^{\gamma}\text{H}$  son similares) y la glicina (Gly) generalmente sólo presenta 2 señales (ocasionalmente los  $^{\alpha}\text{H}$  son distintos y aparece una señal para cada uno de ellos). La prolina (Pro) no aparece en estos espectros por carecer de protón HN (figura 2.4).

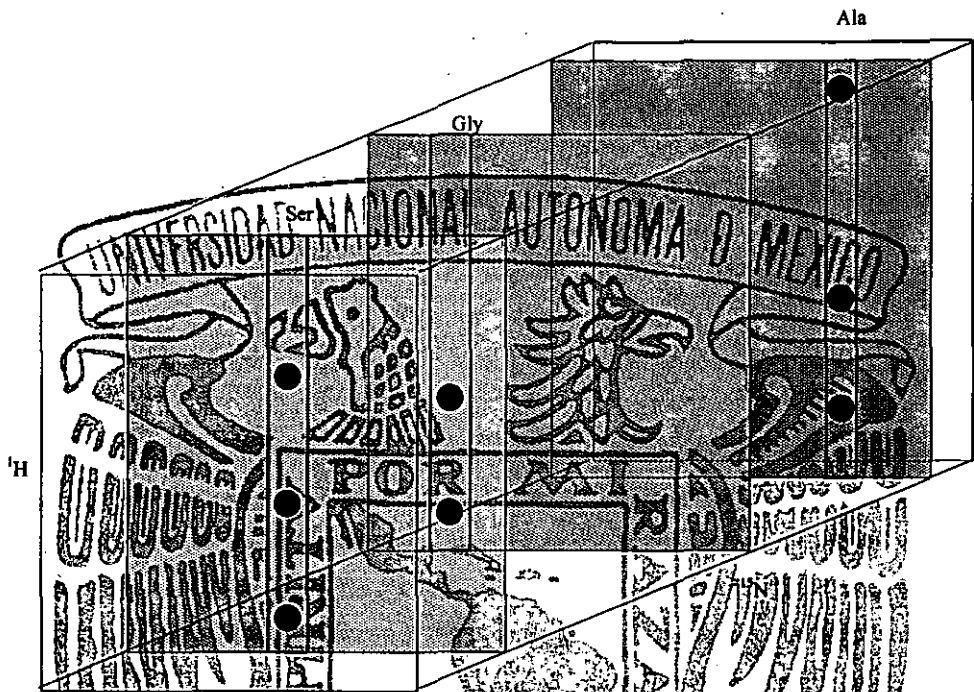


Figura 3.1. Esquema de un espectro 3D  $^1\text{H}$ - $^{15}\text{N}$ -TOCSY-HSQC con tres franjas: serina, glicina y alanina. Obsérvese que los sistemas de espín de estos aminoácidos se encuentran dispersos por el desplazamiento químico de  $^{15}\text{N}$ . La gran ventaja de los espectros 3D es que disminuye el traslape de señales típico de los experimentos 2D. En el esquema se presentan dos aminoácidos cuyo sistema de espín nos permite identificarlos fácilmente: alanina y glicina; la franja de la serina muestra un sistema de espines que comparten 8 aminoácidos (Asp, Asn, Cys, Ser, His, Phe, Trp, Tyr) por lo que es difícil asignarlo a uno de ellos, sin más información.

**Secuenciación de las franjas asignadas a cada residuo, para su ordenación dentro de la estructura primaria, utilizando el espectro 3D  $^1\text{H}$ - $^{15}\text{N}$ -NOESY-HSQC.**

Para efectuar la asignación precisa de cada una de las franjas a un aminoácido particular, se trabaja con el espectro 3D  $^1\text{H}$ - $^{15}\text{N}$ -NOESY-HSQC. El efecto nuclear Overhauser (NOE) proporciona señales de correlación entre dos protones que se encuentran próximos en el espacio (la distancia máxima a la que podemos observar este efecto es aproximadamente 5 Å), sin importar el número de enlaces entre ellos, a diferencia del experimento TOCSY en el que los núcleos necesariamente tienen que estar acoplados y a menos de 7 enlaces de distancia. Otra característica importante del NOE es que se observan señales de correlación entre protones interresiduos (dos residuos de aminoácidos distintos) e intrarresiduo (del mismo residuo de aminoácido), mientras que en el TOCSY sólo se observan señales intrarresiduo.

En el experimento 3D  $^1\text{H}$ - $^{15}\text{N}$ -NOESY-HSQC utilizamos los desplazamientos químicos de las señales que fueron asignadas previamente con el experimento 3D  $^1\text{H}$ - $^{15}\text{N}$ -TOCSY-HSQC (debido a que las 3 dimensiones utilizadas en ambos experimentos son las mismas). El experimento NOE correlaciona protones que se encuentren a menos de 5 Å de distancia; los protones de amida ( $\text{HN}$ ) del esqueleto de la proteína se encuentran entre 2.8 Å (distancia promedio  $\text{H}_i\text{N}_i - \text{H}_{i+1}\text{N}$  en las  $\alpha$ -hélices) y 4.3 Å (distancia promedio  $\text{H}_i\text{N}_i - \text{H}_{i+1}\text{N}$  dentro de las hojas plegadas  $\beta$ ), el experimento NOE, presentará señales de correlación entre el protón  $\text{H}_i\text{N}$  (del  $i$ -ésimo residuo) y el protón  $\text{H}_{i+1}\text{N}$  (de aminoácido posterior al  $i$ -ésimo residuo) que utilizaremos para determinar la secuencia; además, la distancia entre el  $^1\text{H}-\text{H}_{i+1}\text{N}$  es de 3.5 Å en una  $\alpha$ -hélice y de 2.2 Å en una hoja plegada  $\beta$ , esta señal de correlación nos proporciona gran ayuda para la determinación del orden progresivo (ordenación de los aminoácidos en dirección del extremo amino terminal hacia el carboxilo terminal) de las franjas, uniéndolos en cadenas. Otras señales que nos son de gran utilidad en el proceso de formación de las cadenas son  $^1\text{H}-\text{H}_i\text{N}$ ,  $^1\text{H}-\text{H}_{i+1}\text{N}$  y  $^8\text{H}-\text{H}_i\text{N}$ ; con ellas confirmamos la correcta ordenación de las cadenas de franjas.

Una vez ordenadas las franjas en cadenas, utilizaremos la estructura primaria de la proteína para asignar de manera única e inequívoca cada franja a un aminoácido particular. En este proceso tomamos de cualquier cadena una franja  $x$  (que asignaremos como  $i$ ) asignada a un grupo de aminoácidos (Glu, Gln ó Met) y le asignamos provisionalmente uno de ellos (Met, por ejemplo); buscamos en la estructura primaria las posiciones en las que se presenta la metionina (todos los aminoácidos se encuentran más de 1 vez en la SV40); elegimos una de las posiciones en las que se presente y observamos cuál es el aminoácido siguiente; examinamos si el sistema de espines de la  $i+1$ -franja coincide con el que observaríamos para el aminoácido presente en la secuencia. Si el sistema de espines de la franja  $i+1$  no coincide con el aminoácido presente en la secuencia, entonces elegimos la siguiente posición en la que encontramos al aminoácido Met. Ubicada la nueva posición, volvemos a observar el aminoácido  $i+1$ ; nos regresamos a la  $i+1$  franja y observamos si el sistema de espines de ésta coincide con las que presentaría el aminoácido  $i+1$  presente en la estructura primaria. Si nuevamente no coinciden las señales con el sistema de espín, elegimos la posición siguiente y continuamos de esta manera, hasta que se terminen todas las posibilidades para el aminoácido  $i = \text{Met}$ . En el caso de que no hayan coincidido, entonces probamos asignando la franja  $i$  a otro aminoácido de ese grupo, por ejemplo Glu; repetimos el proceso de observar el aminoácido  $i+1$  y verificar con el sistema de espines de

la franja  $i+1$ . Cuando encontremos que el sistema de espines de dos franjas adyacentes coinciden con dos aminoácidos contiguos en la estructura primaria, revisamos la  $i+2$  franja, asignada a algún grupo de aminoácidos, cuyo sistema de espines debe coincidir con el aminoácido  $i+2$ .

Existen muy pocas probabilidades de que una secuencia específica de tres aminoácidos se repita en la estructura de una proteína (la probabilidad de que una secuencia de tres aminoácidos se repita dos veces en una misma proteína es de 0.0146 %), por lo que al identificar y asignar las franjas al aminoácido 18 (p. ej.)  $i+1$  al 19 e  $i+2$  al 20, se facilita la asignación de las franjas que se encuentran alrededor de esta secuencia (...  $i-2$ ,  $i-1$ , ...,  $i+3$ ,  $i+4$ , ...) pues sólo tendremos que observar la estructura primaria de la proteína e ir asignando (y verificando) el aminoácido que corresponda con la franja observada.

#### Identificación y asignación de las señales de correlación a larga distancia.

Concluido el trabajo de asignación secuencial (identificación del aminoácido al que corresponde cada franja), pasamos a la parte más difícil del proceso de asignación de señales: encontrar las correlaciones de larga distancia, que nos indicarán la proximidad de dos protones (separados por una distancia menor a 5 Å) que se ubican en dos residuos de aminoácido lejanos en la estructura primaria; de estas correlaciones se desprenderán las restricciones para obtener la estructura terciaria de la proteína. Este proceso requiere de mucho tiempo de minucioso análisis de los espectros, ya que el desconocer la estructura terciaria de la proteína no nos permite saber *a priori* cuáles son los residuos de aminoácido que se encuentran cercanos y cuáles son sus señales de correlación.

Los protones intrarresiduo se encuentran a una distancia menor de 5 Å, razón por la cual el experimento NOESY presentará (en la mayoría de los casos) las mismas señales de correlación observadas en el experimento TOCSY. Sin embargo, el experimento 3D  $^1\text{H}$ - $^{15}\text{N}$ -NOESY-HSQC mostrará señales que no se presentan en el espectro TOCSY, que son el producto de la correlación entre protones de aminoácidos distantes, y serán estas señales las que más nos interesan para la obtención de las restricciones de larga distancia, con las que fijaremos la estructura terciaria de nuestra proteína. Por ejemplo, la señal del protón  $^a\text{H}$ , en la franja de  $\text{H}_i\text{N}$ , por su proximidad espacial con los protones del aminoácido  $j$ , mostrará una señal de correlación

con el aminoácido *j*; esta señal será muy parecida a la que se muestra en  $H_1N$  pero se presentará en la franja del  $H_1N$  ubicada en el desplazamiento químico de  $^{15}N$  (figura 3.2).

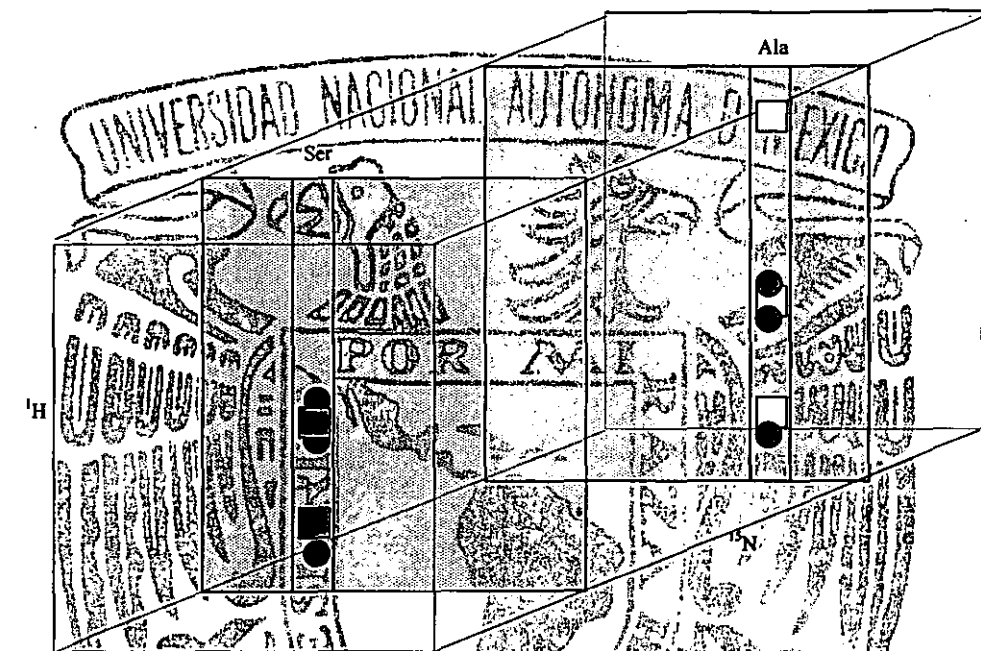


Figura 3.2. Espectro 3D  $^1H$ - $^{15}N$ -NOESY-HSQC. En él se observan las señales de correlación intrarresiduo (similares a la que muestra el espectro TOCSY) y señales que no pertenecen al sistema de espín: estas señales de correlación de larga distancia las convertiremos en restricciones para calcular la estructura terciaria de la proteína. Los círculos negros representan las señales de correlación intrarresiduo (TOCSY) de una serina, los cuadros blancos representan las correlaciones intrarresiduo de una alanina y los cuadros y círculos grises las señales de correlación de larga distancia (NOE).

Tratar de encontrar la correlación de estas dos señales observando uno por uno los 128 planos que conforman el espectro, sería un proceso muy lento, por lo que es preferible utilizar las subrutinas integradas en el programa Xeasy que facilitan el trabajo de búsqueda de las señales de correlación. Una de estas subrutinas de Xeasy nos muestra un corte transversal del espectro 3D a través de la dimensión de  $^{15}N$ , dimensión utilizada en la dispersión de las señales de  $^1H$  (los planos desplegados por el programa muestran planos cuyas coordenadas son de desplazamientos químicos de  $^1H$ , figura 3.3). Las señales de correlación forman un cuadrado, dentro del espectro se escogen los desplazamientos químicos que forman dicho cuadrado y se hace un corte transversal del espectro, es más sencillo de trabajar el espectro de 1 dimensión de protón (a lo largo de los desplazamientos químicos de nitrógeno) lo que nos permite comparar fácilmente

estas cuatro regiones; cuando dos señales de larga distancia correlacionan, dos picos coinciden en el desplazamiento químico en las que se presentan las señales de correlación de larga distancia.

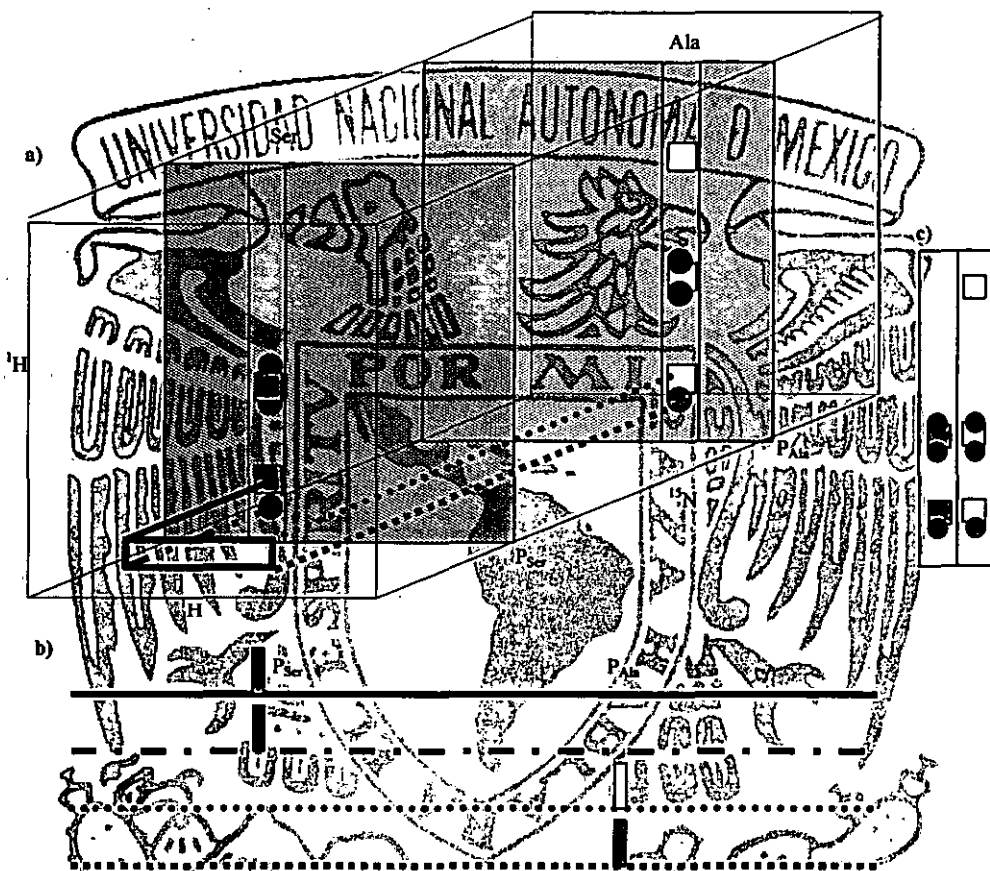


Figura 3.3.- Esquema de la subrutina para la localización de las señales de correlación de larga distancia a través del espectro 3D. Tomamos el ejemplo de la figura 3.2: una serina y una alanina. Si nos encontramos en el plano de la serina ( $P_{\text{ser}}$ ), marcamos los desplazamientos químicos donde esperamos encontrar las señales de correlación con la alanina (a). Obtenemos de esos 4 puntos, el trazo del espectro, a través del desplazamiento de  $^{15}\text{N}$  (b); en los espectros 1D y 2D obtenemos el desplazamiento químico de  $^{13}\text{C}$  donde aparecen las 2 señales con el mismo  $\delta^{13}\text{C}$ . Regresamos al espectro 3D y colocamos el espectro en el desplazamiento  $\delta^{13}\text{C}$  encontrado (que corresponderá al desplazamiento químico de la alanina,  $P_{\text{ala}}$ ). Cortamos las franjas en  $P_{\text{ser}}$  y  $P_{\text{ala}}$ , colocándolas en un solo plano (como si fuera un espectro 2D), que facilita la interpretación (c).

Adicionalmente a la anterior, otro programa (xppm) busca señales, alrededor de un desplazamiento químico determinado (con un pequeño intervalo de certidumbre), de la base de datos que se ha generado en el proceso de asignación y los aminoácidos en los que se encuentran las señales encontradas. De esta forma obtenemos la señal del protón con el que probablemente esta correlacionando ( $^{\alpha}\text{H}$ ,  $^{\beta}\text{H}$ ,  $^{\gamma}\text{H}$ , etc.) y los desplazamientos químicos en los que se encuentra.

Con los desplazamientos químicos nos trasladamos a ese plano del espectro y cortamos una franja, colocamos juntas ambas franjas para facilitar la comparación de estas señales y la determinación de que efectivamente se traten de señales de correlación o sólo es una de tantas señales que tienen desplazamientos químicos parecidos. Las señales de correlación de larga distancia (entre los aminoácidos A y B) generalmente están acompañadas por diversas correlaciones entre los dos:  $H_A N-H_B N$ ,  $H_A N-\alpha H_B$ ,  $H_A N-\beta H_B$ ,  $H_A N-\gamma H_B$ ,  $\alpha H_A-H_B N$ ,  $\beta H_A-H_B N$ ,  $\gamma H_A-H_B N$ , etc., además, es característico en los elementos regulares de estructura secundaria, por la proximidad entre A y B, que el aminoácido A-1 esté cerca de B+1', el aminoácido A-2 esté cerca al aminoácido B+2, etc., con lo cual se confirman las asignaciones a los elementos regulares de la estructura secundaria (figura 4.7).

**Identificación de los elementos regulares de la estructura secundaria de la proteína SV-40 (1-108), a través del índice de proton.**

Los elementos regulares de la estructura secundaria podemos identificarlos fácilmente utilizando la variación en los desplazamientos químicos de los protones alfa ( $^1H$ ) (debido a que los diferentes ambientes de densidad electrónica que prevalecen en una  $\alpha$ -hélice, en una hoja plegada  $\beta$  o en una región sin estructura secundaria definida modifican sustancialmente el desplazamiento químicos de estos protones). Como se mencionó anteriormente, el desplazamiento químico de los núcleos es muy sensible a las diferencias en el ambiente químico que lo rodean. Si una región se encuentra en una zona en la que se tiene mucho contacto con el disolvente ( $H_2O$ ) los desplazamientos químicos serán muy parecidos a los que se observan cuando se estudian oligopéptidos; un fenómeno similar sucede en las hojas plegadas  $\beta$  en la que su estructura lineal queda muy expuesta al disolvente; en estas regiones los desplazamientos químicos de los  $^1H$  se encuentran a campos más bajos (mayores ppm). Las  $\alpha$ -hélices son estructuras que tienen un menor contacto con el disolvente: estas estructuras se enrollan sobre sí mismas alrededor de un eje, adquiriendo una conformación compacta en la región interna y las cadenas laterales de los aminoácidos excluyen al disolvente de esta región (en la zona interna de la hélice se encuentran los protones  $^1H$ ); la presencia de las cadenas laterales de carácter hidrófobo, ejercen un efecto de protección magnética sobre los  $^1H$  modificando su desplazamiento químico hacia campos altos (menores ppm).

Por lo anterior, es posible determinar fácilmente las regiones que poseen determinada estructura secundaria y los aminoácidos que conforman dicha región. Para ello obtenemos el



desplazamiento químico de los  $^2\text{H}$  de cada aminoácido; si su desplazamiento químico se modifica de su valor típico en al menos 0.1 ppm podemos intuir que dicho aminoácido pertenece a una región de estructura secundaria definida. Para establecer a qué elemento regular pertenece una serie de protones, basándonos en las diferencias de desplazamiento químico, se construye el gráfico de índice de protón  $I^p$  en el que se observan un grupo de aminoácidos cuyo desplazamiento químico sufrió una modificación hacia campo alto (modificación debida a la presencia de una  $\alpha$ -hélice), hacia campos más bajos (modificación atribuida a la presencia de una hoja plegada  $\beta$ ) o aquellas zonas en las que no se modifica sustancialmente el desplazamiento químico, correspondiente a una región sin estructura secundaria definida.

Para hacer la gráfica de índice de protón primero obtenemos los desplazamientos químicos de los  $^2\text{H}$  asignados durante la fase de secuenciación. Se obtiene la diferencia de los desplazamientos químicos observados ( $\delta_{\text{obs}}$ ) y de los desplazamientos químicos estándares ( $\delta_{\text{std}}$ ) y se asigna un valor dependiendo de las siguientes condiciones:

$$\begin{aligned} \text{si } \delta_{\text{obs}} - \delta_{\text{std}} < -0.1 & \quad \text{entonces } I^p = -1 \\ \text{si } -0.1 \leq \delta_{\text{obs}} - \delta_{\text{std}} \leq 0.1 & \quad \text{entonces } I^p = 0 \\ \text{si } \delta_{\text{obs}} - \delta_{\text{std}} > 0.1 & \quad \text{entonces } I^p = 1 \end{aligned}$$

Si una serie de aminoácidos (de al menos 4 residuos) tienen un valor de índice de protón igual a -1, entonces esta región forma parte de una  $\alpha$ -hélice. Si la serie de al menos 3 residuos de aminoácidos dan una serie de valores de 1, significa que esta región posee una estructura de hoja plegada  $\beta$ . Cualquier otra serie de valores es indicio de una región sin estructura secundaria definida.

Al hacer y analizar la gráfica de índice de protón (figura 3.4), encontramos que el fragmento de los aminoácidos 1 a 108 de la proteína SV40 está constituida únicamente por  $\alpha$  hélices; se determinó que las cuatro regiones de  $\alpha$  hélices se extiende a lo largo de los siguientes residuos de aminoácido:

Hélice	Aminoácidos
$\alpha$ -hélice I	9-19
$\alpha$ -hélice II	24-41
$\alpha$ -hélice III	49-71
$\alpha$ -hélice IV	89-103

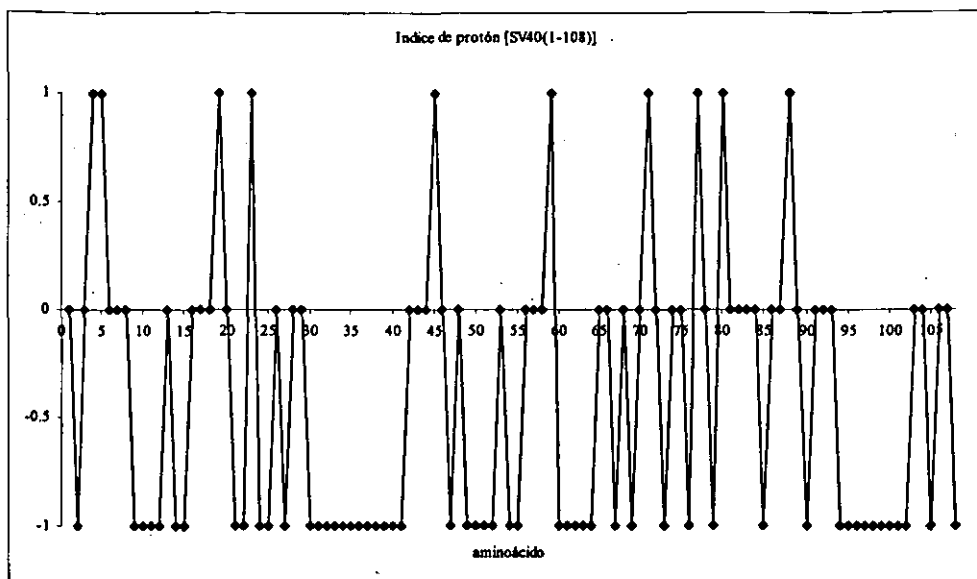


Figura 3.4. Gráfica de índice de protón, en la que se muestran las regiones que conforman la estructura secundaria. Para esta gráfica se utiliza el desplazamiento químico de los  $^7\text{H}_1$ : si  $(\delta_{\text{obs}} - \delta_{\text{ind}}) < -0.1$ , asignamos un valor de índice -1; si  $(\delta_{\text{obs}} - \delta_{\text{ind}}) > 0.1$ , el índice es 1; si  $-0.1 \leq (\delta_{\text{obs}} - \delta_{\text{ind}}) \leq 0.1$ , el índice es 0. Índices continuos nos indican la presencia de elementos regulares de estructura secundaria; si hay al menos 4 índices -1 continuos, indican la presencia de una  $\alpha$ -hélice; si hay al menos 3 índices 1 se tratará de una región tipo hoja plegada  $\beta$ . En cualquier otra serie de índices no es posible afirmar la existencia de un elemento de estructura secundaria.

Con la gráfica de índice de protón no se encontró la presencia de hojas plegadas  $\beta$ ; las regiones restantes de la proteína no poseen una estructura secundaria definida.

Además de los datos de índice de protón, se utiliza el patrón de señales de correlación NOE de corta distancia, para confirmar la presencia de elementos con estructura secundaria definida.<sup>7,9,10,12</sup> La interpretación cualitativa de las señales NOE secuenciales nos permite distinguir los elementos de estructura secundaria; por ejemplo, las  $\alpha$ -hélices se caracterizan por la presencia de señales muy intensas  $\text{H}_i\text{N}-\text{H}_{i+1}\text{N}$ , además de señales de intensidad media (o débil)  $^{\alpha}\text{H}_i-\text{H}_{i+1}\text{N}$ ,  $^{\alpha}\text{H}_i-\text{H}_{i+3}\text{N}$ ,  $\text{H}_i\text{N}-\text{H}_{i+2}\text{N}$  y ocasionalmente se observa  $^{\alpha}\text{H}_i-\text{H}_{i+4}\text{N}$ . Mientras que las hojas plegadas- $\beta$  se caracterizan por señales muy débiles  $\text{H}_i\text{N}-\text{H}_{i+1}\text{N}$  y muy intensas  $^{\alpha}\text{H}_i-\text{H}_{i+1}\text{N}$ .<sup>7,9,10</sup>

### *Cálculo de las estructuras.*

El proceso de elucidación de estructuras de proteínas es un trabajo cíclico: una vez que hemos asignado el número suficiente de señales, obtenemos de ellas las restricciones para efectuar los cálculos, obteniendo estructuras que se ajusten a las restricciones. Aquellas restricciones que no se ajusten a la estructura calculada producirán violaciones e incrementarán la energía total del sistema en cálculos de dinámica molecular. Las restricciones con el mayor número de violaciones se revisarán para determinar la causa de la violación, siendo las más comunes la asignación errónea o una integración inadecuada que proporciona una distancia menor. Cuando la asignación se haya verificado calcularemos nuevamente la estructura para determinar nuevas violaciones, o para continuar con el proceso de nuevas asignaciones y mejorar la estructura.

Para calcular las estructuras se utilizó el programa CNS, que funciona mediante el algoritmo de geometría de las distancias y posteriormente efectúa un refinamiento estructural, disminuyendo la energía total del sistema.



### Análisis de resultados y discusión

En cualquier determinación estructural por RMN es necesario primero asociar cada una de las señales de los espectros con un protón o grupo de protones de la molécula,<sup>6,7,8,9,10,12</sup> a grandes rasgos, hay dos estrategias de asignación de señales de resonancia para determinar la estructura de las proteínas que dependen principalmente de su peso molecular: A) la estrategia homonuclear, en la que se utilizan experimentos en dos dimensiones (COSY, TOCSY, NOESY, principalmente), utilizada cuando la proteína tiene un peso molecular menor a 5 kDa (50 residuos de aminoácido, aproximadamente). B) la estrategia heteronuclear, utilizada cuando al incrementar el tamaño de la proteína se sobrepasan los 7 kDa de peso molecular; este incremento en el tamaño implica que el número de protones se eleva proporcionalmente al número de residuos y por la gran cantidad de protones presentes en la proteína se incrementa la complejidad de los espectros en dos dimensiones provocando un traslape de señales tal que es imposible asignarlas a un protón (o grupo de ellos, como un CH<sub>3</sub>, por ejemplo), condición imprescindible para determinar la estructura (tanto secundaria como terciaria) de la proteína. Además, incrementar el tamaño de las moléculas causa que se muevan lentamente, incrementando el tiempo de correlación y provocando que el tiempo de relajación transversal disminuya, mejorando la eficiencia de la relajación transversal.<sup>5,9,10,12</sup> Ampliar el tiempo de relajación transversal incrementa el ancho de las señales causando dos problemas: 1) el traslape de las señales empeora, dificultando aún más la asignación de las señales y 2) la eficiencia de la transferencia de información entre espines acoplados (escalarmente) se empobrece cuando el ancho de las señales de resonancia sobrepasa la magnitud de la constante de acoplamiento espín-espín. Esta situación cambia cuando se utilizan experimentos heteronucleares de 3 y 4 dimensiones (3D <sup>1</sup>H-<sup>15</sup>N TOCSY-HSQC, 3D <sup>1</sup>H-<sup>15</sup>N-NOESY-HSQC). Las proteínas con una mayor cantidad de residuos de aminoácidos (como el fragmento de la SV40 estudiado en el presente trabajo de 108 aminoácidos) son estudiadas cuando se obtienen marcadas uniforme o selectivamente con <sup>13</sup>C, <sup>15</sup>N o con ambos. Estas marcas selectivas favorecen la aplicación de experimentos en 3 ó 4 dimensiones que aprovechan las constantes de acoplamiento heteronucleares a un enlace que mejora la transferencia de magnetización y que evita el traslape de señales.

La metodología para la asignación de las señales de RMN involucró los siguientes pasos: obtener los desplazamientos químicos de <sup>15</sup>N para ubicar los planos en los experimentos 3D, asignar las franjas a un grupo de aminoácidos, encontrar las correlaciones que nos permitan

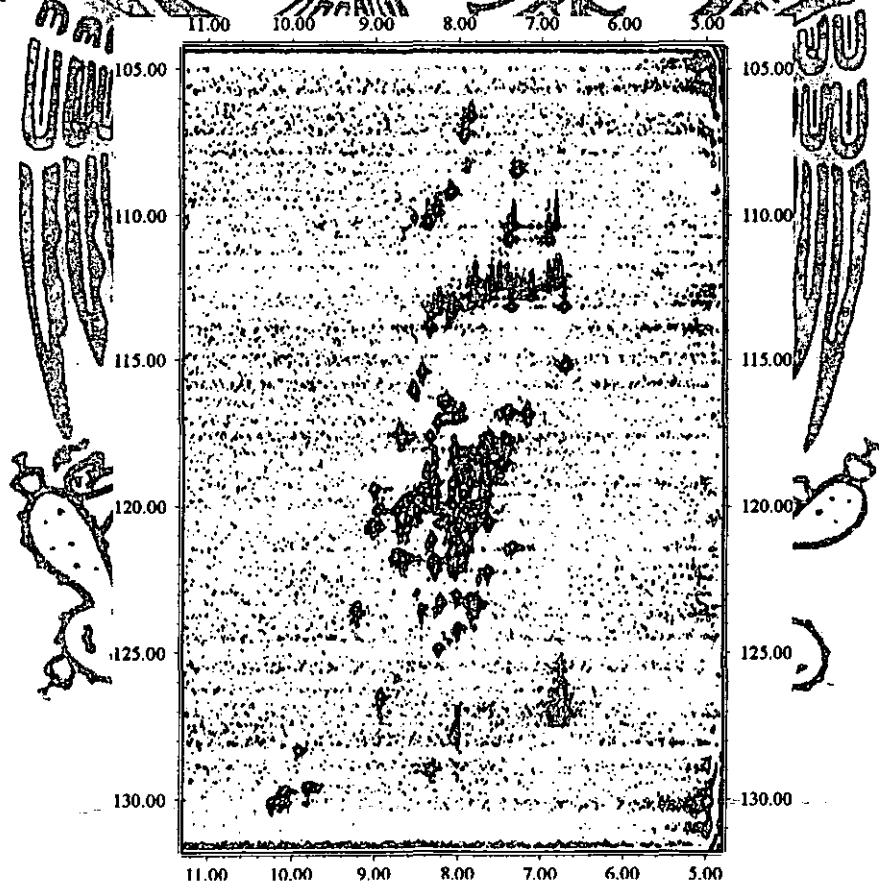
ordenar las franjas en cadenas, hacer la asignación de cada franja a un aminoácido dentro de la secuencia primaria de la proteína, encontrar las señales de correlación de larga distancia, determinar las regiones con estructura secundaria definida y conjuntar todos los datos obtenidos en restricciones adecuadas para calcular la estructura de la proteína.

En la *figura 4-1* se muestra el espectro  $2D\ ^1H\text{-}^{15}N\text{-HSQC}$  del que se obtuvieron los desplazamientos químicos de  $^{15}N$  para ubicar las señales en el espectro  $3D\ ^1H\text{-}^{15}N\text{-TOCSY-HSQC}$ . En la *figura 4-2* se muestran franjas representativas de este espectro; observese que las primeras 4 franjas corresponden a aminoácidos relativamente fáciles de asignar (Ala, Thr), su sistema de espines es único, por lo que su asignación no representa mayores complicaciones. En esta figura también se observa que las 4 siguientes franjas son muy similares entre sí, pero corresponden a aminoácidos distintos. Este hecho de que aminoácidos distintos presenten espectros similares es lo que dificulta la asignación; por lo que sólo suponemos que la franja es debida a algún aminoácido de cierto grupo. Además de estas asignaciones provisionales, lo único que podemos inferir es el tipo de protones; es decir, el desplazamiento químico de la señal indica que se trata de un protón alfa ( $H^a$ , si se encuentra alrededor de 4.3 - 4.8), beta ( $H^b$ , si se encuentra alrededor de 2.8 - 3.8 ppm), de un protón gamma ( $H^g$ , de valina o isoleucina si se encuentra alrededor de 0.9 - 1.4 ppm) o  $H^d$  de isoleucina o leucina (si se encuentra alrededor de 1.4 - 1.8 ppm).

Así se asignaron 103 señales de HN; 4 aminoácidos presentes en la SV40 son prolinas que carecen de protón de amida (HN), no mostrando señal en los experimentos 3D. No fue posible encontrar la señal de lo que se determinaría como la asparagina-56. Además se asignaron cerca de 580 señales de correlación intrarresiduo ( $H^a\text{-}H^a$ ,  $H^b\text{-}H^b$ ,  $H^g\text{-}H^g$  o  $H^d\text{-}H^d$ ).

En la primera parte del trabajo sólo se asignaron algunas franjas a un aminoácido particular (algunas alaninas y treoninas) y principalmente se asignaron las señales a los diferentes tipos de protones, además de que a cada franja se asignó a un grupo de aminoácidos. A continuación se procede a asignar secuencialmente las franjas en cadenas utilizando el experimento  $3D\ ^1H\text{-}^{15}N\text{-NOESY-HSQC}$ ; la distancia entre los protones  $H^N$ ,  $H^a$ , y  $H^b$  de un aminoácido  $i$  y el protón  $H^N$  del aminoácido siguiente ( $i+1$ ) siempre es menor a 5 Å, entre ellos se presentarán señales de correlación NOE (*figura 4-3*) que utilizamos para organizarlas

secuencialmente a partir del amino terminal, progresivamente hacia el carboxilo terminal y con las franjas en orden {amino terminal- $H_2N$ ,  $H_1$ ,  $H_2$ ,  $H_3$ ,  $H_4$ , ...,  $H_n$ -COOH-carboxilo terminal} utilizamos la estructura primaria de la proteína para asignar las franjas a un aminoácido específico. En la *figura 4-1* se muestra la región de 4-8 ppm (regiones donde aparecen las señales de  $H^\alpha$  y  $H_N$ ) de 8 franjas del espectro 2D  $^1H$ - $^{15}N$ -HSQC en las que se observa que la señal de  $H^\alpha$  siempre aparece en la franja de  $H_{i+1}N$ , nunca en la franja de  $H_iN$ , si ordenamos las franjas en cadenas para utilizar la estructura primaria de la proteína y continuar el proceso de asignación, pero ahora trabajando en la identificación de las franjas para reconocer el residuo al que pertenece.



**Figura 4-1.-** Espectro 2D  $^1H$ - $^{15}N$ -HSQC. En él se aprecia que la dispersión de las señales en la dimensión de  $^{15}N$  abarca de alrededor de 105 ppm hasta cerca de 130 ppm (aproximadamente 25 ppm); en contraste, la dispersión en la dimensión de  $^1H$  abarca sólo 3 ppm: de 7 hasta aproximadamente 10 ppm. Por ello las señales en la dimensión de nitrógeno presentan menos traslape, mientras que en la dimensión de hidrógeno hay más señales en el mismo desplazamiento químico. Esta es la gran ventaja de los experimentos heteronucleares multidimensionales: la utilización de los desplazamientos químicos del heteronúcleo evita el traslape de las señales.

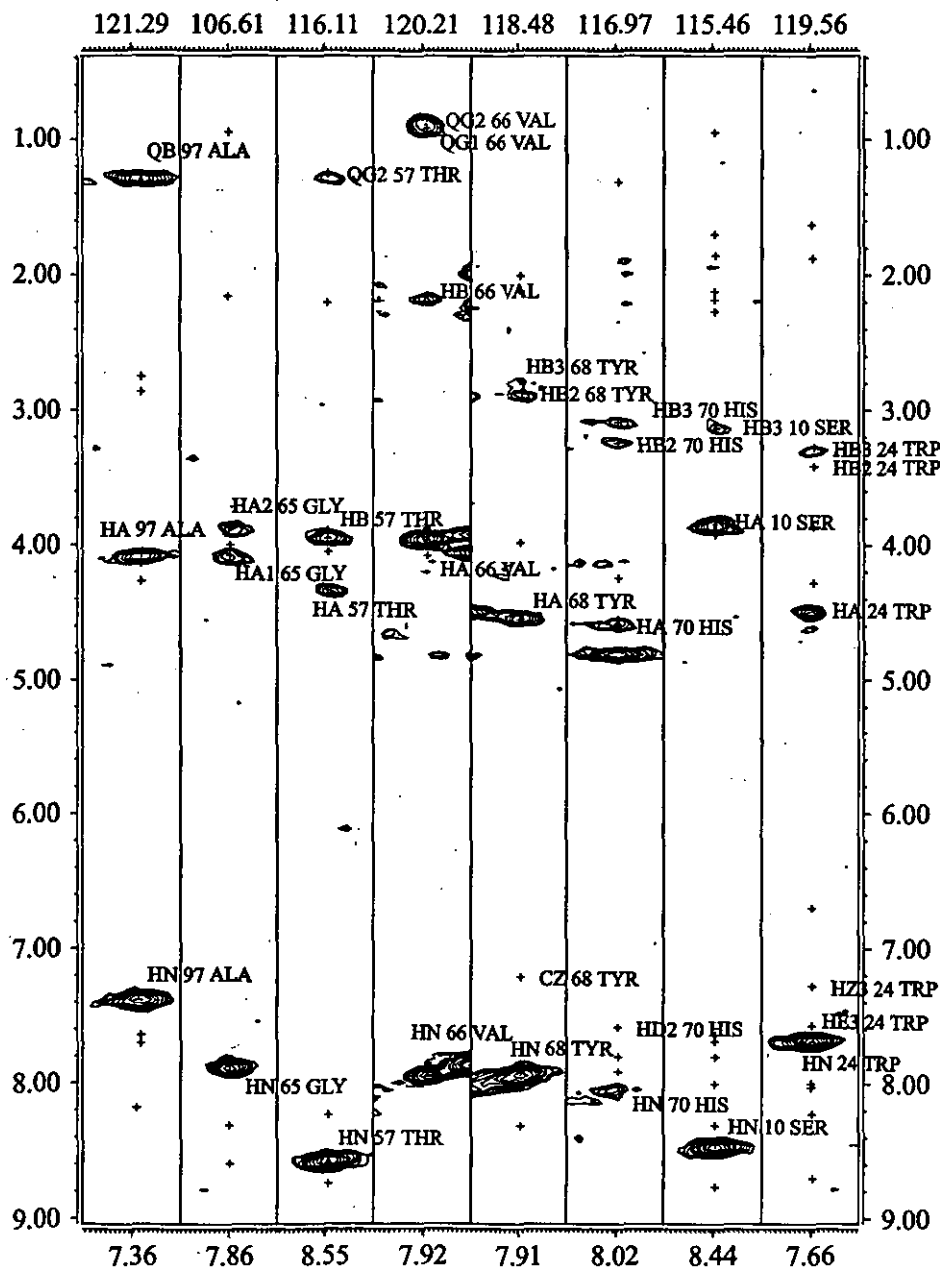


Figura 4-2.- Se muestran 8 franjas del experimento 3D  $^1\text{H}$ - $^{15}\text{N}$ -TOCSY-HSQC, que corresponden a: 1) Ala 97, 2) Gly 65, 3) Thr 57, 4) Val 66, 5) Tyr 68, 6) His 70, 7) Ser 10 y 8) Trp 24. Estas franjas son cortes de diversos planos del espectro 3D colocados al mismo nivel. Las primeras cuatro franjas son relativamente fáciles de asignar, los sistemas de espín únicos hace posible una identificación sin muchos problemas (por el número de señales y el desplazamiento químico de cada una de ellas). Las siguientes 4 corresponden a aminoácidos de un mismo grupo que no son fáciles de asignar. En la parte superior se observa el desplazamiento químico de  $^{15}\text{N}$  de cada una de las franjas, a los lados y en la parte inferior se observan los desplazamientos químicos de  $^1\text{H}$ .

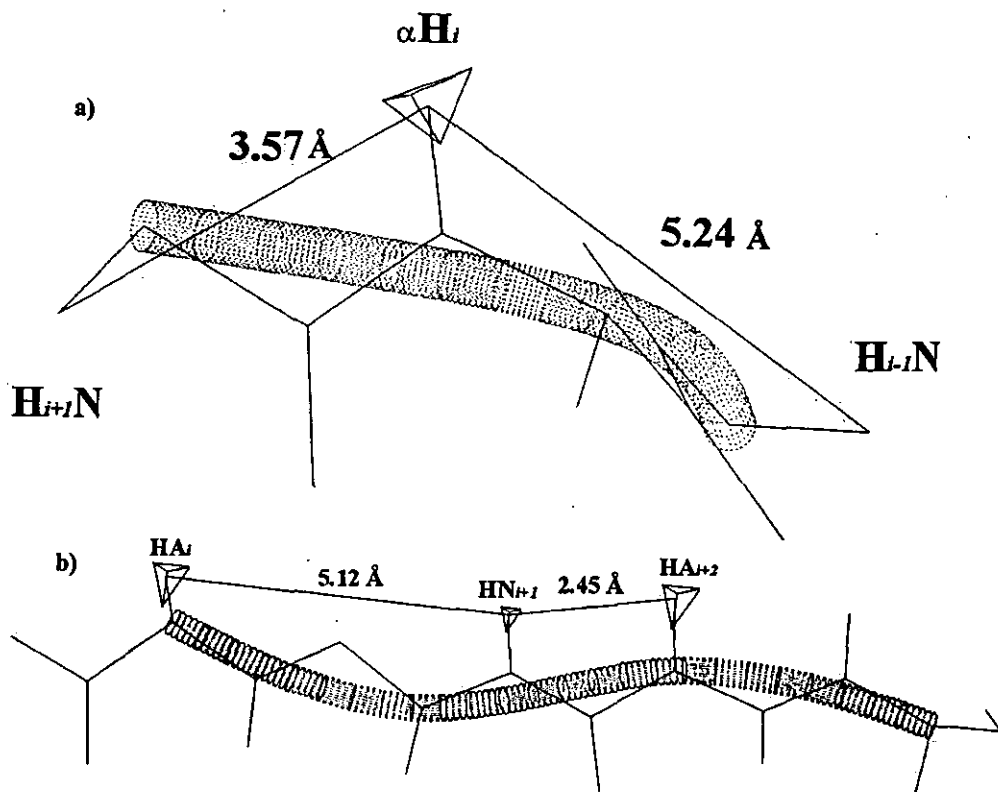
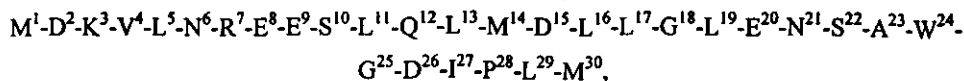


Figura 4-3.- Esquema en el que se muestran las distancias promedio entre los diferentes protones dentro de las dos estructuras secundarias más comunes a)  $\alpha$  hélice y b) hoja plegada  $\beta$ . En ambos casos puede observarse que la distancia  $\text{H}_i\text{-H}_{i+1}\text{N}$  es menor a 5 Å (presentará señal de correlación en los espectros NOE), mientras que la distancia  $\alpha\text{H}_i\text{-H}_{i-1}\text{N}$  es mayor a 5 Å, distancia límite a la que aparecen las señales de correlación NOE.

Para asignar secuencialmente las franjas a un aminoácido particular utilizamos las cadenas que previamente hicimos con ayuda del experimento 3D  $^1\text{H}\text{-}^{15}\text{N}\text{-NOESY}\text{-HSQC}$ ; para determinar la estructura terciaria de las proteínas mediante RMN es requisito indispensable que previamente conozcamos la estructura primaria de la proteína en estudio. Por ejemplo, asignemos secuencialmente una cadena de la que desconocemos su ubicación dentro del fragmento de proteína compuesta por los siguientes 30 aminoácidos:



los números indican su posición dentro de la estructura primaria; tomemos la cadena correspondiente a los 10 primeros aminoácidos:



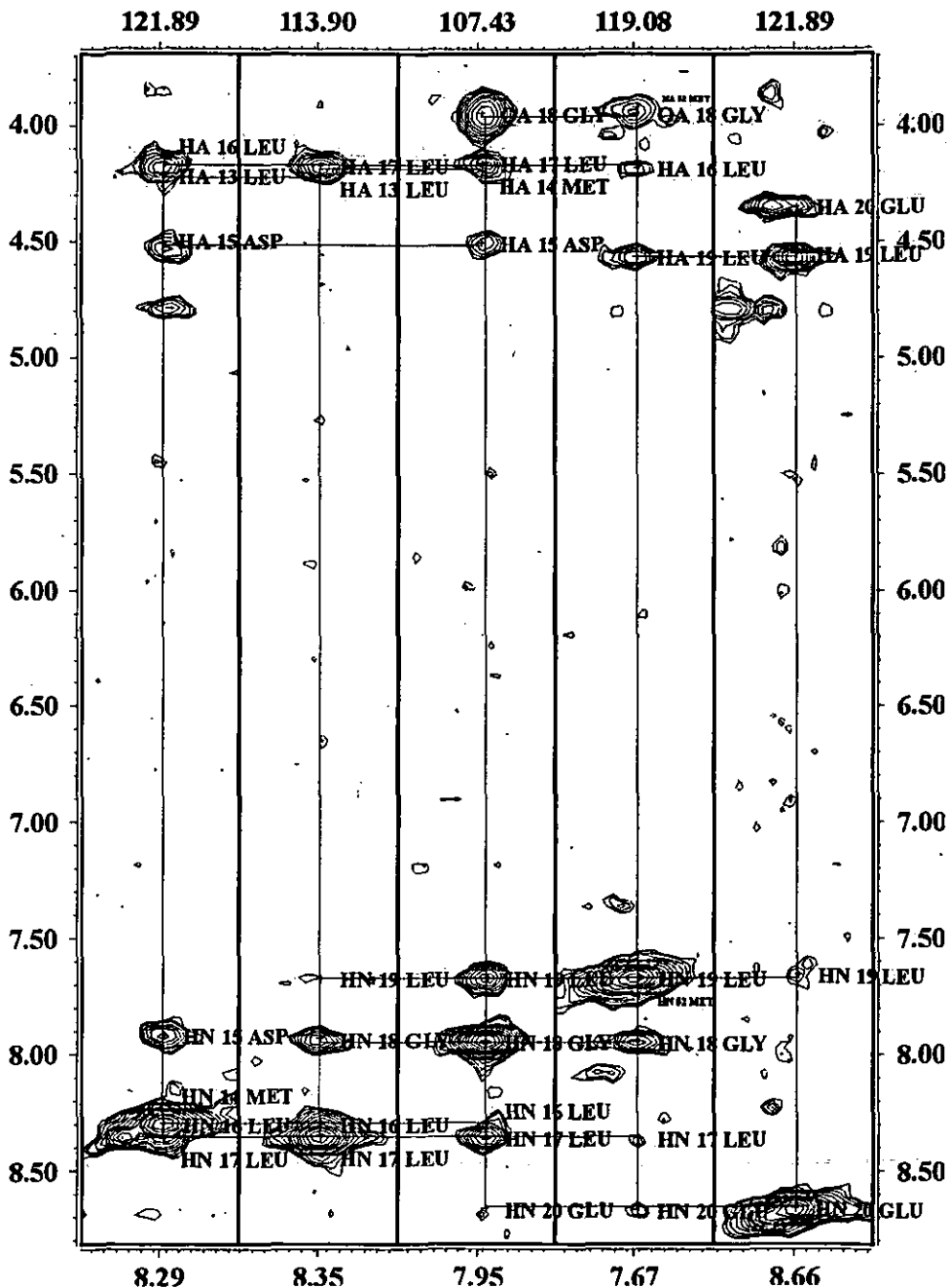


Figura 4-4.- Región del espectro 3D  $^1\text{H}$ - $^{15}\text{N}$ -NOESY-HSQC de 4 ppm a 8 ppm donde aparecen las señales de correlación utilizadas en la secuenciación de las franjas. Las señales más importantes para este fin son las señales de correlación  $^1\text{H}_i\text{-}^1\text{H}_{i+1}\text{N}$  y  $^1\text{H}_i\text{-}^1\text{H}_{i+1}\text{N}$ .

## M-D-K-V-L-N-R-E-E-S,

asignada provisionalmente (con ayuda del experimento 3D  $^1\text{H-}^{15}\text{N-TOCSY-HSQC}$ ) a los siguientes aminoácidos o grupos de ellos:

{QEM}-{DNCSHFVY}-K-V-{L}-{DNCSHFVY}-R-{QEM}-{QEM}-{DNCSHFVY}

entre corchetes están agrupados los aminoácidos que comparten sistema de espines similares; elegimos cualquier franja para iniciar el trabajo, supongamos que iniciamos con la primera de esta cadena seleccionada (*figura 4.5, pág. 53*): {QEM}. Comenzamos suponiendo que la franja es debida al aminoácido Q (Gln); revisamos la estructura primaria buscando todas las posiciones donde se presenta este aminoácido, en nuestro ejemplo sólo se presenta en la posición 12. En la estructura primaria se observa que la glutamina 12 ( $\text{Q}^{12}$ ) está precedida por la lisina 11 ( $\text{L}^{11}$ ) y después está la lisina 13 ( $\text{L}^{13}$ ). Revisando los espectros observamos que en la cadena, después de la franja de " $\text{Q}^{12}$ ", la franja no muestra un sistema de espines parecido al de la lisina, sistema compuesto por 6 señales, sino que en la franja sólo hay 2 señales. Debido a que el sistema de espines no coincide con las señales presentes en la segunda franja observada, descartamos que el fragmento corresponda a la región de 12. Ante ello, cambiamos y suponemos que la primera franja elegida pertenece ahora a otro aminoácido del mismo grupo E (Glu). Nuevamente revisamos las posiciones de la estructura primaria que contienen a dicho aminoácido: 8, 9 y 20. Elegimos trabajar primero con el aminoácido 8 y observamos que a continuación está el ácido glutámico 9. Revisando la segunda franja de la cadena observamos que ésta contiene 2 señales y el Glu tiene un sistema de espines de 4 señales, por lo que esta cadena tampoco corresponde a esta región de la proteína. Ahora elegimos trabajar con el otro ácido glutámico: el aminoácido nueve. A continuación está la serina 10, cuyo sistema de espín coincide con la segunda franja de la cadena, ambas poseen dos señales en los desplazamientos químicos típicos. Ya que coinciden los dos primeros aminoácidos con las franjas correspondientes, observamos la siguiente franja que debe de coincidir con el tercer aminoácido de la secuencia elegida. En este caso se trata de la leucina 10, cuyo sistema de espines no coincide con el observado: hay el número de señales adecuado, pero no están en los desplazamientos químicos que corresponden al sistema de espines de la leucina, por lo que se desecha esta región y se estudia la última posición donde se presenta el ácido glutámico. Después del Glu 20 en la estructura primaria está la Asn 21, cuyo sistema de espines coincide con la segunda franja de la cadena. Sin embargo, nuevamente para el tercer

aminoácido no coincide el sistema de espines esperado con el que se muestra en la franja: el aminoácido 22 es una alanina, con un sistema de espines de 2 señales y la tercera franja muestra 6. Descartados los dos primeros aminoácidos del grupo continuamos con la metionina, la cual se haya presente en tres posiciones de nuestra proteína: en 1, en 14 y en 30. En este caso podemos descartar de principio que se trate de la metionina 30: nuestro fragmento sólo consta de 30 aminoácidos y la cadena muestra que hay más aminoácidos después de la metionina, por lo que no se puede tratar de ésta. Así, sólo se analizan los fragmentos de las metioninas 1 y 14. El segundo aminoácido de la secuencia es ácido aspártico cuyo sistema de espines coincide perfectamente con el que se presenta en la segunda franja de la cadena; examinamos el siguiente aminoácido, una lisina, y observamos en la tercera franja que el número de señales y los desplazamientos químicos coinciden con el sistema de espines de la lisina. Continuamos revisando y encontramos que el siguiente aminoácido (valina 4) coincide con sistema de espines de la franja 4. Si se revisan los siguientes aminoácidos se encontrará que en éstos siempre coincidirán los sistemas con las señales que se muestran en las franjas, logrando asignar estas franjas a los primeros 4 aminoácidos de la secuencia.

En el ejemplo anterior se expuso el método que se siguió para la asignación secuencial de las franjas dentro de una cadena cualquiera; esta estrategia se simplifica si comenzamos la asignación con franjas de aminoácidos con sistema de espín único: son característicos los sistemas de la glicina (G), alanina (A), arginina (R), treonina (T), lisina (K) y valina (V). La asignación del fragmento anterior se simplificaría mucho si se inicia la asignación por la lisina 3, valina 4, arginina 7, glicina 18, 25 ó alanina 23; y en este caso particular el trabajo se simplifica más por que hay dos aminoácidos de sistema de espín característico (la lisina 3 y la valina 4) contiguos, esos dos aminoácidos son el punto de partida para la asignación, fácilmente se logra la identificación de dos de los tres aminoácidos con los que comenzamos la asignación secuencial.

Cuando ya tenemos asignadas tres o cuatro franjas consecutivas se facilita el trabajo, por que sólo hay que seguir la secuencia primaria para asignar las siguientes franjas y únicamente tenemos que verificar que el número y posición de las señales concuerden con el sistema de espines del aminoácido. Lo anterior se fundamenta en teoría de permutaciones: la probabilidad de que un arreglo ordenado de 3 objetos (por ejemplo, la siguiente secuencia de aminoácidos: SLQ) suceda, teniendo 20 posibles objetos (aminoácidos), es de 1 en 6840. La probabilidad para que se presente un arreglo de 4 objetos (del mismo grupo de 20 posibilidades) es de 1 en 116280. Esto

quiere decir que al identificar una secuencia de 3 aminoácidos dentro de una cadena, tenemos una certidumbre del 99.985 %; si la secuencia identificada es de 4 tendremos una certidumbre del 99.999 %.<sup>5</sup>

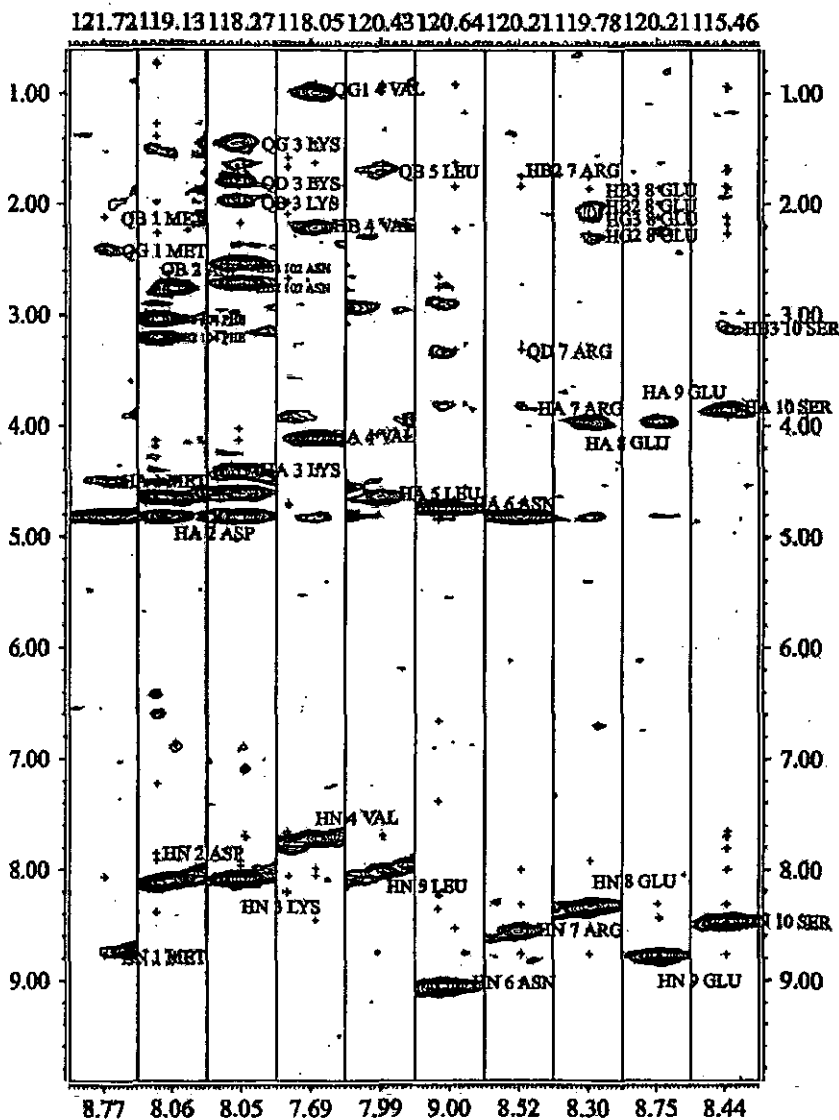


Figura 4.5.- Se muestra una cadena con 10 franjas del experimento 3D <sup>1</sup>H-<sup>15</sup>N-TOCSY-HSQC, correspondientes a la secuencia M<sup>1</sup>-D<sup>2</sup>-K<sup>3</sup>-V<sup>4</sup>-L<sup>5</sup>-N<sup>6</sup>-R<sup>7</sup>-E<sup>8</sup>-E<sup>9</sup>-S<sup>10</sup>. La asignación secuencial de cada una de estas franjas es un proceso cíclico en el que se comparan las señales de una franja con el sistema de espín de un aminoácido de la secuencia; si el sistema de espín coincide con el número de señales presentes en la franja, se compara la siguiente franja con el aminoácido que está a continuación en la secuencia; si no coinciden, se elige otro aminoácido de la secuencia primaria y se repite el proceso de comparación. Esto se continúa hasta que las señales de las franjas coinciden con los sistemas de espín de los aminoácidos de una secuencia.

Con la asignación de las señales se definen las regiones de estructura secundaria ( $\alpha$  hélices, hojas plegadas  $\beta$  ó regiones de estructura aleatoria) y los aminoácidos que pertenecen a cada una de éstas.<sup>18,19</sup> El desplazamiento químico es un fenómeno que depende esencialmente del ambiente químico que rodea a los núcleos y la existencia de grandes diferencias en densidad electrónica entre las 3 principales estructuras locales que conforman las proteínas, fundamenta el uso de este importante factor en la asignación de las regiones de estructura secundaria. Uno de los protones más sensibles a modificar su desplazamiento químico en función de las diversas regiones de estructura secundaria es el protón  $\alpha$ , cuyo valor en regiones de estructura aleatoria se toma como referencia. El arreglo espacial de las  $\alpha$  hélices ejercen un efecto de protección magnética sobre los  $H^\alpha$  afectando los desplazamientos químicos, trasladándolas hacia un campo más alto. Sucede lo contrario en una hoja plegada  $\beta$ , donde ésta ejerce un efecto de desprotección magnética que desplazará las señales hacia campo bajo. Estas modificaciones de los desplazamientos químicos de los valores característicos de una región de estructura aleatoria, se ven claramente en una gráfica de índices de protón donde se observan dichas desviaciones a lo largo de la secuencia de la proteína. Esta se construye como se indicó en la sección de *Desarrollo Experimental* (figura 3.4).

Las diferentes regiones de estructura secundaria se deducen cuando se presenta un grupo continuo de al menos cuatro índices idénticos. Las regiones de la proteína donde se presentan 4 índices  $-1$  continuos, indican que estos aminoácidos forman parte de una  $\alpha$  hélice; las hojas plegadas  $\beta$  se deducen cuando se presentan regiones con 3 índices  $+1$  continuos. En la SV40 (1-108) no observamos este tipo de regiones, suponemos que esta proteína no posee regiones de hoja plegada- $\beta$ . En la gráfica se observan regiones continuas de índices  $-1$ , donde es clara la presencia de estructuras de hélice; sin embargo, en algunas regiones hay unas zonas discontinuas dentro de la tendencia de índices, asignando las estructuras helicoidales de manera continua. Ello se debe a que en la gráfica de índice de protón se asigna un índice de  $\pm 1$  a los desplazamientos químicos que exceden el intervalo de  $\pm 0.1$  ppm del valor estándar; sin embargo, esto no toma en cuenta la tendencia del desplazamiento químico que se acerca al límite establecido, sin sobrepasarlo; supondríamos con ello que la estructura regular termina algunos aminoácidos antes, como puede observarse en la comparación de las gráficas (figura 4-6). Pese a este pequeño inconveniente, es preferible utilizar la gráfica "indexada" porque en ella se observan claramente las tendencias de los índices de protón, facilitando la delimitación de las regiones de estructura secundaria. De esta

manera encontramos que la proteína SV40 en la región de 1-108 está formada por 4  $\alpha$  hélices que van de los residuos 9 - 19, 24 - 41, 49 - 71 y 89 - 103.

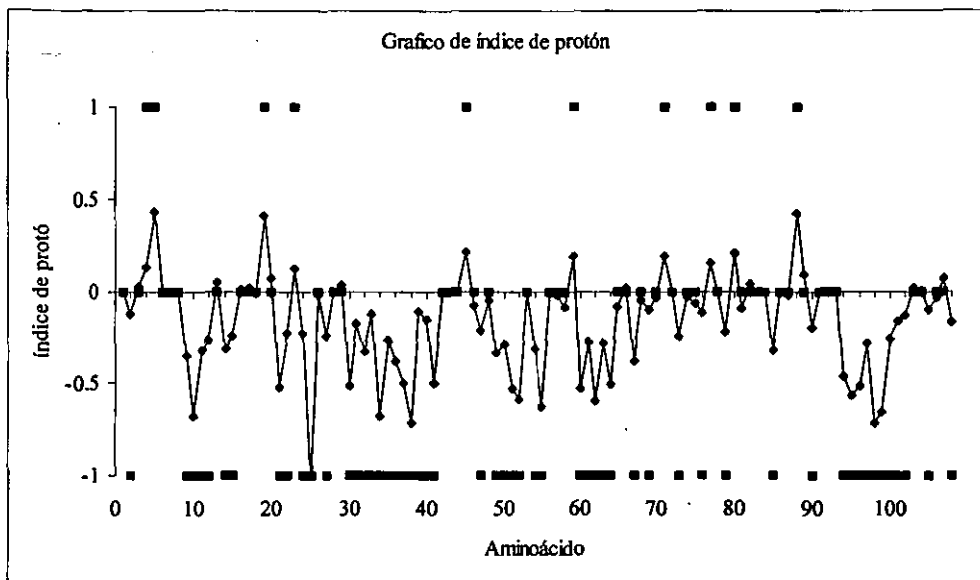


Figura 4-6.- Comparación de los desplazamientos químicos del  $\alpha$  hélice con y sin índice. En ésta gráfica se muestra que los índices de protón no toman en "cuenta" la tendencia del desplazamiento químico y por ello no toma en cuenta algunos residuos que pertenecen al elemento regula de estructura secundaria.

En adición a los índices de protón, se utilizó la intensidad de las señales de correlación NOE de corta distancia para definir con mayor precisión los residuos que pertenecen a las  $\alpha$ -hélices, además de que con ello confirmamos que se trataba de este elemento de estructura secundaria. Observamos la presencia y la intensidad de las señales como se indica en la literatura:<sup>5,7,9,10</sup> señales de correlación muy intensas  $H_iN-H_{i+1}N$  y señales de intensidades mediana o débil  ${}^{\alpha}H_i-H_{i+1}N$  (generalmente mediana),  ${}^{\alpha}H_i-H_{i+2}N$  y  ${}^{\alpha}H_i-H_{i+4}N$  (débiles) y  ${}^{\alpha}H_i-H_{i+3}N$  (muy débiles). Con estos dos datos no hubo lugar a dudas en cuanto a la asignación de las hélices.

La última fase del proceso de obtención de la estructura terciaria de una proteína es la identificación de las correlaciones de larga distancia, que contribuyen principalmente en la determinación de la estructura terciaria. Para ello es importante la identificación de las regiones de estructura secundaria: conocer los aminoácidos que forman una región con estructura definida nos permite suponer que al encontrar una señal de correlación dentro de esta región, los aminoácidos adyacentes mostrarán correlaciones hacia la misma región o zonas muy cercanas;

las señales de correlación aparecerán regularmente, en patrones característicos, permitiéndonos deducir regiones de estructura secundaria que se encuentran próximas, dirigiendo la búsqueda de señales a las zonas adyacentes a las regiones donde encontramos la señal de correlación (figura 4-7).

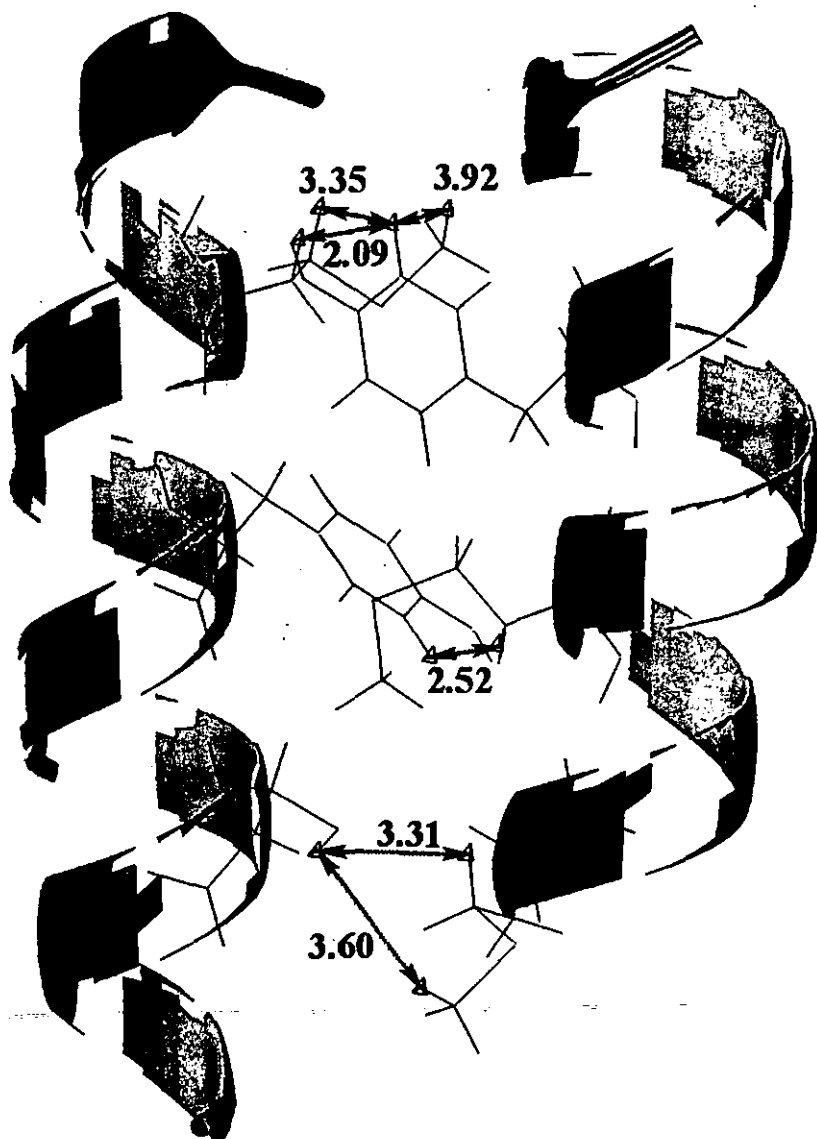


Figura 4-7.- Esquema donde muestra el orden de las regiones de estructura secundaria el cual provoca que se encuentren señales de correlación con regularidad. Obsérvese que los aminoácidos de una hélice próximos a otra muestra una regularidad en su cercanía, por lo que se espera que las señales NOE también aparezcan con regularidad.

Para encontrar las señales de correlación de larga distancia, buscamos en diferentes planos del espectro 3D (figura 4-8). En los experimentos 2D cuando dos protones correlacionan, las señales forman un cuadrado; en los experimentos 3D cuando dos protones correlacionan no es tan sencillo encontrar dichas señales. Supongamos el caso de una glicina próxima en espacio a una alanina, lejanos en la secuencia. La franja de la glicina se encuentra en el desplazamiento químico  $^{15}\text{N}_G$ , correspondiente al plano X y la alanina se encuentra en el desplazamiento químico de  $^{15}\text{N}_A$ , que corresponde al plano Y. En cada uno de los planos se muestran las señales del sistema de espín característico de cada aminoácido y las señales debidas a la correlación espacial entre ellos (figura 3.2). Puede observarse en las franjas que las señales del espectro NOESY incluyen las del propio aminoácido y las del aminoácido con quien se correlaciona; por ello primero deben identificarse las señales del sistema de espín con ayuda del TOCSY y las señales remanentes se asignan como correlaciones de larga distancia. En la franja de la glicina se observan las señales de este aminoácido, las correlaciones secuenciales y tres señales adicionales, las correlaciones con los protones de la alanina. El procedimiento para encontrar los aminoácidos que correlacionan utilizando los subprogramas de Xeasy, se describió en la sección de Desarrollo Experimental (figura 3.3). Idealmente se espera que en las franjas aparezcan señales de correlación entre todos los protones de los aminoácidos, pero no sucede así, por que no todos los protones se encuentran a la distancia requerida para que la señal de NOE sea apreciable (menos de 5 Å); sin embargo, son muchos los protones que se encuentran a la distancia necesaria para observar la mayoría de las señales de correlación.

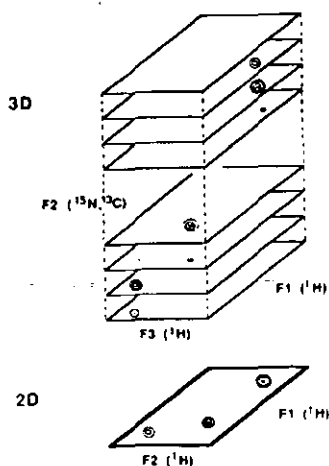


Figura 4-8.- Los experimentos 3D vistos como un conjunto de espectros 2D (planos) que se extienden a lo largo del desplazamiento químico del heteroátomo ( $^{15}\text{N}$  ó  $^{13}\text{C}$ ).



La zona que presentó mayor dificultad en asignar fue la que comprenden los aminoácidos 75-89, que muestra muy pocas señales. No fue posible determinar la razón por la que dicha región no muestra ni siquiera señales de correlación intrarresiduo; creemos que es una zona que carece de estructura secundaria definida, cambiando constantemente de conformación que no permite una transferencia de magnetización adecuada para la observación de señales. Sin embargo, los espectros con los que contamos no son concluyentes al respecto, son necesarios más experimentos para poder concluir si esta región tiene una estructura secundaria definida o no, además de si correlaciona con algún otra región de la proteína.

El parámetro esencial en la determinación de la estructura terciaria de las proteínas son las restricciones obtenidas del experimento NOE, que al no poderse obtener con precisión, se agrupan en tres categorías principales y a cada una se le asignan valores límite superior e inferior: señales muy intensas 1.8-2.8 Å, señales con intensidad intermedia 1.8-3.5 Å o si la señal es débil 1.8-5 Å. A pesar de estas categorías, está plenamente probado que la resolución de la estructura terciaria esta dada por la cantidad de restricciones:<sup>9</sup> más que la precisión de las restricciones, el número de éstas es la que mejora la resolución de la estructura. Por ello clasificar cualitativamente en tres grupos las distancias no afecta el proceso de obtención de la estructura.

En el caso de las constantes de acoplamiento sucede algo muy distinto. Es muy importante medir las constantes con la mayor precisión posible, un leve cambio en la constante modifica bruscamente el ángulo diedro; una gran cantidad de constantes de acoplamiento nos proporcionará estructuras con el mayor refinamiento posible. Es factible medir las constantes de acoplamiento con una gran precisión,<sup>20</sup> pero para ello se requieren de espectros que tengan la mayor resolución espectral posible, requiriendo éstos de un gran tiempo de equipo.

Durante los primeros cálculos de las estructuras únicamente se utilizaron restricciones de distancia (NOE) y puentes de hidrógeno (que estabilizan las hélices y mantiene su conformación). Estos cálculos iniciales proporcionaban estructuras con hélices mal plegadas. Este problema se resolvió utilizando valores de ángulo diedro reportados, basándonos en los resultados de las gráficas de índice de protón que muestran las regiones de estructura secundaria, que en el caso de la SV40 son  $\alpha$  hélices.<sup>9, 10, 14</sup>

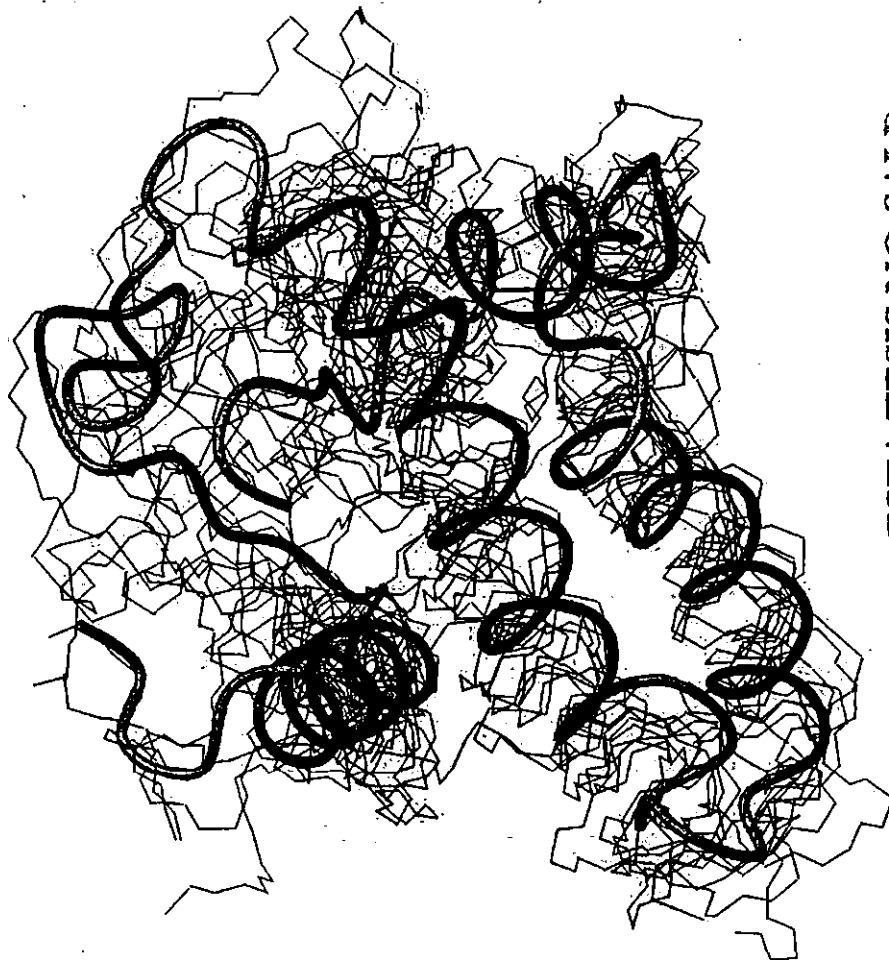
La mayoría de las restricciones utilizadas para calcular la estructura terciaria de las proteínas provienen de los datos de NOE y representan la distancia límite a la que se pueden encontrar dos protones; la distancia mínima entre dos protones se toma como la suma de los radios de van der Waals. Adicionalmente a las restricciones de distancia, se introducen a los cálculos restricciones de ángulos diedros y restricciones de puentes de hidrógeno.

Se han desarrollado una gran variedad de métodos para calcular las estructuras de proteínas a partir de las restricciones obtenidas por RMN; el objetivo de todos estos programas es determinar las posiciones de los átomos de la proteína que satisfagan las restricciones obtenidas de los diferentes experimentos de RMN. Los datos de RMN no definen una estructura única, debido a que sólo se conoce un pequeño subconjunto de restricciones (no se pueden determinar todas las distancias, el NOE está limitado a distancias menores a 5 Å) y estas distancias representan diferentes intervalos de valores permitidos; además, sólo se conocen algunos ángulos diedros y puentes de hidrógeno. Por ello, el cálculo es repetido múltiples ocasiones para obtener un conjunto de estructuras (de baja energía) consistentes con los datos experimentales. Un "buen" conjunto de estructuras minimiza las violaciones a las restricciones y la raíz cuadrada de la desviación media (RMSD), entre los miembros del conjunto.<sup>5,6,7,8,9,10,21,22</sup>

A veces el RMSD global no es una buena medida de la precisión de la estructura calculada, ya que éste no es capaz de discriminar entre estructuras que en promedio son pobremente reproducidas de estructuras reproducidas con precisión, excepto en regiones que no poseen estructura secundaria definida (debida al movimiento rápido). En algunas ocasiones es preferible medir el RMSD de amplias regiones o segmentos, tratando de abarcar la mayor parte de la estructura. Sólo cuando se tiene un gran conjunto de datos experimentales (10 restricciones o más por residuo, que en una proteína de 100 aminoácidos implica cerca de 1000 señales asignadas) puede definirse con precisión toda la estructura de la proteína y en estos casos es útil obtener el RMSD global.

En la *figura 4-9* se muestra la superposición del esqueleto de las 10 estructuras con menor energía, calculadas a partir de los datos de distancia (NOE), ángulos diedros y puentes de hidrógeno. En esta figura puede observarse la estructura general de la proteína, que a pesar de las pocas restricciones se pliegan "uniformemente", reproduciendo una sola estructura. Las regiones con mayor dispersión son debidas a la falta de restricciones, el espectro no presenta muchas

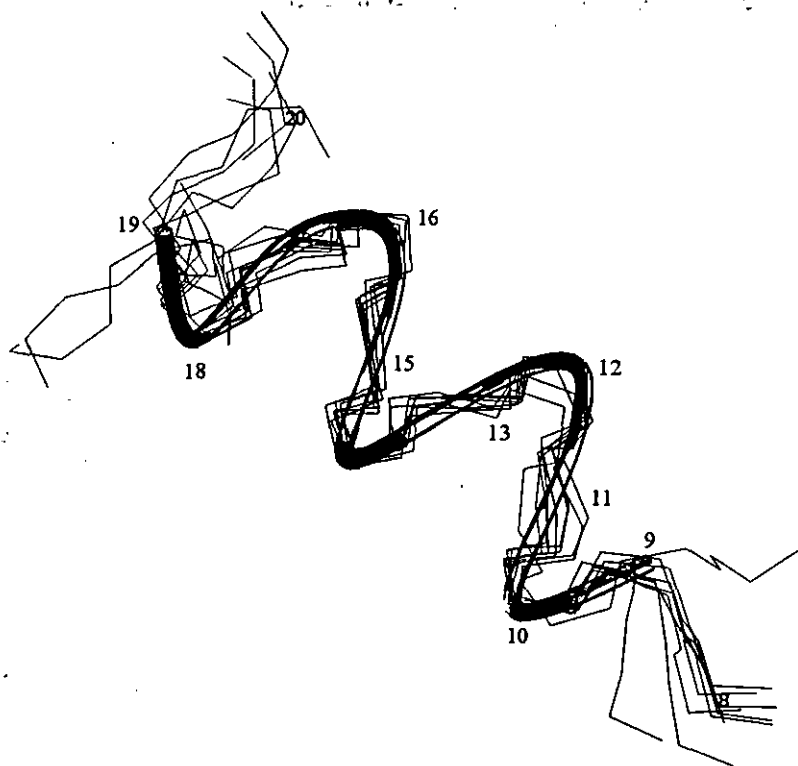
señales para los aminoácidos de esa zona, por lo que no fue posible obtener restricciones que determinen con precisión una estructura. El RMSD general es muy alto, producto de la gran dispersión entre estructuras calculadas con pocas restricciones; al determinar el RMSD de diversas regiones por separado, obtenemos valores muy bajos de RMSD, que indican un plegamiento adecuado de la estructura con las restricciones asignadas. La superposición de las estructuras, comparando ciertas regiones, ayudó a identificar fácilmente las restricciones que producen violaciones al cálculo, que fueron corregidas mediante la re-interpretación de los espectros.



ESTA TESIS NO SALE  
DE LA BIBLIOTECA

**Figura 4-9.-** Muestra la superposición del esqueleto de las 10 estructuras con menor energía. A pesar de que aún existen diferencias significativas entre algunas estructuras, éstas muestran una tendencia general de empaquetamiento de la estructura, en todos los casos ésta tiende a compactarse alrededor en dos largas hélices antiparalelas.

Así la *figura 4-10* representa la región formada por la  $\alpha$  hélice que abarca los aminoácidos 9-19 plegándose uniformemente en las 10 estructuras calculadas, el RMSD para esta región es de 0.460. En la *tabla 4-1* se presentan los RMSD para las diversas regiones de la proteína con estructura secundaria definida. En los casos de las hélices 9-19, 24-41 y 89-103 se tiene un RMSD aceptable, todos están por debajo de 1.000.

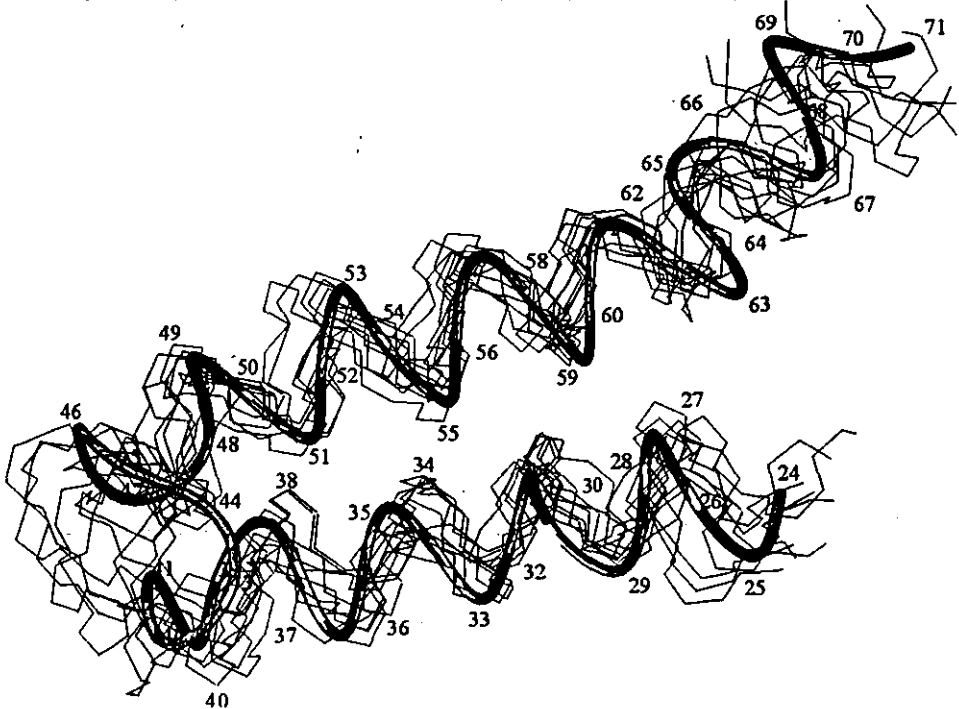


**Figura 4-10.-** Superposición del esqueleto de la región  $\alpha$  hélice comprendida entre los aminoácidos 9-19; el plegado de esta región en las 10 estructuras calculadas es uniforme, existiendo pocas diferencias entre éstas estructuras calculadas.

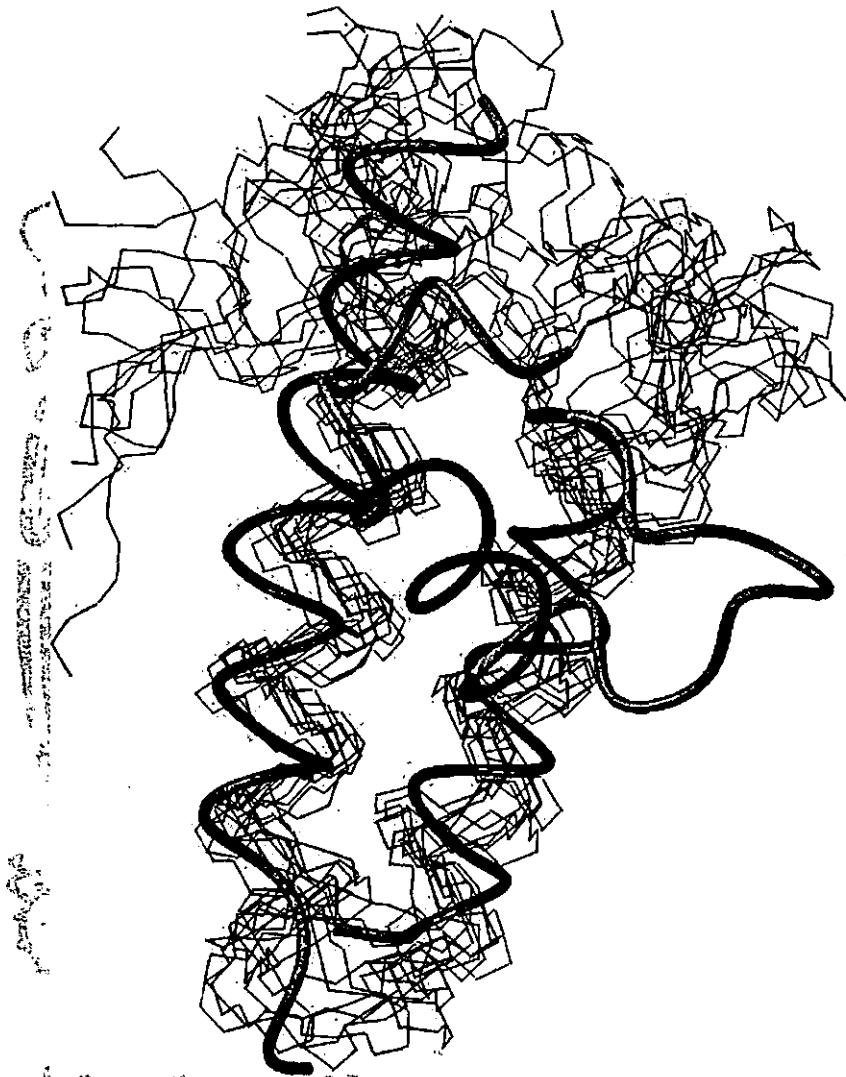
Región	RMSD
9 - 19	0.460
24 - 41	0.840
49 - 71	1.651
89 - 103	0.602
24 - 71	2.207
1 - 108	5.461

**Tabla 4-1.-** RMSD de la superposición del esqueleto de 10 estructuras calculadas con menor energía en las regiones con estructura secundaria definida.

La región más interesante es la comprendida entre los aminoácidos 24 y 71, es la región más extensa en la que se determinó una estructura secundaria definida, en esta zona se encuentra la mayor cantidad de señales en el espectro NOE, comprobando que una mayor cantidad de restricciones produce una mejor definición de la estructura. La región está formada por dos hélices antiparalelas: la primer hélice abarca del aminoácido 24 hasta el 41 (siendo la segunda hélice de la estructura general) y la siguiente hélice va del aminoácido 49 hasta el 71 (tercera dentro de la estructura, *figura 4-11*). En un trabajo anterior se trató de determinar la estructura terciaria fragmento 1-135 de la SV40, pero el grupo del Dr. Wagner sólo determinó la estructura de los residuos 1-71; encontramos que, a grandes rasgos, el fragmento 1-108 conserva la estructura del fragmento 1-71 (*figura 4-12*); por la cantidad de restricciones determinadas, la región 24-71 es la que presenta un mayor parecido en ambos fragmentos, calculados independientemente. De las 10 estructuras con menor energía, sólo 3 muestran que la región 1-23 posee una orientación distinta a la estructura determinada por el grupo de Harvard; en esta región no se asignaron muchas señales, por ello la estructura presenta mayor dispersión y una baja reproducibilidad. En general se observa que este fragmento presenta una estructura similar a la determinada por el grupo del Dr. Wagner.



**Figura 4-11.-** Superposición del esqueleto de las 10 mejores estructuras en la región de los aminoácidos 24-71, el plegamiento de esta zona muestra dos largas  $\alpha$  hélices antiparalelas.



**Figura 4-12.-** Superposición del esqueleto de las 10 mejores estructuras y de la estructura terciaria del fragmento 1-71 de la sv40, determinado por Han Jung. En general ambas estructuras son muy similares y sólo en 3 de las estructuras calculadas, el fragmento 1-21 posee una orientación distinta a la calculada para el fragmento 1-71.

Las regiones amino y carboxilo terminal generalmente carecen de una estructura secundaria definida, son zonas con una alta movilidad. En la proteína estudiada (SV40[1-108]) no hay señales de correlación de corta y/o larga distancia (NOE) en las regiones 1 - 8, 72 - 88 y 104 - 108 y son escasas las señales de correlación intrarresiduo (TOCSY); esta carencia de señales provoca que no se puedan definir regiones con estructura secundaria; se cree que la alta movilidad de estas regiones provoca que no se definan elementos regulares de estructura.

### Conclusiones

Basados en la gráfica de índice de protón y confirmando por la presencia de señales de correlación NOE de corta distancia  $H_iN - H_{i+1}N$  intensas,  ${}^{\alpha}H_i - H_{i+1}N$ ,  ${}^{\alpha}H_i - H_{i+2}N$ ,  ${}^{\alpha}H_i - H_{i+3}N$ ,  ${}^{\alpha}H_i - H_{i+4}N$  y  $H_iN - H_{i+2}N$ , de intensidades media o débil, se encontró que el fragmento de la proteína SV40 [1-108] está formado únicamente por  $\alpha$  hélices, determinando que los aminoácidos de dichas regiones de estructura secundaria son:

Hélice	Aminoácidos
I	9 - 19
II	24 - 41
III	49 - 71
IV	89 - 103

Se descarta la presencia de elementos regulares de hoja plegada- $\beta$ , ya que la gráfica de índice de protón no muestra la tendencia de desplazamientos químicos característicos, además de que no se presentan los patrones característicos de las señales de correlación NOE de corta distancia (carencia de señales de correlación  $H_iN - H_{i+1}N$  y señales  ${}^{\alpha}H_i - H_{i+1}N$  muy intensos).

El grupo del Prof. Wagner ha estudiado la región 1-135 de la proteína SV40, sin haber encontrado la estructura secundaria completa. Únicamente, pudieron encontrar el plegamiento del fragmento comprendido entre los residuos 1-71, el cual es similar al de la misma región del fragmento 1-108. Especialmente, en los residuos 24-71, las diferencias observadas entre estas proteínas se atribuyen al número escaso de restricciones que han sido asignadas en esa zona.

El RMSD del esqueleto de las regiones de estructura secundaria definida es bastante bajo, lo que indica que en estas zonas las restricciones son suficientes para ajustarlas; sin embargo, el RMSD de todo el esqueleto es elevado, quedando pendiente su refinamiento, lo cual sería un tema de tesis de un grado superior.

**Bibliografía**

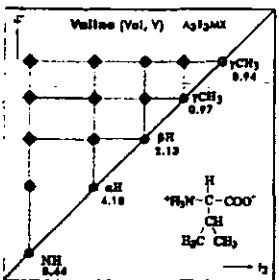
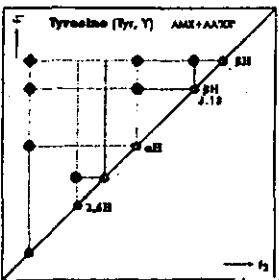
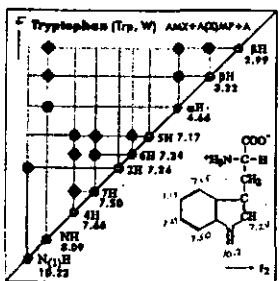
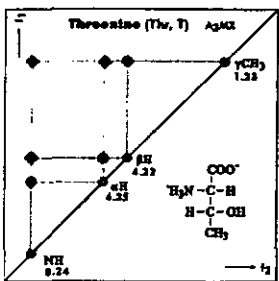
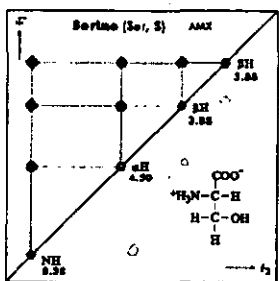
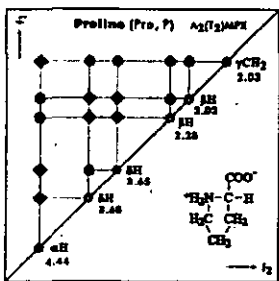
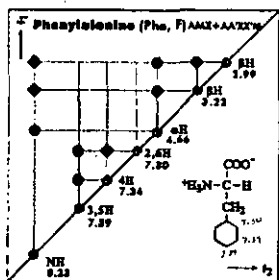
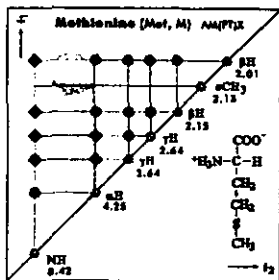
- 1) Kirk – Othmer. *Encyclopedia of Chemical Technology*. J. I. Kroschwitz. Wiley-Interscience, EE.UU., 1992. Vol 20, pp. 395 - 427.
- 2) <http://www.chimerx.com/sv40.htm>
- 3) <http://rex.nci.nih.gov/massmedia/pressreleases/simianvirus.html>
- 4) <http://www.nci.nih.gov>
- 5) *NMR in Drug Design*. D. J. Craik. CRC, EE.UU., 1996.
- 6) K. Wütrich. *Science*, 1989, 243, 45 - 50
- 7) M. Gronenborn, G. M. Clore. *Anal. Chem.*, 1990, 62, 2 - 15
- 8) M. Clore, A. M. Gronenborn. *Science*, 1991, 252, 1390 - 1399
- 9) *NMR of Protein and Nucleic Acids*. K. Wütrich. John Wiley & Sons, EE.UU., 1986.
- 10) *Biomolecular NMR Spectroscopy*. J. N. S. Evans. Oxford University Press, Oxford, 1995.
- 11) M. Eberstadt, G. Gemmecker, D. F. Mierke, H. Kessler. *Angew. Chem. Int. Ed. Engl.*, 1995, 34, 1671 - 1695
- 12) *Protein NMR Spectroscopy. Principles and Practice*. J. Cavanagh, W. J. Fairbrother, A. G. Palmer III, N. J. Skelton. Academic Press, Londres, 1995
- 13) <http://www.mol.biol.ethz.ch/wutrich/software/xeasy>
- 14) <http://cns.csb.yale.edu>
- 15) <http://www.mol.biol.ethz.ch/wutrich/software/molmol>
- 16) L. Szilágyi, O. Jardetzky. *J. Magn. Res.*, 1989, 83, 441 - 449
- 17) A. Pastore, V. Saudek. *J. Magn. Res.*, 1990, 90, 165 - 176
- 18) D. S. Wishart, B. D. Sykes, F. M. Richards. *Biochemistry*, 1992, vol. 31, No. 6, 1647 - 1651
- 19) D. S. Wishart, B. D. Sykes, F. M. Richards. *J. Mol. Biol.*, 1991, 222, 311 - 333
- 20) A. Garza-García, G. Ponzanelli, F. Del Río-Portilla. *J. Magn. Res.*, 2000, 148, 214 - 219

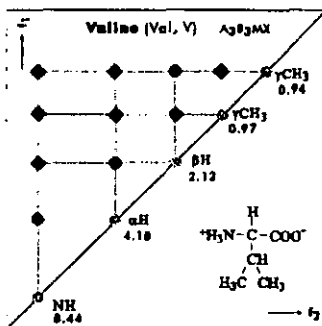
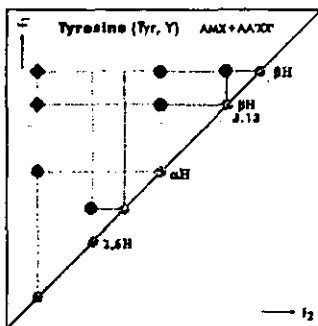
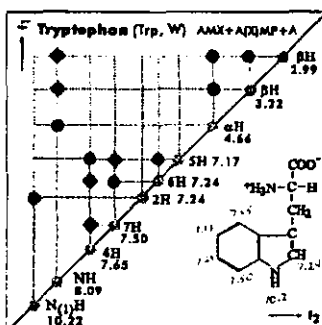
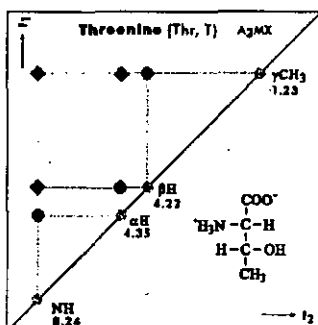
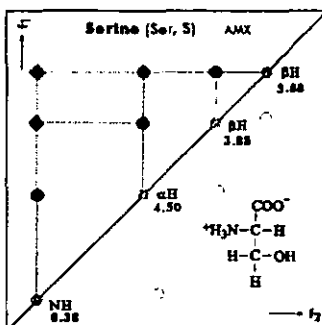
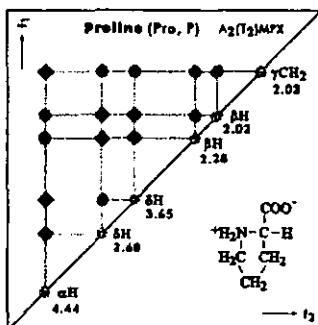
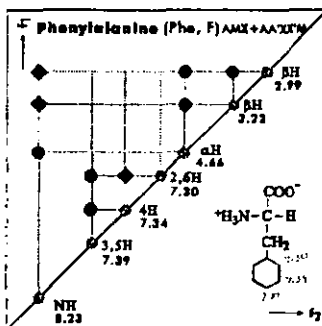
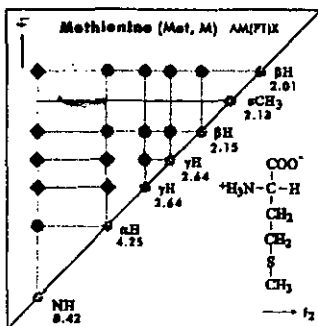


- 21) T. Brünger, G. M. Clore, A. M. Gronenborn, R. Saffrich, M. Nilges. *Science*, 1993, 261, 328 - 331.
- 22) G. Hyberts, M. S. Goldberg, T. F. Havel, G. Wagner. *Prot. Sci.*, 1992, 1, 736 - 751.
- 23) M. Billeter, W. Braun, K. Wütrich. *J. Mol. Biol.*, 1982, 155, 321 - 346
- 24) D. Marion, L. E. Kay, S. W. Sparks, D. A. Torchia, A. Bax. *J. Am. Chem. Soc.*, 1989, 111, 1515 - 1517
- 25) D. G. Davis, A. Bax. *J. Am. Chem. Soc.*, 1985, 107, 2820 - 2821
- 26) L. Braunschweiler, R. R. Ernst. *J. Magn. Res.*, 1983, 53, 521 - 528

Anexo 1

Espectros TOCSY de los 20 L-aminoácidos más comunes. Los círculos corresponden a las señales de correlación COSY; los rombos y los círculos corresponden al sistema de espín observado en los espectros TOCSY.





### Anexo 2

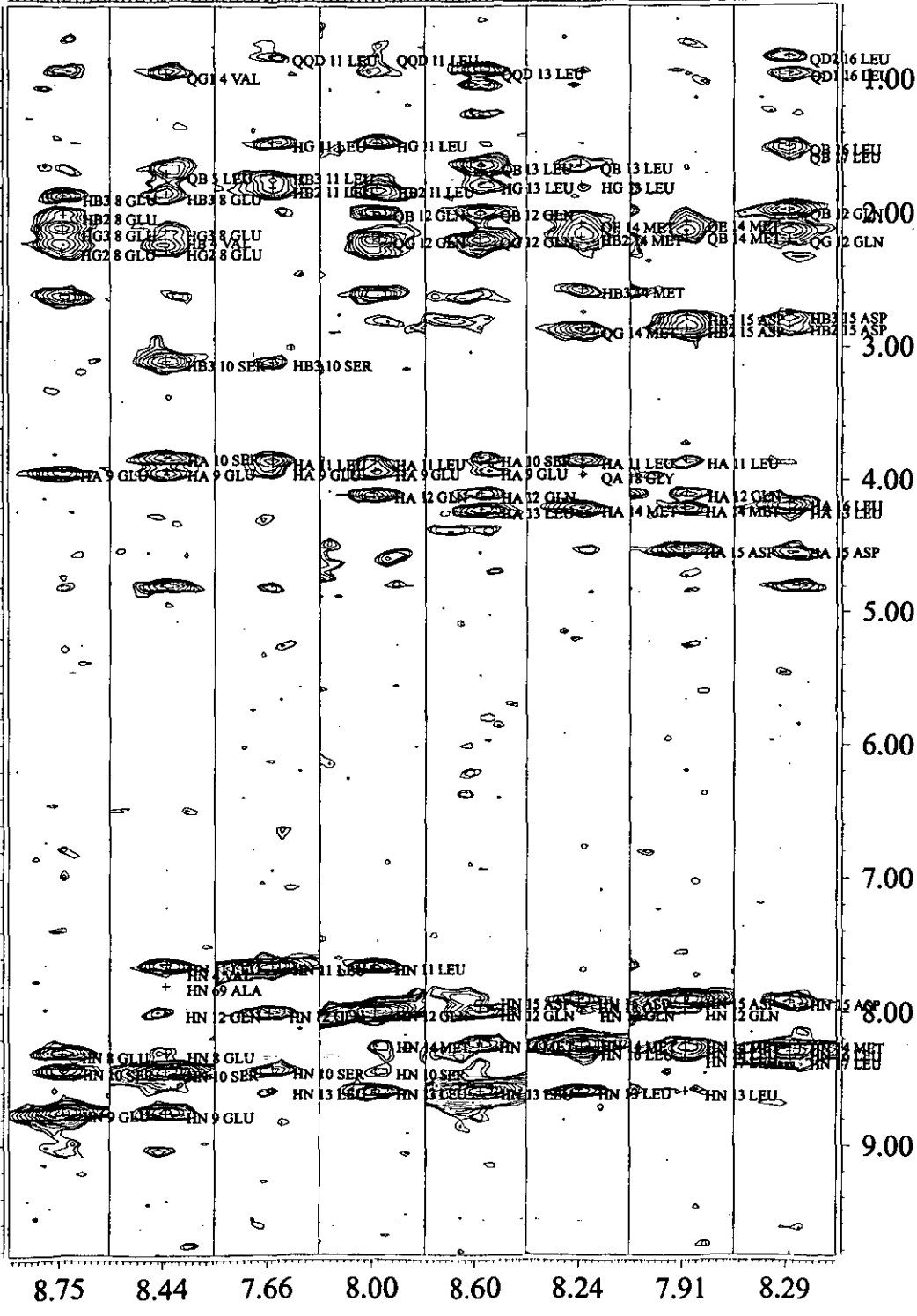
A continuación se muestran las franjas del espectro 3D  $^1\text{H}$ - $^{15}\text{N}$ -NOESY-HSQC, utilizadas en la determinación de los elementos regulares de estructura secundaria y de la estructura terciaria de baja resolución, del fragmento 1 – 108 de la proteína SV40. En ellos, sólo se muestra la asignación de las señales en el eje y (que corresponde a una dimensión de  $^1\text{H}$ ); en el programa Xeasy se designa a los protones como:

$\text{H}^{\text{N}}$	HN
$\text{H}^{\alpha}$	HA
$\text{H}^{\alpha}, \text{H}^{\alpha'}$	QA
$\text{H}^{\beta}, \text{H}^{\beta'}$	HB2, HB3
$\text{H}^{\beta}$	QB
$\text{H}^{\gamma}, \text{H}^{\gamma'}$	HG2, HG3
$\text{H}^{\gamma}$	QG
$\text{H}^{\delta}, \text{H}^{\delta'}$	HD2, HD3
$\text{H}^{\delta}$	QD
$\text{H}^{\epsilon}$	HE

la designación QB indica que los protones beta son equivalentes. El eje x del espectro, es el desplazamiento químico de  $^1\text{H}$  que se indica debajo de cada franja; el eje z es el desplazamiento químico de  $^{15}\text{N}$  indicado en la parte superior de cada franja.

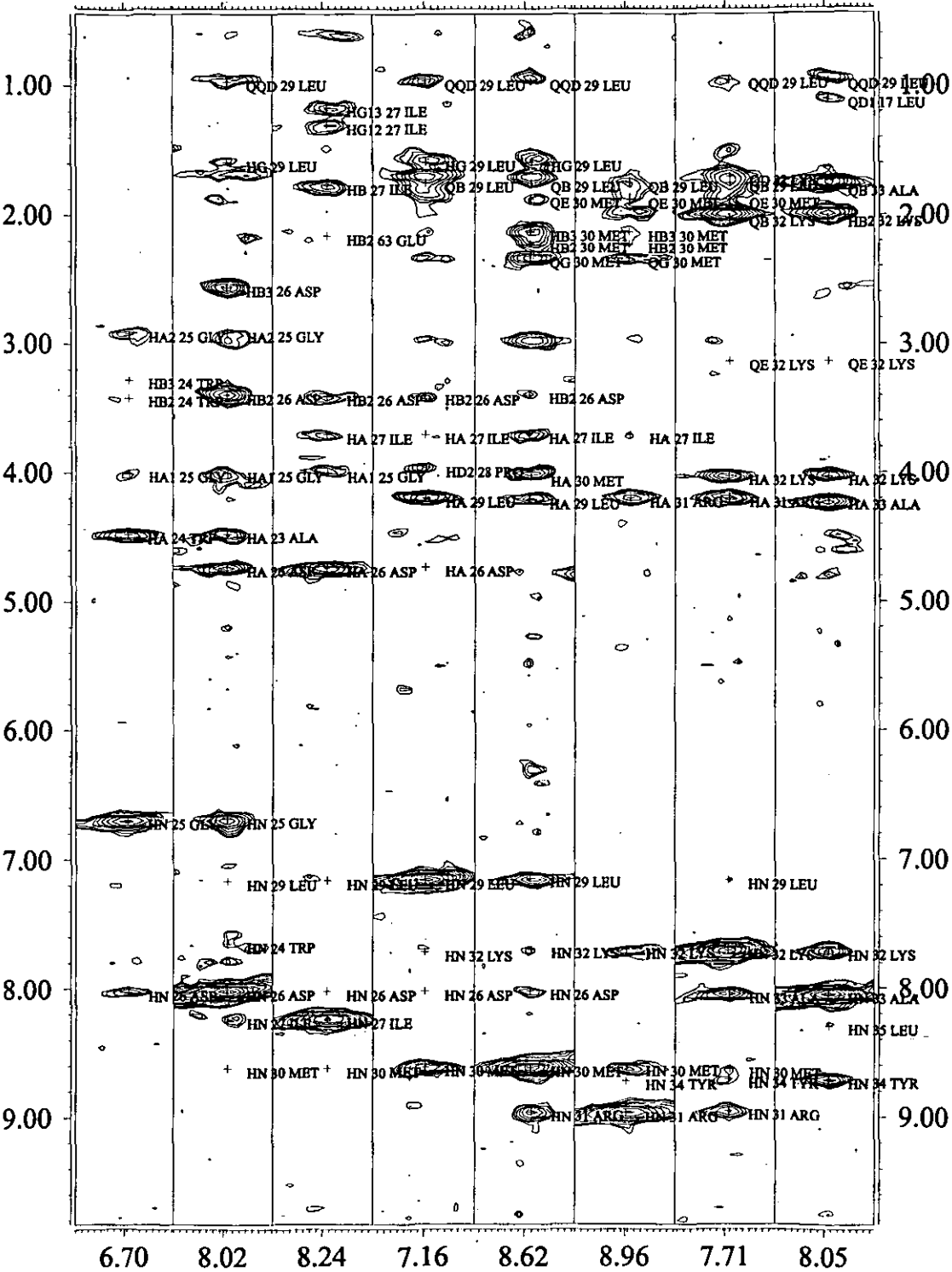


120.16 115.41 122.32 118.87 119.73 117.14 118.65 121.89



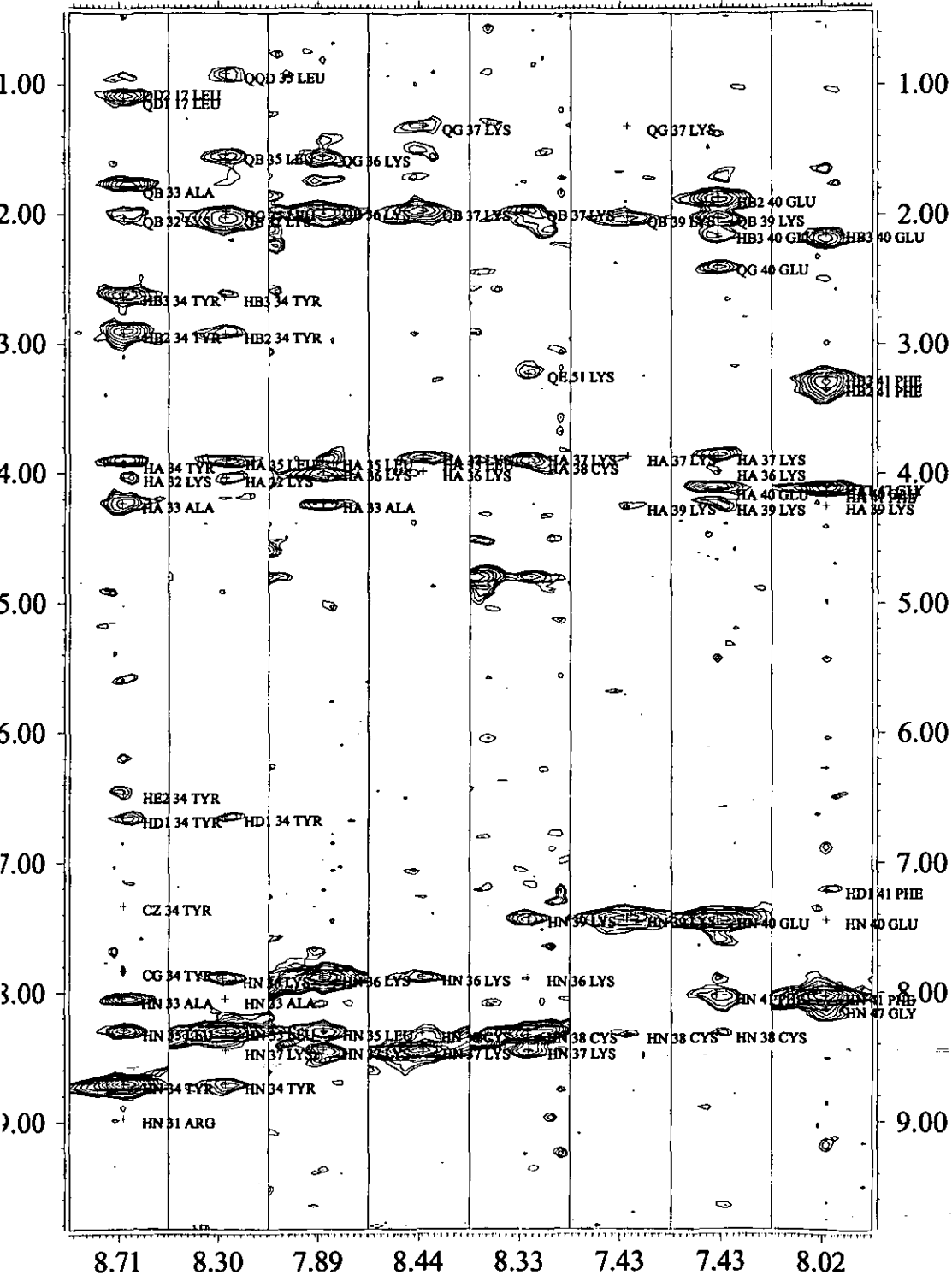


115.20 120.81 124.91 116.92 120.38 120.16 118.44 122.10

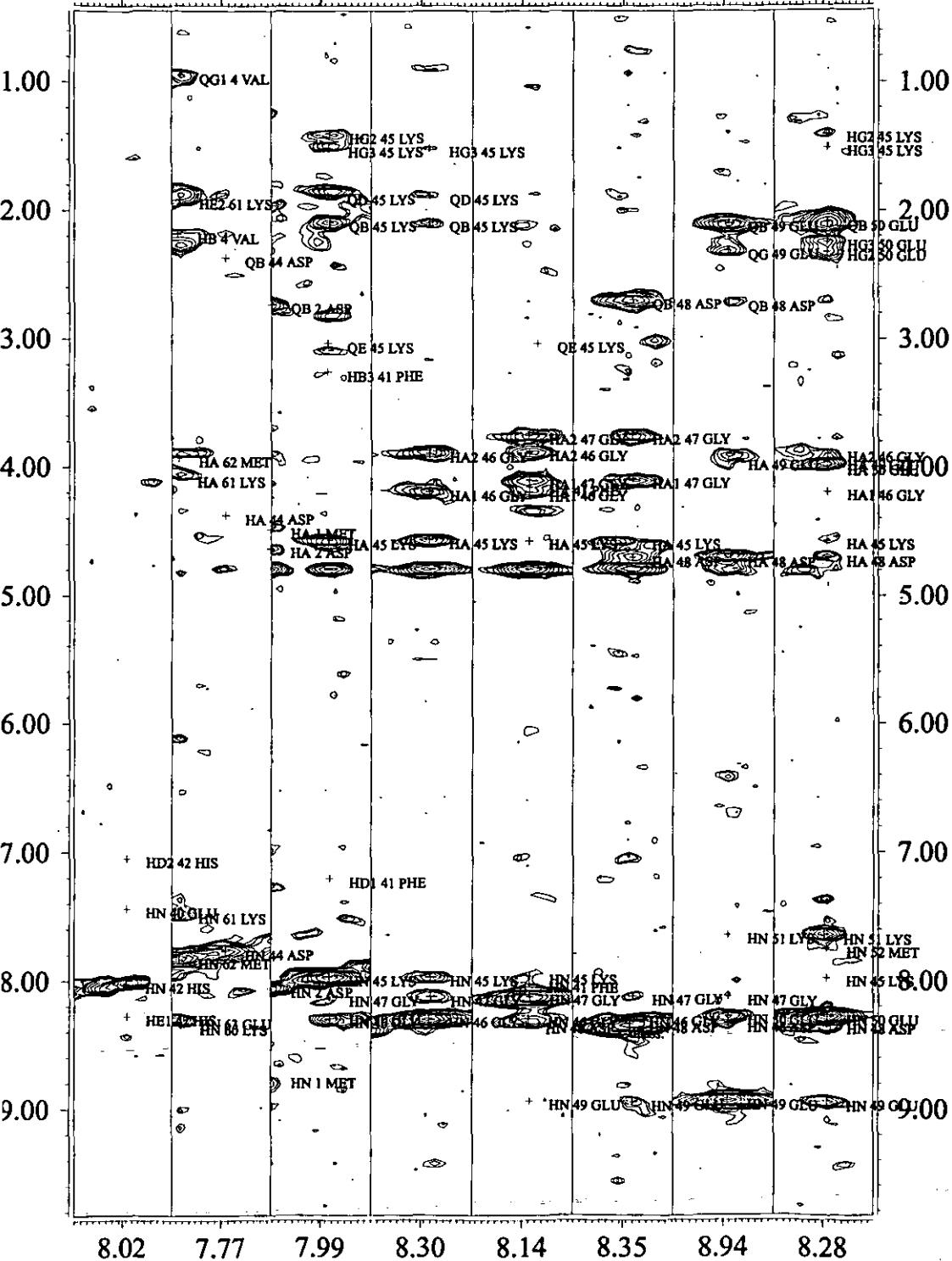




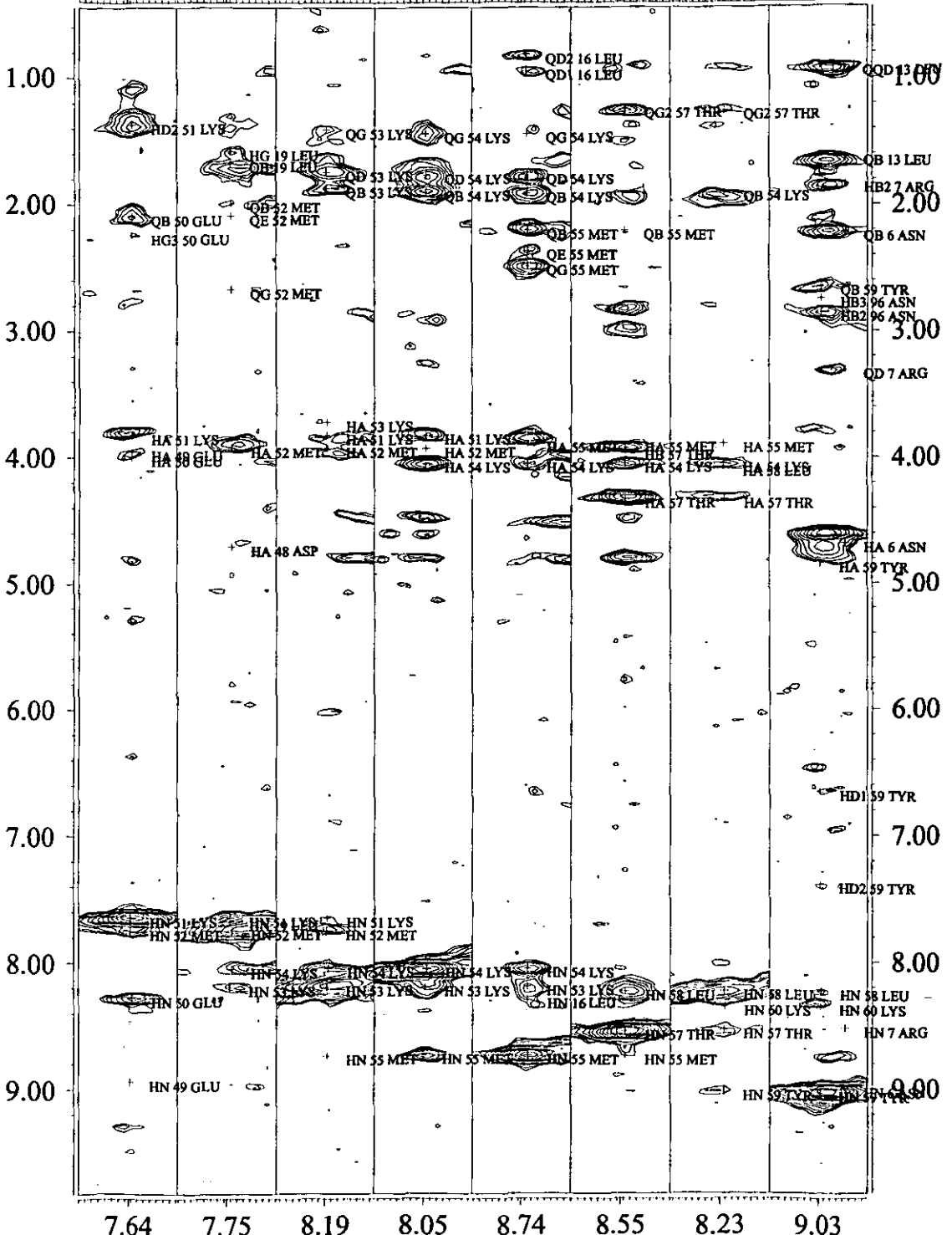
120.81 119.95 119.30 119.51 118.87 116.71 117.79 122.97



127.29 119.95 119.08 109.80 109.37 121.24 126.42 119.08



120.59 118.22 120.16 121.24 117.57 116.06 120.59 120.59



7.64

7.75

8.19

8.05

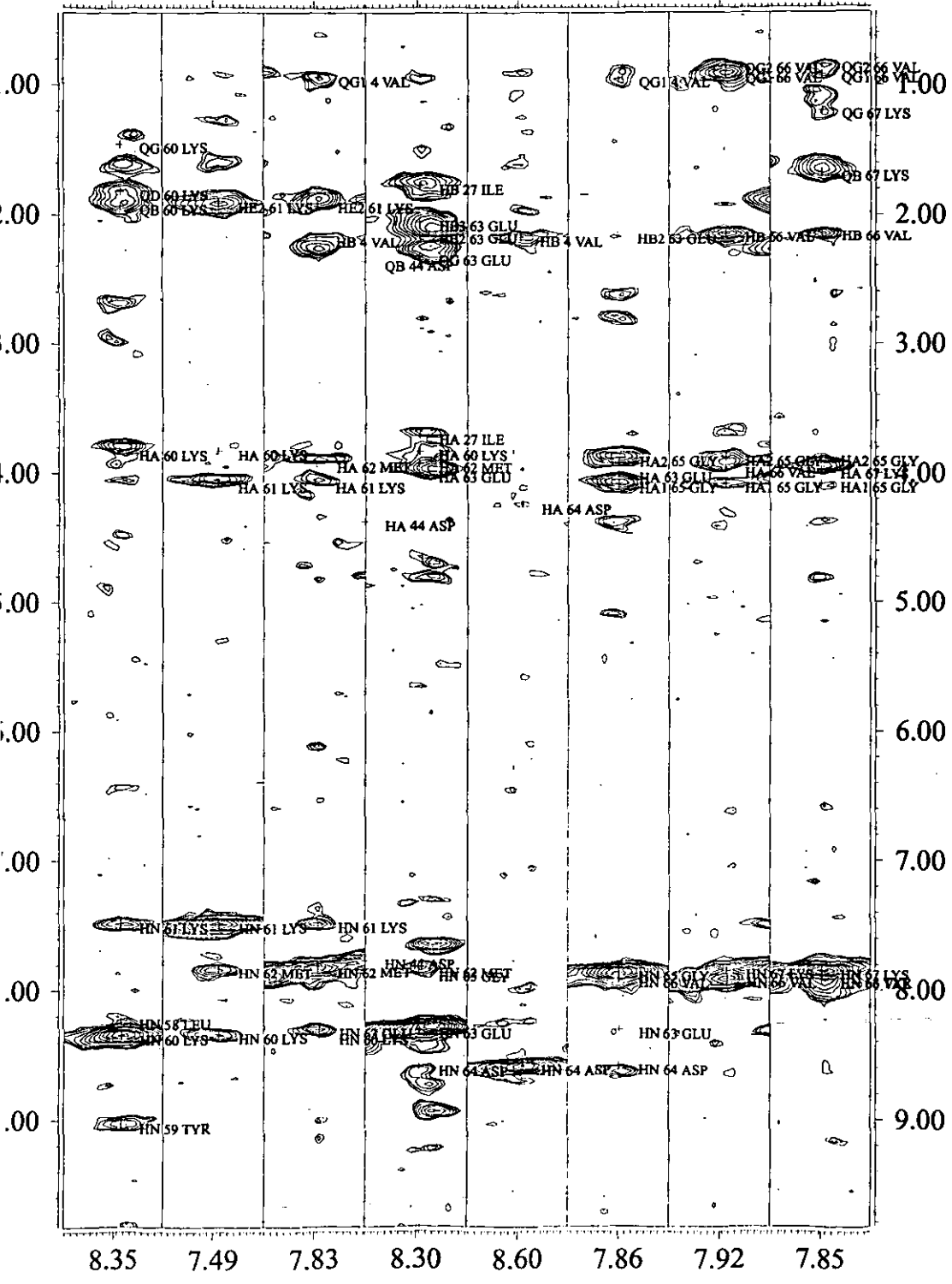
8.74

8.55

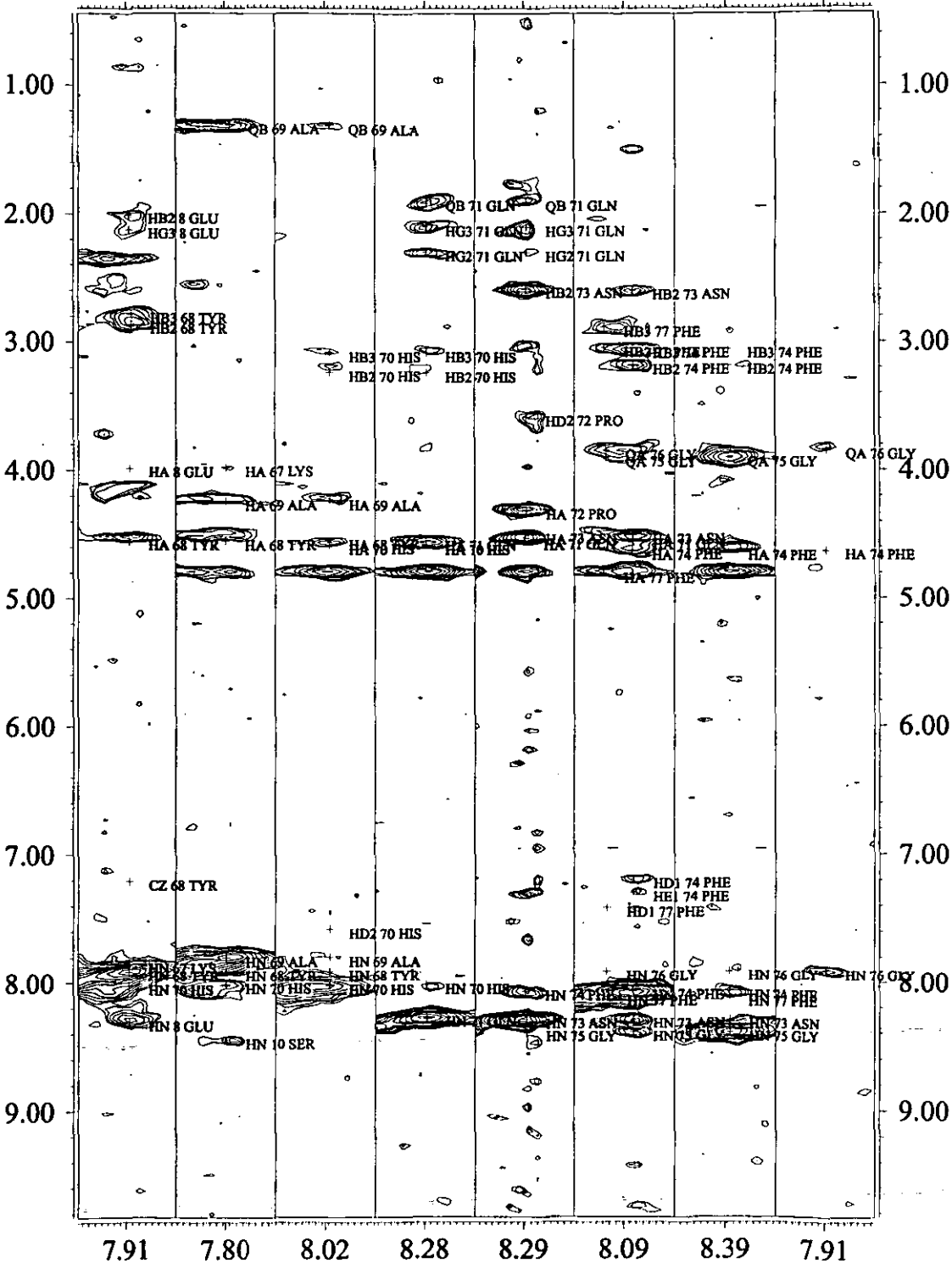
8.23

9.03

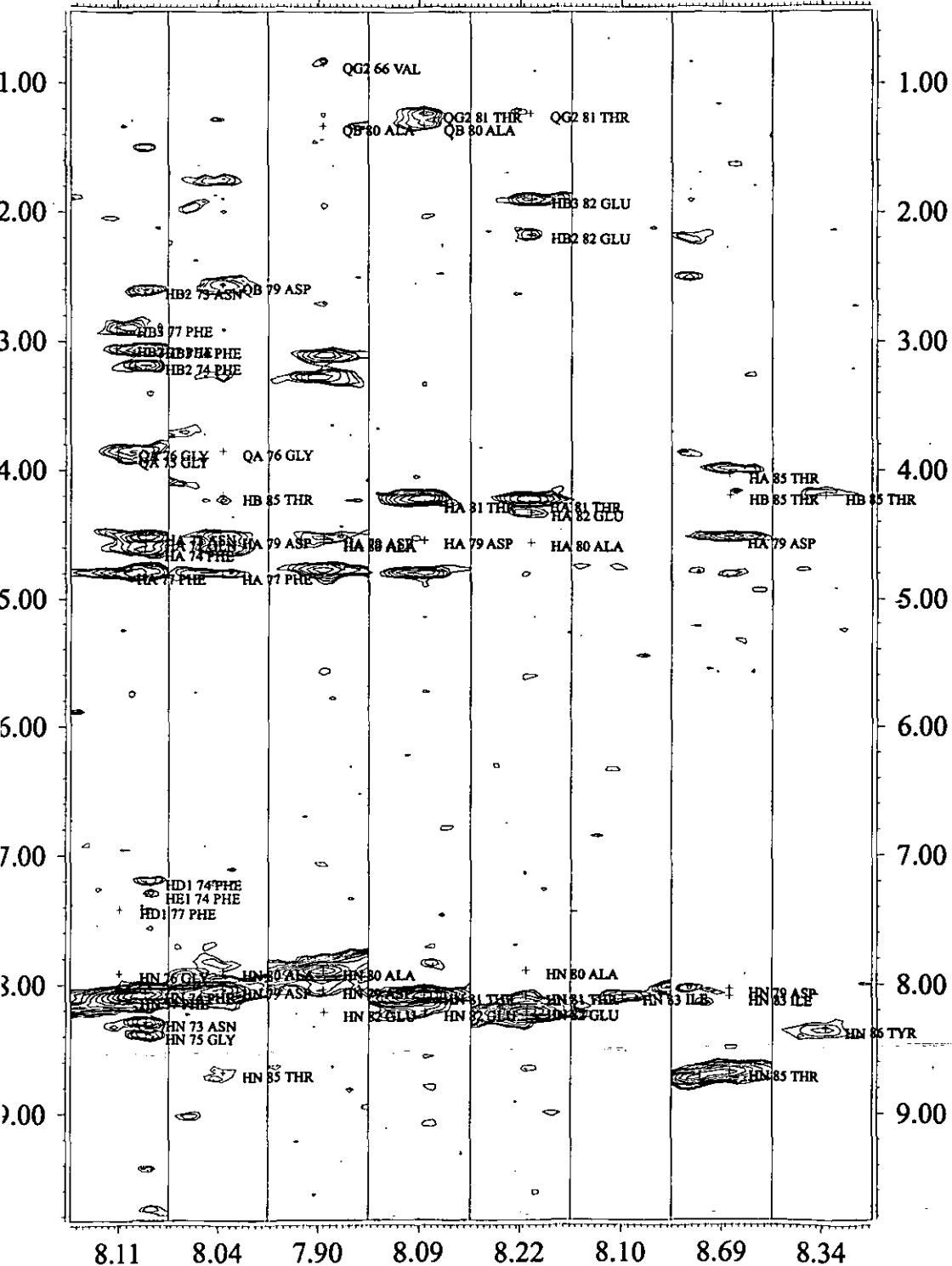
117.57 118.65 119.95 119.30 119.51 106.56 120.16 120.81



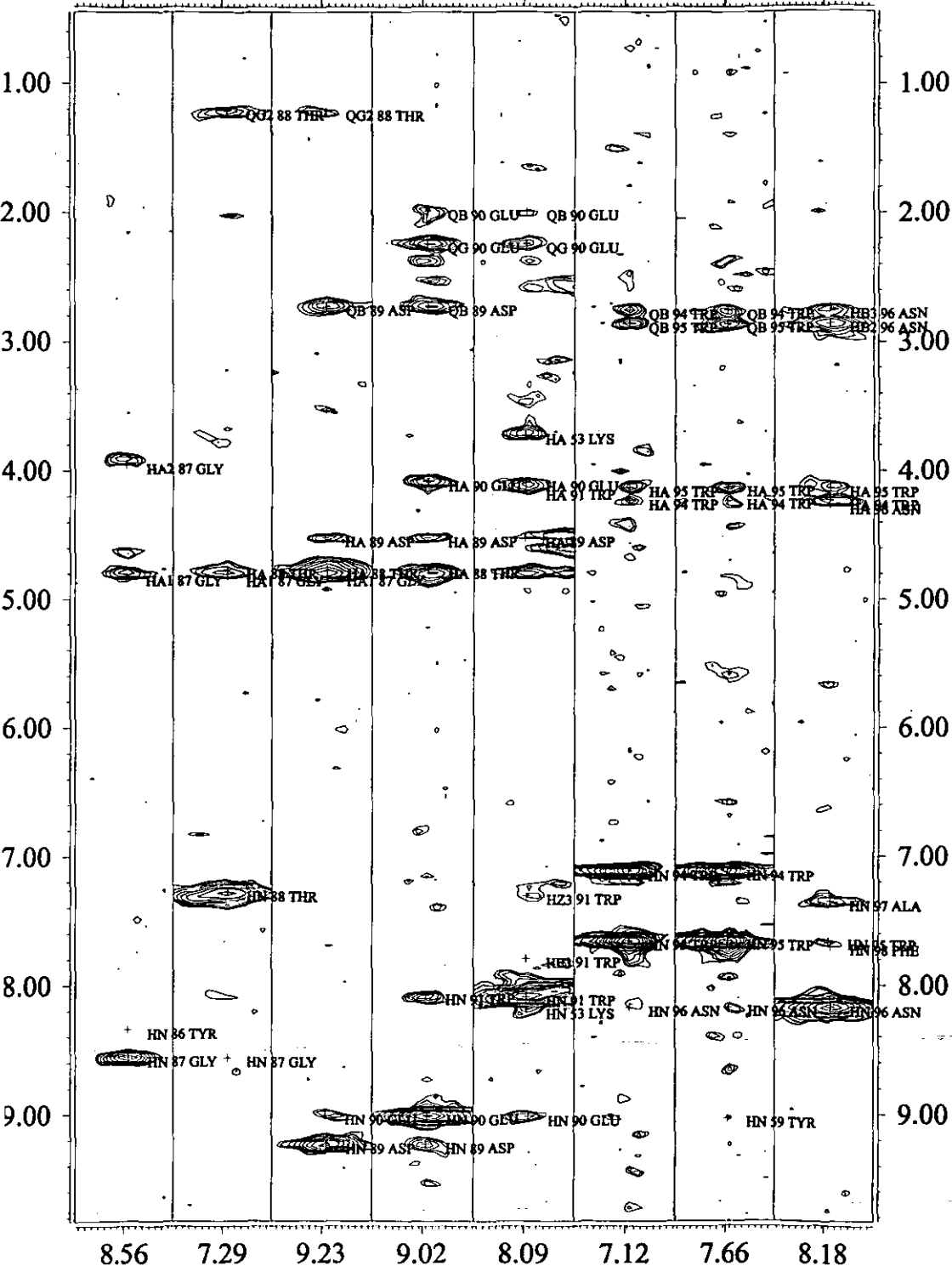
118.44 123.40 116.92 122.10 118.44 119.95 110.23 108.50



119.95 121.89 123.18 113.25 123.18 124.70 117.79 129.01

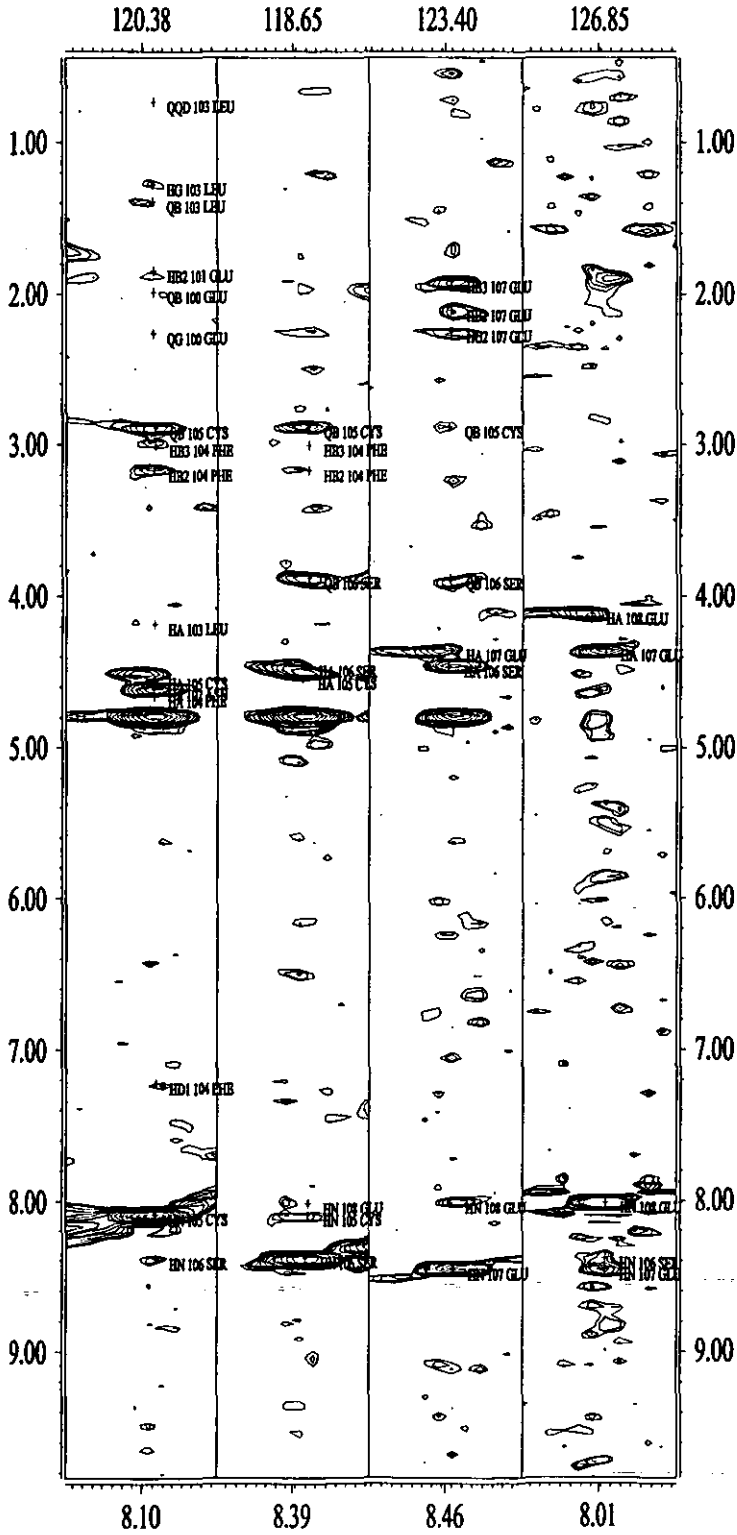


110.02 108.50 123.62 119.51 121.67 112.82 112.82 116.49









## Anexo 3

Hoy en día, en la determinación de la estructura terciaria de las proteínas utilizando RMN, el análisis de los espectros se efectúa con la ayuda de computadoras y de programas adecuados para ello. XEASY es un programa muy versátil, que nos permite manejar espectros en 2, 3 ó 4 dimensiones y asignar las señales de correlación a protones específicos.

En el transcurso del trabajo de asignación, las señales que son asignadas se almacenan en un archivo que es una base de datos en la que se guarda la información acerca del pico (al que el programa le asigna un número), desplazamientos químicos y los números de los picos asociados a esta correlación. Así mismo, en el programa se genera otra base de datos que asocia cada pico (su número) con el tipo de protón asignado ( $H^N$ ,  $H^\alpha$ ,  $H^\beta$ ,  $H^\gamma$ ,  $H^\delta$ ,  $H^\epsilon$ ). Cuando se integra el espectro, se asocia a cada una de las señales de correlación un volumen (o área, dependiendo de si es un espectro en 2D ó 3D) relativo, producto de la intensidad de las señales en el espectro.

Cuando se ha asignado un número suficiente de señales se efectúan cálculos de prueba que nos permiten inspeccionar el plegamiento de la proteína y observar si existen errores en la asignación (esto lo sabemos por que una correlación mal asignada provocará que varias restricciones no se cumplan, a las cuales se les llama *violación* del cálculo); lo que nos obliga a revisar y re-interpretar el espectro o continuar con la asignación si es que no se produce ninguna violación; en caso de haber asignado *todas* (o casi todas) las señales del espectro procedemos al cálculo final de la estructura. Para ello necesitamos transformar la información contenida en los archivos generados durante la asignación (archivos de XEASY) a un formato adecuado para el programa que efectúa los cálculos (CNS); el programa que efectúa la transformación de la base de datos de XEASY al formato que se utilizará en CNS es CALIBA. Para poder cambiar el formato de los archivos de XEASY (que sólo contiene correlaciones entre números) CALIBA utiliza los siguientes archivos que contienen la información adecuada para la transformación:

1° /usr/local/xeasy/src.new/ecepp.lib, archivo que contiene las distancias entre átomos, dentro de los residuos de aminoácidos;

2° archivo.seq, este es un archivo que contiene la estructura primaria (secuencia) de la proteína; debe de estar en formato de tres letras mayúsculas seguido por la posición que ocupa dicho aminoácido dentro de la secuencia (ALA 1 MET 2 CYS 3 THR 4 ARG 5 ...);

3° archivo.prot, archivo generado en XEASY que contiene la siguiente información del pico: # de pico, desplazamiento químico en la dimensión adecuada, tipo de pico ( $N$ ,  $HN$ ,  $H\alpha$ ,  $H\beta$ , etc.) y residuo al que pertenece;

4° archivo.peaks, archivo de XEASY que contiene los parámetros de correlación y los volúmenes (o área) de la señal de correlación NOE;

5° cal.cal, archivo de CALIBA con los parámetros de calibración necesarios para generar el archivo que se utilizará en los cálculos con CNS.

De este proceso se obtiene un archivo que contiene correlaciones entre *tipos de átomos* (p. ej., el  $HN$  del residuo 34 está a 3.5 Å del  $HA$  [alfa] del residuo 70) el cual reconoce CNS para

calcular estructuras, plegando a la proteínas de tal forma que se satisfagan las distancias obtenidas de los espectros NOESY.

CNS es un programa modular, integrado para efectuar múltiples tareas (cálculo de estructuras a partir de datos de RMN o de difracción de rayos X, refinamiento de las estructuras, transformación de datos, etc.) y que genera los diversos archivos que necesita para efectuar la tarea requerida.

CNS puede utilizar una interfase de red (archivos *web* manejados desde los navegadores comunes) para generar los archivos de entrada (*archivo.inp*) que contienen los parámetros que el programa necesita para efectuar las tareas requeridas, generando un archivo de salida (*archivo.out*) que contiene información global de los resultados del cálculo y archivos que el programa va generando.

Para el cálculo de estructuras de proteínas con datos de RMN, CNS necesita de dos archivos auxiliares: *archivo.mtf* y *archivo\_extended.pdb*. El archivo *mtf* contiene información necesaria para el programa CNS y el archivo *extended.pdb* contiene la estructura totalmente extendida (lineal) de la proteína, esto es porque el cálculo necesita una estructura inicial y para los primeros cálculos es mejor comenzar de una estructura con la geometría de los enlaces adecuados, pero completamente lineal para no influir al cálculo. En el proceso de refinamiento, ya se podrá utilizar, como estructura de partida, una que contenga el plegamiento calculado.

Para crear uno de los archivos auxiliares, obtenemos de la interfase de red el archivo de inicio *generate\_mtf.inp*, que necesita de un *archivo.seq* que especifica la estructura primaria de la proteína. Obtenido el archivo *inp*, se inicia un cálculo en CNS de la siguiente manera:

```
cns < generate_mtf.inp > generate_mtf.out
```

el programa creará los archivos *generate\_mtf.out* y el *archivo.mtf*.

Con el archivo *mtf*, podemos crear el otro archivo auxiliar, que contiene la estructura lineal de la proteína (*extended.pdb*). Este también se genera con un archivo de entrada obtenido de la interfase de la red. El archivo *extended.inp* necesita del *archivo.mtf*:

```
cns < extended.inp > extended.out
```

creando los archivos *extended.out* y *extended.pdb*, éste último contiene la estructura lineal.

Con los archivos de restricciones (distancias derivadas de NOE, ángulos diedros derivados de constantes de acoplamiento, desplazamiento químico, puentes de hidrógeno, etc.), obtenidos de la interpretación de los espectros de RMN, y estos archivos, podemos efectuar cálculos de la estructura de la proteína.

*annealing.inp* es el archivo que le indica a CNS que calcule la estructura de la proteína, considerando las restricciones guardadas en los archivos adecuados. Estas restricciones se almacenan en forma de tablas de datos (*archivo.tbl*), algunos de los cuales son:

*Archivo con datos de distancias, obtenidas de los espectros de efecto nuclear Overhauser:*

assign (resid # and name H\$) (resid # and name H\$)  $d_{\max}$   $d_{\min}$  0.5

# son los aminoácidos que muestran la correlación;

H\$ son los protones entre los que se presenta la correlación

$d_{\max}$  y  $d_{\min}$  son las distancias entre los protones, determinadas del experimento NOE (la distancia mínima es de 1.8 Å)

0.5 es la incertidumbre de la distancia máxima.

*Archivo con datos de puentes de hidrogeno, responsables de mantener la estructura secundaria, los puentes de hidrogeno se deducen de la gráfica de índice de protón:*

assign (resid #a and name O) (resid #b and name HN) 1.95 0.35 0.35

assign (resid #a and name O) (resid #b and name N) 2.90 0.40 0.40

#a y #b nuevamente representan los aminoácidos entre los que se establece el puente de hidrógeno; en las  $\alpha$  hélice los puentes de hidrógeno se establecen entre el carbonilo de un aminoácido  $i$  y el HN del aminoácido  $i+4$ , por lo que #b debe de cumplir con este requisito.

1.95 y 2.90 son las distancias entre los núcleos señalados;

0.35 y 0.40 son las incertidumbres.

*Archivo con restricciones de ángulo diedro, estos ayudan en el plegamiento correcto de los elementos de estructura secundaria; este archivo proporciona los ángulos diedro ( $\phi$ ) que se forman entre el plano del residuo #a y el plano del residuo siguiente #b:*

Assign (resid #a and name c) (resid #b and name n)

(resid #b and name ca) (resid #b and name c) 1.0 -57.0 15.0 2

1.0 es un parámetro que utiliza CNS;

-57 es al ángulo diedro para una  $\alpha$  hélice;

15 es la incertidumbre de dicho ángulo;

2 es un parámetro utilizado por CNS.

Para calcular las estructuras en CNS utilizamos el comando:

```
cns < anneal.inp > anneal.out
```

El resultado de estos cálculos es una familia de estructuras de la proteína, cada una de las estructuras estará en un archivo *protein.pdb*. La comparación de las estructuras generadas durante el cálculo nos permite determinar si existe una correlación que obligue a la estructura a

plegarse de una manera inadecuada, lo cual produce una *violación* al cálculo; de ser así, tenemos que revisar la asignación que provoca la violación y corregirla mediante la re-interpretación del espectro. Si no se han producido violaciones, la comparación de las estructuras nos permite visualizar la tendencia del plegamiento de la proteína y centrar la búsqueda de restricciones en las regiones de la proteína que van mostrando interacciones. Por eso, necesitamos de un programa que nos facilite la visualización y análisis de las estructuras; el programa Molmol se desarrolló como una herramienta que facilita la visualización y análisis de estructuras de proteína (en formato de *archivos:pdb*). Este es un programa muy dinámico, el cual no sólo facilita la visualización de la estructura de la proteína; permite la sobreposición de varias estructuras de la proteína y calcular la dispersión (RMSD) de la estructura global o de diferentes regiones; puede seleccionar átomos y calcular la distancia entre ellos; puede calcular superficies de isopotencial, hacer gráficas de Ramachandran, etc.

Para sobreponer, comparar y determinar el RMSD de  $n$  estructuras debemos utilizar los siguientes dos comandos:

```
SelectAtom '#1-n:x-y and bb'
```

$n$  es el número de estructuras que se quieren comparar (2, 5, 10, etc.);

$x-y$  indica que residuos se quieren comparar (1-108, del 1 al 108, 9-19, del 9 al 19, etc.)

$bb$  indica que sólo tome en cuenta la cadena principal (el esqueleto, *backbone*).

```
Fit to _first 'selected'
```

con este comando Molmol efectúa la sobreposición de las estructuras en la región indicada y muestra la desviación (RMSD).

En muchas ocasiones deseamos estudiar una región de la proteína y debemos *esconder* el resto de la misma, especialmente si queremos determinar la desviación de una región; si se quiere ver una región de la proteína o esconder esta región, utilizamos el comando:

```
SelectRes '#1-n:x-y'
```

y del menú de opciones se escoge la opción que se desea (*invisible* [para esconder], *ribbon* [para ver la estructura en forma de cintas], etc.).

En algunas ocasiones deseamos resaltar únicamente cierto tipo de átomos, lo que nos facilita su selección para calcular la distancia entre ellos; esto lo logramos con el siguiente comando:

```
SelecAtom '#1-n:x-y@H&'
```

donde H& representa el tipo de átomo que deseamos resaltar (HN, HA, HB, CA carbono  $\alpha$ , CB, CO, etc.), y del menú de opciones podemos escoger la que más nos convenga.