



Casa abierta al tiempo

UNIVERSIDAD AUTÓNOMA METROPOLITANA

**MODELADO MOLECULAR DE
QUIMOPAPAÍNA**

TESIS QUE PRESENTA
ALBERTA JAQUELINE PADILLA ZÚÑIGA

PARA OBTENER EL GRADO DE
MAESTRO EN QUÍMICA

NOVIEMBRE DE 1993

**UNIVERSIDAD AUTÓNOMA METROPOLITANA
UNIDAD IZTAPALAPA
DIVISIÓN DE CIENCIAS BÁSICAS E INGENIERÍA
DEPARTAMENTO DE QUÍMICA**

UNIDAD IZTAPALAPA

Av. Michoacán y La Purísima, Col. Vicentina, Iztapalapa, D.F. C.P. 09340. Tel.: 686-03-22

AGRADECIMIENTOS

Deseo expresar mi sincero agradecimiento a todas las personas que de una manera directa o indirecta estuvieron involucradas en la realización de este trabajo de tesis:

Agradezco al Dr. R. W. Pickersgill su colaboración al ceder tan generosamente el archivo de las coordenadas cristalográficas de la proteinasa Ω , de primerísima importancia en este trabajo.

A las Áreas de Química Cuántica y Fisicoquímica Teórica, agradezco muy ampliamente todas las facilidades que amablemente me brindaron para el uso de sus recursos de cómputo requeridos por el programa BIOGRAF, y de manera especial agradezco al Quím. Rodolfo Álvarez Bustamante el apoyo técnico otorgado durante el desarrollo del presente trabajo, siempre de manera expedita y desinteresada.

A mis compañeros del Área de Biofisicoquímica les agradezco el haberme permitido emplear sus resultados experimentales y por el ambiente de cordialidad que siempre ha imperado en nuestro grupo.

De manera particular deseo expresar mi gran agradecimiento al Dr. Andrés Hernández Arana por sus valiosas observaciones y su incondicional disposición en la asesoría del presente trabajo y al Dr. Arturo Rojo Domínguez por sus consejos, el tiempo y el trabajo concedidos a la elaboración de esta tesis.

Í N D I C E

CAPÍTULO 1	INTRODUCCIÓN	1
CAPÍTULO 2	MÉTODOS	8
	Selección de la estructura base	9
	1) Homología en composición	9
	2) Homología en secuencia. Alineamiento de pares de secuencias	10
	3) Significancia de la homología	11
	Modelado molecular	12
	Superficie de energía potencial	12
	Campo de fuerzas	14
	Minimización de energía o mecánica molecular	16
	Gradientes conjugados	18
	Dinámica molecular	19
	Método de Verlet	20
	Cálculo de áreas superficiales y volúmenes moleculares	24
	Superposición de estructuras	24

CAPÍTULO 3	RESULTADOS Y DISCUSION	25
	Selección de la estructura base	25
	a) Homología en composición	25
	b) Homología en secuencia	25
	c) Significancia de la homología en secuencia	27
	Modelado molecular de la quimopapaína	30
	a) Construcción del modelo inicial	30
	b) Minimización de energía	32
	c) Convergencia hacia la conformación de equilibrio	33
	Evaluación de la calidad estereoquímica del modelo	34
	Análisis de la estructura del modelo de la quimopapaína	46
	Comparación estructural del modelo con otros miembros de las P&V	52
	Aplicaciones del modelo propuesto	56
CAPÍTULO 4	CONCLUSIONES	78
BIBLIOGRAFIA		81

CAPÍTULO I

INTRODUCCIÓN

Por lo conocido acerca de las proteínas, todos los esfuerzos dedicados al estudio de sus propiedades y aplicaciones resultan justificados. Si como se ha dicho, el DNA es un proyecto de vida y las proteínas los "ladrillos" y el "mortero" que hacen posible todas las funciones vitales (Doolittle, 1985), no es sorprendente que encontremos a algunas proteínas cumpliendo funciones de reconocimiento molecular para el transporte de metabolitos o como defensa contra organismos patógenos; a otras, como reguladores del ritmo vital, como barrera física y soporte, como sintetizadores bioquímicos, como generadoras de movimiento y aún como materia prima de otras biomoléculas. Y lo más notable es que esta increíble variedad de funciones se verifica dentro de las células en condiciones suaves, con extraordinaria eficiencia y con gran especificidad. De entre las funciones de síntesis, quizá la labor catalítica es la más rica en cuanto a la diversidad de proteínas involucradas. La función enzimática, estrechamente relacionada con el reconocimiento molecular, ha llamado fuertemente la atención por las aplicaciones industriales que pueden derivarse de ella, tanto farmacéuticas como alimentarias y de síntesis química.

En el afán de utilizar a las proteínas como potentes catalizadores de las reacciones biológicas, se ha intentado producir estas moléculas en mayor cantidad que aquella en que se hallan naturalmente, sin embargo su obtención y su purificación posterior, representan un gran reto a la habilidad y experiencia del investigador. Por ello, recientemente se ha acudido a las técnicas de ingeniería de proteínas y de DNA recombinante, en donde se ha hecho evidente el fuerte nexo que existe entre la estructura de una proteína, la preservación de su integridad y sobre todo, su función biológica. Con estos antecedentes es comprensible el gran interés en los estudios estructurales, que de algunos años a la fecha se ha venido presentando en diferentes partes del mundo y que se manifiesta por ejemplo, en el diseño de proteínas de novo (Richardson & Richardson, 1984; Goraj *et al.*, 1990) cuyo objetivo es determinar si las tendencias generales, que han podido deducirse de la observación de un grupo de proteínas con estructura conocida, pueden aplicarse a la generación de proteínas sintéticas con patrones tridimensionales característicos y predeterminados. Aunque estos estudios sólo pretenden obtener

cadenas de proteínas plegadas en una conformación particular, ya se ha alcanzado otra meta aún más ambiciosa que es la de producir en el laboratorio péptidos artificiales con actividad biológica preestablecida (Hahn *et al.*, 1990). Estos trabajos permiten confiar en que el fin último de la ingeniería de proteínas sobre el diseño de estas macromoléculas con funciones específicas y nunca antes observadas en la naturaleza, se encuentra cada vez más cercano.

En la *Ingeniería de Proteínas* están involucrados tanto el diseño y construcción de macromoléculas completas, como la realización de modificaciones puntuales donde se reemplazan, suprimen o adicionan residuos individuales o aún fragmentos de cadena polipeptídica en la secuencia de una proteína natural. Todo esto con el objeto de identificar los sitios de unión de la proteína con su sustrato o ligando y establecer la jerarquía de funciones que cada residuo sostiene en la compleja maquinaria enzimática y de reconocimiento molecular. Así también, las técnicas mutagénicas se han venido empleando en la elucidación del papel que cumplen las cadenas laterales de los aminoácidos en la estabilidad y en el mecanismo de plegamiento de las proteínas que les permite alcanzar su estructura tridimensional biológicamente activa (Altamirano & Calcagno, 1993).

El gran esfuerzo que implica el manipular el DNA celular de un microorganismo para lograr la generación de una proteína específica, mantiene aún vigente la alternativa de estudiar mutantes naturales de una proteína con estrecha homología en secuencia y que conserven un patrón de doblado común. Esto permitiría no sólo identificar a los residuos inmutables de aquéllos que cumplen una función accesoria en la secuencia polipeptídica (como en la ingeniería de proteínas), sino que además, proveería de principios generales para la identificación y caracterización de estos grupos de mutantes, que forman verdaderas familias de proteínas, para sumarse a los importantes estudios sobre la relación estructura-función-estabilidad que se han mencionado antes.

Uno de los problemas más importantes por resolver en biología molecular, es la forma en que la secuencia de aminoácidos de una proteína es capaz de conducir a la macromolécula hacia la estructura tridimensional que determina su función biológica específica. Y aunque éste es un punto de debate actualmente, el hecho es que todos los estudios que se realizan sobre proteínas en el mundo, tienen el

objetivo mediato o inmediato de encontrar las claves estructurales que mantienen la estabilidad molecular y que controlan la maquinaria funcional de las proteínas, e intentar con ello definir las tendencias generales que sigue la naturaleza en el diseño estructural de estas biomoléculas. Por tanto, es necesario contar con un número cada vez mayor de estructuras tridimensionales de alta resolución, que permitan realizar estudios de caracterización sistemáticos para la elucidación de los mecanismos de plegamiento, con las claves moleculares para conocer el proceso de adquisición de la estructura polipeptídica tridimensional única y completamente funcional.

Con la intención de obtener nuevas estructuras de proteínas que permitan establecer las relaciones funcionales entre la secuencia de aminoácidos y la conformación tridimensional, se han venido empleando exitosamente dos métodos experimentales: la difracción de rayos X en cristales de proteína (Perutz, 1992) y la de resonancia magnética nuclear (RMN) de proteínas en solución (McIntosh & Dahlquist, 1990; Clore & Gronenborn, 1991). La primera de estas técnicas requiere monocristales de proteína estables a los rayos X y con un orden molecular que permita la adecuada difracción de la radiación incidente. Por su parte, la técnica de RMN precisa del manejo de soluciones de proteína en concentraciones relativamente altas (1-2 mM), valores de pH normalmente menores a 6 (Branden & Tooze, 1991), y además, moléculas que contengan un máximo de 150 residuos en su cadena polipeptídica (De Vlieg & van Gunsteren, 1991). En la realidad sólo algunas proteínas pueden cumplir las condiciones experimentales que imponen estas técnicas, y ello es responsable de que se conozcan sólo algunos cientos de estructuras tridimensionales, cuando el número de secuencias completamente determinadas que se tiene actualmente, se eleva a varias decenas de miles (Fasman, 1989).

Con el rápido crecimiento de la capacidad de procesamiento de datos y con las velocidades de cálculo numérico de que se dispone actualmente, es ya factible la utilización de programas de cómputo para la aplicación de técnicas como la mecánica y dinámica moleculares, que permiten simular el comportamiento y las propiedades de sistemas constituidos por varios miles de átomos, como las proteínas y ácidos nucleicos, con una muy buena aproximación a los resultados experimentales (Langone, 1991). De hecho, desde finales de los 60's se han venido empleando algoritmos de programación para el modelado molecular de proteínas con gran éxito, y a partir de entonces, se han perfeccionado los métodos

matemáticos que generan y representan gráficamente el comportamiento dinámico y las conformaciones espaciales más estables, que describen fielmente a los sistemas biológicos. Este último punto, constituye la solución concreta al problema de la adquisición de estructuras tridimensionales, ya que los elementos de partida de las técnicas de simulación molecular son relativamente asequibles y salvan muchas de las restricciones sobre la naturaleza molecular, que imponen las técnicas de difracción de rayos X y RMN mencionadas anteriormente; aunque de hecho, las aproximaciones teórica y experimental resultan ser complementarias en la mayoría de los casos. Las aplicaciones más sobresalientes de los programas de simulación además de la generación de modelos moleculares, son las que se refieren a la determinación de parámetros termodinámicos de sistemas poliatómicos (Karplus & Petsko, 1990) y a los cálculos dinámicos a temperaturas superiores a las del medio biológico, para producir la desnaturalización de una proteína; al menos hasta llevar a la estructura al estado de glóbulo fundido (Daggett, 1993).

En el Área de Biofísicoquímica de la UAM-I se está llevando a cabo un proyecto de caracterización termodinámica sobre la familia de enzimas denominada proteasas sulfhidrúlicas vegetales (**PSV**). Esta familia comprende un grupo de enzimas cuya actividad proteolítica depende del grupo tiol de un residuo de cisteína de su secuencia de aminoácidos. Aunque las proteasas sulfhidrúlicas más estudiadas son las que se obtienen de fuentes vegetales, existen análogos estructurales en animales y en algunos microorganismos (Kamphuis *et al.*, 1985a; Moore, 1969). Del grupo de proteasas sulfhidrúlicas o tiólicas vegetales, la papaína es la enzima más ampliamente estudiada y por tanto, la mejor conocida hasta ahora (Glazer & Smith, 1971; Brocklehurst *et al.*, 1981; Baker & Drenth, 1987). Este tipo de enzimas puede ser obtenido ya sea del látex, el endocarpio y el zumo, o bien del tallo, las hojas y aún las raíces, de un gran número de frutos y plantas. Por ejemplo, del látex de papaya se extraen la papaína y al menos otros dos tipos de proteasas: la quimopapaína y la papaya peptidasa A, también conocida como proteinasa Ω (Baker & Drenth, 1987); en el látex de las plantas *Ficus spp.*, *Asclepias syriaca* y *glaucescens* y *Calotropis gigantea*, pueden encontrarse las ficinas (Kortt *et al.*, 1974), las asclepáinas (Tablero *et al.*, 1991) y las calotropinas y calotropáinas (Pal & Sinha, 1980) respectivamente. Así mismo, del endocarpio de la piña y del fruto llamado kiwi, pueden purificarse las bromelaínas (Takahashi *et al.*, 1973) y las actinidinas (Carne & Moore, 1978), y de las hojas de la cebada puede obtenerse la aleuraína (Holwerda & Rogers, 1992). Los

distintos tipos de enzimas que se localizan en una sola fuente vegetal, no corresponden a simples productos de degradación sino que constituyen en realidad, tipos particulares de proteínas que pueden ser identificados tanto por las variaciones en sus secuencias de aminoácidos, como por la manifestación de sus diferentes propiedades físicas. Todas ellas guardan una estrecha semejanza u homología en sus secuencias de aminoácidos, con una cantidad de residuos idénticos que se encuentra en el intervalo de 40 a 90%; al efectuar comparaciones entre pares de proteínas. Actualmente se tienen caracterizadas por difracción de rayos X las estructuras de tres proteasas tiólicas vegetales: papaína (Kamphuis *et al.*, 1985b), actinidina (Baker, 1980) y proteinasa Ω (Pickersgill *et al.*, 1991). De la observación de la arquitectura de estas tres proteínas, se sabe que están integradas por dos dominios estructurales de aproximadamente 100 residuos de aminoácido cada uno, lo que corresponde a cerca de 200 residuos por cadena polipeptídica. También se conoce que el sitio activo, localizado en la hendidura interdominio, se encuentra integrado por una triada de aminoácidos: una cisteína del primer dominio, además de una histidina y una asparagina pertenecientes al segundo dominio. En esta triada, el grupo sulfhidrilo de la cisteína interacciona fuertemente con el anillo imidazol de la histidina, el cual a su vez, se encuentra estabilizado por el carbonilo de la cadena lateral de la asparagina. Espacialmente cercanas a estos tres residuos, se han identificado siete regiones coadyuvantes de la función catalítica que son los denominados subsitios S4, S3, S2, S1, S1', S2' y S3' de la enzima. De esta forma, siete residuos del sustrato peptídico (codificados por P4, P3, P2, P1, P1', P2' y P3'), interaccionan con los subsitios catalíticos mencionados (**Fig. 1-1**), y el punto de escisión está situado entre los subsitios S1 y S1'. También se sabe que la diferente afinidad o especificidad por los sustratos en estas enzimas (al menos para papaína y actinidina), está determinada principalmente por la composición y la geometría del subsitio S2 (Drenth *et al.*, 1976; Baker & Drenth, 1987).

En la familia de las proteasas sulfhidrúlicas vegetales se tienen completamente secuenciadas siete enzimas: papaína, actinidina, proteinasa Ω , quimopapaína, bromelaína de tallo, papaya peptidasa IV y aleuraína (Mitchel, 1970, Carne & Moore, 1978, Dubois *et al.*, 1988a, Watson *et al.*, 1990, Ritonja *et al.*, 1989a y b, Whittier *et al.*, 1987, respectivamente). En la **Tabla 1-I** se presenta la extensión de la cadena

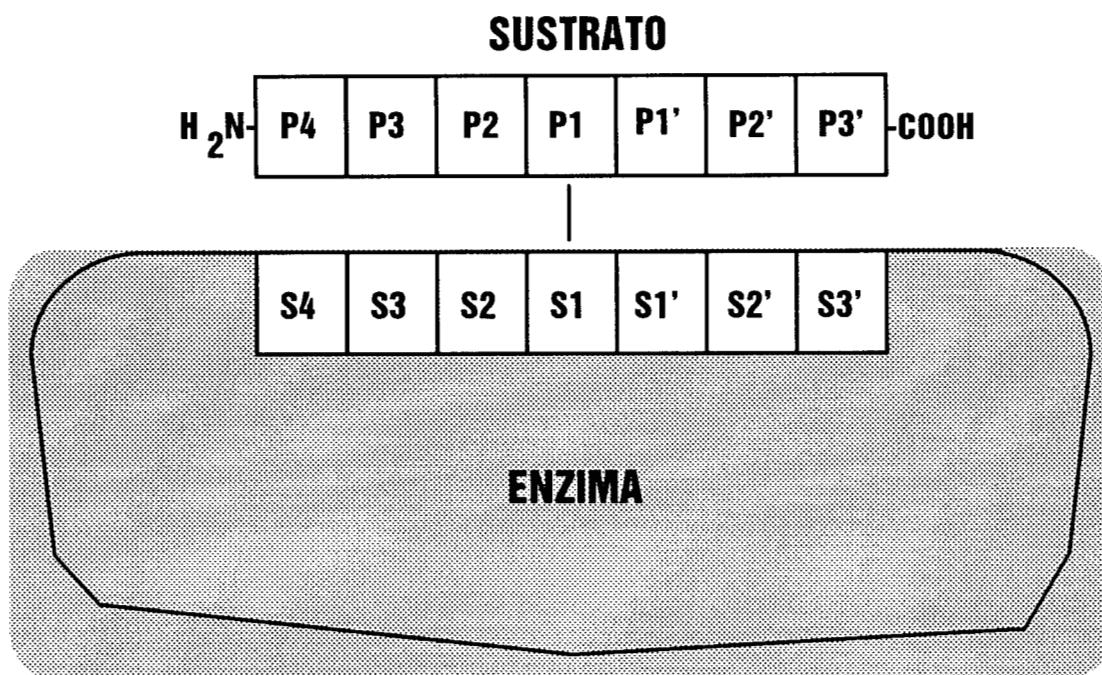


Figura 1-1. Representación de un polipéptido de siete residuos que interaccionan con los subsitios de una proteasa sulfhidrónica, de acuerdo con Baker & Drenth (1987).

polipeptídica y la indicación de las estructuras caracterizadas por difracción de rayos X, de las proteasas tiólicas que actualmente se estudian en nuestro laboratorio.

De las **PSV** completamente secuenciadas y cuya estructura tridimensional aún no se conoce, la quimopapaína es la más estudiada hasta ahora. Se tiene información sobre su actividad catalítica (Buttle & Barret, 1985), reactividad (Enzyme Handbook, 1991), topografía (Watson, *et al.*, 1990, Dubois *et al.*, 1988b, Carey *et al.*, 1983), contenido de estructura secundaria (Solís-Mendiola *et al.*, 1989) y estabilidad termodinámica (Solís-Mendiola *et al.*, 1992), entre otras. Una característica notable al comparar algunas de estas evidencias experimentales con las observadas para las demás **PSV** estudiadas, es que consistentemente la quimopapaína parece alejarse de las tendencias generales seguidas por la familia. Con estos antecedentes y con la intención de contribuir a la completa caracterización de la quimopapaína, el objeto inmediato en este trabajo, es el poder determinar una estructura tridimensional confiable para esta proteína, que ayude a esclarecer muchas de las interrogantes planteadas por los estudios experimentales. Sin embargo, aunque existen reportes

sobre la obtención de cristales de quimopapaína (Jansen & Balls, 1941; Ebata & Yasunobu, 1962) hasta ahora no se han publicado estudios de difracción de rayos X relacionados con ella. Por otra parte, su cadena de 218 residuos de aminoácido, hace impracticable la caracterización estructural por medio de la RMN con la tecnología de que se dispone actualmente. De aquí que la única alternativa viable para salvar los problemas que imponen las técnicas experimentales, es el empleo de la metodología de la simulación por computadora, que ha resultado ser exitosa en la determinación de la geometría tridimensional de una gran variedad de proteínas (Blundell *et al.*, 1988; Karplus & Petsko, 1990; Levitt, 1992; Srinivasan *et al.*, 1993). Así, el objetivo que se persigue en este trabajo es el de proponer una estructura tridimensional que represente lo más fielmente posible a la quimopapaína, a través de las técnicas de simulación por computadora, que emplean métodos de mecánica y dinámica moleculares.

En las siguientes secciones se describirá la metodología seguida para la obtención del modelo tridimensional de la quimopapaína, la estimación de su nivel de confiabilidad y su posterior aplicación a la interpretación de algunos de los resultados experimentales disponibles, sobre todo, de aquéllos que inciden directamente en los aspectos geométricos de las proteínas involucradas.

TABLA 1-I

<i>PROTEASA</i>	<i>NO. DE RESIDUOS</i>
* PAPAÍNA	212
QUIMOPAPAÍNA	218
* PROTEINASA Ω	216
BROMELAÍNA DE TALLO	212
* ACTINIDINA	220
FICINA	?

* Proteínas con estructura tridimensional determinada por difracción de rayos X.

CAPÍTULO 2

MÉTODOS

La similitud conformacional entre proteínas de estructura desconocida se determina normalmente, a través de sus distancias relativas mutuas sobre un árbol filogenético construido con la información estadística de la estructura primaria de un grupo de macromoléculas. Es una característica casi invariante, que las proteínas que guardan alta homología en sus secuencias de aminoácidos presenten patrones de plegamiento tridimensional muy semejantes. Aunque también se han encontrado casos en los que proteínas con una escasa semejanza en sus secuencias manifiestan gran relación geométrica (Ryu *et al.*, 1990; Holmgren & Branden, 1989). Esto sugiere que la diversidad en topología espacial está mucho más limitada que la variabilidad observada en las secuencias de aminoácidos de las proteínas. A pesar del importante progreso que se ha conseguido, sobre de la comprensión del mensaje cifrado en la secuencia de aminoácidos que relaciona a ésta estrechamente con el patrón terciario dentro de una proteína, aún se desconocen en gran medida los principios del doblado de la cadena polipeptídica. Así, el modelado de estructuras basado en la homología en secuencia, constituye el principal medio para predecir la estructura de una proteína mientras no se encuentren disponibles los datos cristalográficos o de RMN (Srinivasan *et al.*, 1993).

Dos requisitos indispensables en el modelado de estructuras homólogas son: a) el que se disponga de la secuencia de aminoácidos de la proteína de estructura desconocida; b) que se cuente con información tridimensional de moléculas de la misma familia estructural a la que pertenece la proteína en cuestión (Browne *et al.*, 1969; Greer, 1981; Blundell *et al.*, 1987). En este trabajo se seleccionó una de las estructuras cristalográficamente determinadas en la familia de las **PSV** para que sirviera de molde a la quimopapaína. Sobre el esqueleto de este molde se sustituyeron todos los residuos que constituyen la secuencia de la quimopapaína, y por técnicas de simulación molecular, se permitió que la molécula híbrida adquiriera la conformación representativa de la proteína en estudio. Finalmente se procedió a evaluar la calidad estereoquímica del modelo obtenido y a confrontar sus características geométricas con la información experimental disponible.

A continuación se describe la metodología seguida en este trabajo para modelar la estructura tridimensional de la quimopapaína.

Selección de la estructura base

Para poder elegir convenientemente de entre las tres proteasas de estructura tridimensional conocida y pertenecientes a las **PSV**, aquella que mejor sirviera de base para modelar a la quimopapaína, se compararon las secuencias de aminoácido de estas enzimas mediante la aplicación de tres criterios de selección estadísticos independientes. Se emplearon las secuencias reportadas por Mitchel *et al.* (1970) para papaína, por Carne & Moore (1978) para actinidina y por Dubois *et al.* (1988) para la proteinasa Ω . Estas secuencias coinciden con las estructuras primarias correspondientes a las mismas enzimas y que se encuentran depositadas junto con sus coordenadas espaciales en los archivos 9PAP, 2ACT y 1PPO (en pre-liberación) del banco de datos cristalográficos (**PDB**) (Bernstein *et al.*, 1977; Abola *et al.*, 1987). Para el caso de la quimopapaína se utilizó la secuencia de aminoácidos determinada recientemente por Jacquet *et al.* (1989) y Watson *et al.* (1990). Los criterios de selección mencionados que se describen a continuación, involucraron comparaciones en la composición de aminoácidos, análisis de la homología en secuencia y determinación de la significancia en esa homología en estructura primaria, para cada proteína caracterizada tridimensionalmente en su confrontación con quimopapaína.

1) Homología en composición.

En este método se determina la diferencia en contenido de aminoácidos que mantienen dos proteínas, a fin de establecer una comparación en composición entre ellas que sirva como primer criterio de homología; independientemente del conocimiento que se tenga de las secuencias individuales de las macromoléculas en estudio. La homología en composición se evalúa a través de la suma de las diferencias en el contenido de cada uno de los 20 aminoácidos constituyentes de una pareja dada de proteínas, de acuerdo con la siguiente expresión:

$$IDC = \sum_{i=1}^{i=20} (compA_i - compB_i)^2 \quad (1)$$

donde **IDC** es el índice de la diferencia en composición y $compA_i$ y $compB_i$ las composiciones del aminoácido i en la pareja de proteínas involucradas **A** y **B**. Estas composiciones pueden ser referidas tanto a la frecuencia con que aparecen los residuos en la cadena polipeptídica, como al porcentaje del total de aminoácidos que corresponde a cada residuo en una proteína dada. En cualquier caso, de acuerdo con la relación (1), un valor de **IDC**=0 significa que las proteínas en comparación tienen idéntica composición de residuos (cero diferencias); y por consiguiente a un mayor valor de **IDC** corresponde una mayor divergencia entre las composiciones de las moléculas comparadas. En este trabajo se aplicó el criterio descrito para confrontar la composición de aminoácidos de la quimopapaína y cada una de las tres proteasas cristalizadas de las **PSV**.

2) Homología en secuencia. Alineamiento de pares de secuencias.

Para llevar a cabo la comparación topológica entre estructuras primarias, se utilizó el algoritmo desarrollado por Myers & Miller (1988) específicamente para el alineamiento de parejas de secuencias de aminoácidos. El criterio de homología en este caso, consistió en la identificación de la estructura que mantuviera el mayor número de residuos idénticos y/o similares, simultáneamente con el menor número de inserciones en su cadena polipeptídica, al ser alineada con la quimopapaína. El método de Myers & Miller para alinear una pareja de secuencias de aminoácidos, que llamaremos nuevamente **A** y **B**, consiste en la determinación del conjunto de operaciones que transforman la proteína **A** en **B** de tal manera que la suma de los costos de cada operación sobre las secuencias, sea la menor. Las tres operaciones definidas son:

- (a) Reemplazo de un aminoácido por otro.
- (b) Supresión de n aminoácidos consecutivos.
- (c) Inserción de n aminoácidos consecutivos.

El costo de reemplazar un residuo por otro se encuentra especificado en una matriz **W** donde cada elemento en el arreglo matricial $w(a,b)$, corresponde al cambio del aminoácido **a** de la primer proteína por el aminoácido **b** de la segunda. En este trabajo se utilizó la matriz **MDM-78** (Schwartz & Dayhoff, 1978) que se basa en la estadística de mutaciones puntuales en secuencias de proteínas homólogas. El costo de insertar o suprimir un conjunto de n aminoácidos en cualquiera de las secuencias por alinear, se calcula a través de la función:

$$F(n) = CA + (n \cdot CU) \quad (2)$$

donde **CA** es el costo de apertura de una inserción y **CU** es el costo por cada aminoácido que deba ser agregado a la secuencia. De acuerdo con lo recomendado por los autores, se utilizaron valores de penalización de 100 unidades para cada uno de estos parámetros de alineamiento.

Los residuos de aminoácido que son considerados similares entre sí por el método de alineamiento empleado en este trabajo fueron: **A**, **S** y **T** (residuos pequeños y polares); **D** y **E** (residuos que pueden adquirir carga eléctrica negativa); **N** y **Q** (residuos polares vinculados con los dos anteriores); **R** y **K** (residuos que pueden adquirir carga eléctrica positiva); **I**, **L**, **N** y **V** (residuos alifáticos); **F**, **Y** y **W** (residuos aromáticos).

3) *Significancia de la homología.*

El tercer criterio empleado en la selección de la estructura de base para el modelo de la quimopapaína, fue la ponderación del parecido en secuencia entre dos proteínas, respecto a lo que pudiera deberse solamente a casos fortuitos o aleatorios. Es decir, de acuerdo con el criterio estadístico, para precisar si dos estructuras presentan homología real que pudiera dar una clara indicación de algún ancestro común, es necesario obviar a todas las proteínas cuya secuencia resulte semejante a la de interés sólo por una combinación estocásticamente afortunada de sus residuos. El algoritmo del método que se utilizó aquí es el diseñado por Needleman & Wunsch (1970), modificado por Dayhoff (1978) y Feng *et al.* (1985). Este método consiste en comparar, en unidades de desviación estándar, la puntuación lograda al alinear las secuencias de aminoácidos de la proteína de interés y de aquella de la que se sospecha un parentesco. Esta puntuación se almacena, e independientemente se vuelven a comparar las secuencias de aminoácidos de estas dos proteínas, pero ahora la secuencia de la segunda macromolécula se mezcla aleatoriamente, aunque sin cambiar la composición, y se evalúa un nuevo marcador. La mezcla de aminoácidos al azar junto con el cálculo de la homología en secuencia en cada iteración, se prolonga a voluntad del usuario. El grupo de valores obtenido de esta manera, genera toda una distribución de eventos de homología aleatoria en los que la proteína de comparación estará incluida; a menos que realmente mantenga cercanía evolutiva con la proteína de interés. Cabe señalar que en el uso de este

método se encontró una gran dependencia del parámetro de significancia con el número de comparaciones aleatorias. Los resultados de este análisis se publicaron en otro lugar (Rojo & Padilla, 1993) y de ellos se establece que el número óptimo de iteraciones que se requiere para llevar a cabo una comparación confiable entre las secuencias analizadas, debe mantenerse entre 80 y 150 mezclas aleatorias. Aunque esta prueba se aplica fundamentalmente a proteínas con baja homología en secuencia, se optó por utilizarla en este trabajo para tener mayor confianza en que la estructura seleccionada sería la base más adecuada para la construcción del modelo de la quimopapaína.

Modelado molecular

Las simulaciones de mecánica y dinámica moleculares, se realizaron en un sistema IBM RS6000 con 16MB de memoria RAM y 16 Mflops de rendimiento en cálculo numérico. Este recurso se utilizó junto con el programa de cómputo BIOGRAF Versión 2.2 (Molecular Simulations Inc., Waltham, MA.) y el campo de fuerzas *Dreiding-II* (Mayo *et al.*, 1990). La paquetería de BIOGRAF permite el manejo de las estructuras depositadas en los bancos de datos cristalográficos, su representación gráfica, su manipulación y análisis, así como la realización de todos los cálculos de energía que permiten la simulación de sistemas macromoleculares.

A continuación se reseñan brevemente los principios teóricos en los que se fundamentan las técnicas de modelado molecular, y se explican los algoritmos de minimización y dinámica moleculares empleados por BIOGRAF.

Superficie de energía potencial

La descripción matemática completa de una molécula es un problema formidable debido a las pequeñas escalas dimensionales y a las grandes velocidades involucradas en los movimientos atómicos. Debido a que aún no se cuenta con una teoría mecánico-cuántica relativista que aporte una descripción adecuada de todos los efectos que se presentan en las moléculas o en sistemas moleculares, es conveniente revisar un planteamiento de la energía en forma simplificada.

De acuerdo con la ecuación de Schrödinger:

$$H\Psi(\mathbf{R},\mathbf{r}) = E\Psi(\mathbf{R},\mathbf{r}) \quad (3)$$

donde H es el Hamiltoniano para el sistema, Ψ es la función de onda y E es la energía (por lo común, Ψ es función de las coordenadas tanto del núcleo como de los electrones). Aunque muy general, esta ecuación es demasiado compleja para cualquier uso práctico, de manera que deben introducirse algunas simplificaciones como la de Born & Oppenheimer (1927) según la cual puede generarse, con buena aproximación, una ecuación para el movimiento electrónico que sea independiente de la que describe al movimiento nuclear, y que dependa sólo paramétricamente de las posiciones de los núcleos atómicos:

$$H\Psi(\mathbf{r};\mathbf{R}) = E\Psi(\mathbf{r};\mathbf{R}) \quad (4)$$

En esta ecuación se define una energía conocida como la superficie de energía potencial, que es función únicamente de las coordenadas nucleares.

La segunda ecuación que describe el movimiento nuclear en la superficie de energía potencial es:

$$H\Phi(\mathbf{R}) = E\Phi(\mathbf{R}) \quad (5)$$

donde Φ es ahora una función que depende solamente de las coordenadas nucleares.

Si se requiere el estudio de la evolución en el tiempo de una molécula, debe resolverse la ecuación (5), sin embargo la ecuación (4) debe ser resuelta previamente, lo cual no resulta trivial; por ello, para resolver la ecuación (4) se utiliza normalmente un ajuste empírico a la superficie de energía potencial. La solución a la ecuación mecánico-cuántica (5) es llamada dinámica cuántica, pero dado que puede considerarse a los núcleos como objetos relativamente pesados al compararlos con los electrones, los efectos cuánticos resultan despreciables y en tal caso, al introducir una aproximación, la misma ecuación (5) puede reemplazarse por la ecuación del movimiento de Newton, donde se calculan las velocidades que experimenta cada átomo de la molécula como función del tiempo, en cada conformación:

$$-\frac{\delta E}{\delta \mathbf{R}} = m \frac{\delta^2 \mathbf{R}}{\delta t^2} \quad (6)$$

La solución de la ecuación (6) mediante el ajuste empírico a la superficie de energía potencial, es llamada dinámica molecular. La mecánica molecular o minimización de energía, que es un método paralelo a la dinámica molecular, por su parte, ignora la evolución de los sistemas en el tiempo y se especializa en la búsqueda de geometrías particulares, sus energías asociadas y otras propiedades estáticas.

Campo de fuerzas

El campo de fuerzas es el ajuste empírico a la superficie de energía potencial y está integrado tanto por la serie de ecuaciones que describen las diferentes interacciones moleculares, como por el conjunto de parámetros que resultan del ajuste a la superficie de energía potencial (Ermer, 1976; Hagler, 1985). Los campos de fuerzas empleados comúnmente para la descripción de las moléculas, emplean una combinación de coordenadas internas (distancias y ángulos de enlace, torsiones y distorsión en centros de inversión) para describir la contribución de las interacciones covalentes a la superficie de energía potencial. Además, estos campos utilizan las distancias interatómicas para estimar las interacciones van der Waals, electrostáticas y de puentes de hidrógeno. La forma de las funciones es muy variable, y van desde simples ecuaciones cuadráticas hasta funciones de Morse, expansiones de Fourier y potenciales Lennard-Jones, entre otras. El objetivo de un campo de fuerzas es describir, con precisión razonable, a una gran cantidad de moléculas diferentes. De esta manera, el campo de fuerzas interpola y extrapola a partir de la información de un pequeño conjunto de moléculas empleadas para su parametrización, hacia un conjunto mayor de estructuras y moléculas relacionadas. Cabe señalar que en un campo de fuerzas, tanto la forma de las funciones, como los parámetros en sí mismos, representan la única gran aproximación en el modelado molecular. La calidad del campo de fuerzas, su aplicabilidad y su habilidad para predecir las propiedades estimadas en la simulación, determina directamente la validez de los resultados.

La parte fundamental de un programa de simulación molecular es el cálculo de la energía potencial para una determinada distribución espacial de los átomos. El

cálculo de esta energía, junto con sus derivadas respecto a las coordenadas atómicas, aporta la información necesaria para la representación de los movimientos atómicos reales.

Existen diferentes campos de fuerzas que se emplean comúnmente para la simulación de una gran cantidad de moléculas tanto orgánicas como inorgánicas. Entre ellos destacan: *Dreiding-II* (Mayo *et al.*, 1990), *AMBER* (Wiener *et al.*, 1984; 1986), *CVFF* y *CFF91* (Biosym Technologies, 1993) y *CHARMM* (Brooks *et al.*, 1983).

Como se había mencionado, las coordenadas reales de los átomos de una molécula combinadas con el campo de fuerzas, describen la superficie de energía potencial a través de una expresión general que considera las contribuciones de interacciones enlazadas (o covalentes) y no enlazadas de la siguiente manera:

$$E_{\text{potencial}} = E_{\text{covalente}} + E_{\text{no covalente}} \quad (7)$$

donde:

$$E_{\text{covalente}} = E_{\text{enlaces}} + E_{\text{ángulos}} + E_{\text{torsiones}} + E_{\text{inversiones}} \quad (8)$$

(1) (2) (3) (4)

$$E_{\text{no-covalente}} = E_{\text{vdW}} + E_{\text{Coulomb}} + E_{\text{p-hidrógeno}} \quad (9)$$

(5) (6) (7)

La expresión matemática para los términos de las ecuaciones (8) y (9) está determinada por el campo de fuerzas específico que se utilice. En este trabajo se empleó el campo de fuerzas *Dreiding-II* que forma parte del programa de simulación molecular BIOGRAF, y en el que se definen las contribuciones a la energía potencial de la siguiente forma:

$$E_{\text{covalente}} = \sum K_e (R - R_0)^2 + \sum K_a (\theta - \theta_0)^2 + \sum K_t (\cos\varphi - \cos\varphi_0)^2 + \sum K_j (\cos\omega - \cos\omega_0)^2 \quad (10)$$

(1) (2) (3) (4)

$$E_{\text{no-covalente}} = \sum D_o \left\{ \left[\frac{R_o}{R} \right]^{12} - 2 \left[\frac{R_o}{R} \right]^6 \right\} + \sum C_o \left\{ \frac{Q_i Q_j}{\epsilon R_{ij}} \right\} +$$

(5) (6)

$$\sum D_o \left\{ 5 \left[\frac{R_o}{R} \right]^{12} - 6 \left[\frac{R_o}{R} \right]^{10} \right\} \tag{11}$$

(7)

En las expresiones anteriores los términos (1) a (4) representan la energía de deformación de las distancias y ángulos de enlace (R y θ), ángulos de torsión o diedros (φ) e interacciones fuera del plano o de inversión (ω), respectivamente. Los términos (5) a (7) corresponden a interacciones entre átomos no enlazados: van der Waals, representada por una función tipo Lennard-Jones 12-6; una representación coulombica de las interacciones electrostáticas (donde la constante dieléctrica, ϵ , es dependiente de la distancia entre las cargas eléctricas para simular el efecto de apantallamiento del solvente) y finalmente, un término explícito para representar a las interacciones por puentes de hidrógeno, también con una función tipo Lennard-Jones pero con exponentes 12-10. Todas las constantes presentes forman parte de la parametrización del campo de fuerzas. La **figura 2-1** muestra esquemáticamente las interacciones atómicas consideradas por la generalidad de los campos de fuerzas.

Minimización de energía o mecánica molecular

Un método de gran utilidad para explorar la superficie de energía potencial consiste en encontrar conformaciones que sean puntos estables sobre esa superficie. La minimización tiene como objetivo primordial, el determinar una configuración espacial en la que la fuerza neta sobre cada átomo sea cercana a cero. De esta manera, las conformaciones estables pueden determinarse simplemente llevando al valor mínimo la energía del sistema. Sin embargo se presenta una serie de problemas al realizar esta tarea, sobre todo en el caso de proteínas y otras macromoléculas, ya que generalmente existen varios mínimos en la superficie de energía potencial donde las fuerzas sobre cada átomo son cercanas a cero, por lo que la identificación del mínimo global no resulta trivial. Por otra parte, aunque teóricamente es fácil determinar la convergencia del proceso, en la práctica se dificulta su identificación, debido a que el método aproxima asintóticamente la

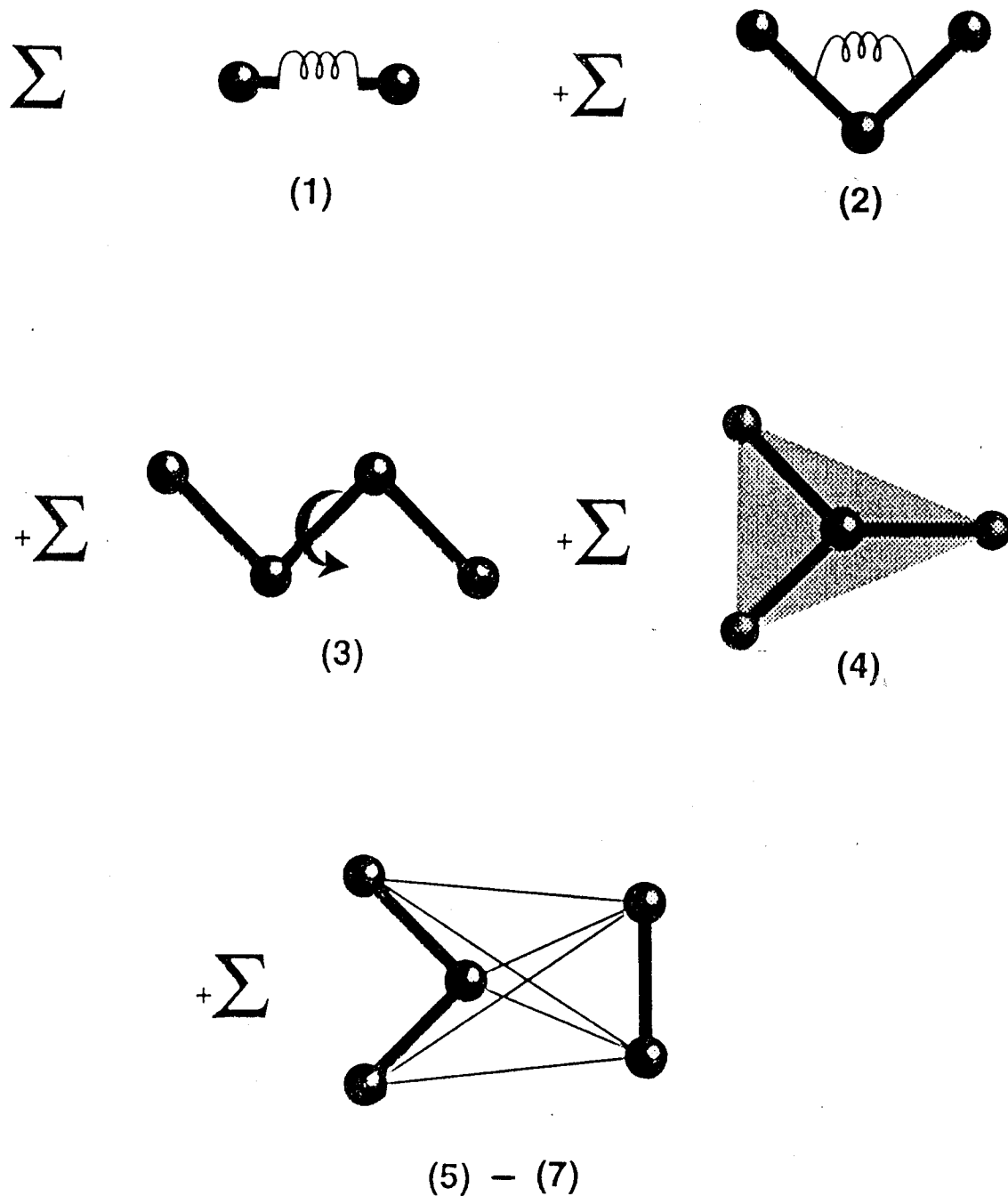


Figura 2-1. Representación gráfica de los términos de las ecuaciones (10) y (11). Las contribuciones a la energía potencial que se incluyen son: 1) distancia de enlace; 2) ángulo de enlace; 3) torsiones o ángulos diedros; 4) interacciones fuera del plano o inversiones; 5)-7) interacciones van der Waals, coulombicas y de puentes de hidrógeno.

energía del sistema a su valor mínimo, y el usuario debe decidir en qué momento se está ya suficientemente cerca de este punto.

La minimización de una molécula se realiza en dos etapas; primero se evalúa la expresión de energía como función de las coordenadas para una conformación inicial dada, y después se ajusta esa conformación de tal manera que la energía alcance su menor valor. Este mínimo de energía, puede encontrarse en una o en varias iteraciones (que pueden llegar a ser de cientos), dependiendo de la naturaleza del algoritmo de minimización, de la forma de la ecuación que expresa la energía y, del tamaño de la molécula. La eficiencia del proceso de minimización se juzga tanto por el tiempo requerido para evaluar la función de energía, como por el número de ajustes estructurales o iteraciones necesarias para converger a un mínimo. La mayoría de los algoritmos de minimización, asumen que la superficie de potencial es aproximadamente armónica, ya que aún las superficies anarmónicas son descritas convenientemente en el límite de convergencia en el mínimo. Dada una ecuación que describe la superficie de energía para un sistema y un punto de inicio sobre ella (conformación inicial), los métodos de minimización deben determinar la dirección hacia el mínimo y la distancia que lo separa de él. Una dirección razonable para iniciar el cálculo, es simplemente la que produce el mayor cambio de la función en ese punto. Los algoritmos más comunes de minimización, en orden de eficiencia, son: búsqueda lineal, máximos descensos, gradientes conjugados y Newton-Raphson (Biosym Technologies, 1993). En este trabajo se utilizó el método de gradientes conjugados debido a que es el más eficiente a nuestro alcance, pues el algoritmo de Newton-Raphson requiere una cantidad de memoria que excede la capacidad de cualquier computadora actual en el caso de macromoléculas.

Gradientes conjugados

Este método utiliza un algoritmo que produce una base completa de direcciones mutuamente conjugadas, de tal forma que cada paso sucesivo refina continuamente la dirección de búsqueda hacia el mínimo de energía. El algoritmo principia con la determinación de la dirección de máximo cambio de la superficie de energía, \mathbf{h}_0 , igualándolo con la derivada o gradiente, \mathbf{g}_0 , en un punto de inicio cualquiera. En esa dirección busca iterativamente la conformación de mínima energía, y actualiza las coordenadas a ese punto. Las direcciones siguientes, \mathbf{h}_{i+1} , son calculadas a través de la fórmula:

$$\mathbf{h}_{i+1} = \mathbf{g}_{i+1} + \gamma_i \mathbf{h}_i \quad (12)$$

donde \mathbf{g}_{i+1} es el gradiente en los puntos sucesivos, y γ_i es un escalar definido como:

$$\gamma_i = (\mathbf{g}_{i+1})^T(\mathbf{g}_{i+1})/(\mathbf{g}_i)^T(\mathbf{g}_i) \quad (13)$$

Todos los cálculos de minimización de energía de este trabajo, se realizaron con las opciones de mecánica molecular del programa BIOGRAF. El criterio de convergencia utilizado consistió en que la magnitud del gradiente de energía entre dos etapas consecutivas de cálculo, fuese menor o igual a $2 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$.

Dinámica molecular

La dinámica molecular resuelve las ecuaciones del movimiento para sistemas atómicos, lo cual se traduce como la representación de la evolución en el tiempo de los movimientos moleculares (también llamada trayectoria). Esta técnica permite a la molécula remontar máximos de energía para explorar múltiples conformaciones estables (mínimos locales de la superficie de potencial). La dependencia con el tiempo es simultáneamente la fortaleza y la debilidad de este método, ya que sólo pueden simularse de manera práctica, trayectorias del orden de 10^{-12} a 10^{-9} segundos (de picosegundos a nanosegundos). Como se había mencionado previamente, en la metodología de simulación de la dinámica molecular se resuelve la ecuación del movimiento de Newton para cada átomo del sistema: $\mathbf{F}_i = m_i \mathbf{a}_i$, donde \mathbf{F}_i es la fuerza que actúa sobre el i -ésimo átomo, m_i su masa y \mathbf{a}_i su aceleración. El cálculo de la fuerza, necesario para resolver esta ecuación, puede obtenerse directamente de la derivada de la energía potencial E respecto a las coordenadas \mathbf{r}_i ; esta es la razón de expresar a la energía potencial en una forma analítica y diferenciable:

$$\mathbf{F}_i = - (\delta E / \delta \mathbf{r}_i) = m_i (\delta^2 \mathbf{r}_i / \delta t^2) \quad (14)$$

En principio, con una expresión adecuada de la energía potencial y con las masas conocidas de todos los átomos, sería posible resolver la ecuación diferencial (14) para calcular futuras posiciones atómicas en el tiempo (esto es, determinar la trayectoria). Desafortunadamente, la física clásica sólo proporciona soluciones exactas

a esta ecuación para sistemas de 1 ó 2 partículas independientes. La resolución de sistemas mayores requiere el empleo de métodos numéricos, que aportan soluciones aproximadas.

El movimiento de un átomo particular puede expresarse mediante su desarrollo en una serie de Taylor. Si la posición al tiempo t es $\mathbf{r}(t)$, la posición después de un pequeño intervalo de tiempo Δt será:

$$\mathbf{r}(t+\Delta t) = \mathbf{r}(t) + (\delta\mathbf{r}/\delta t)\cdot\Delta t + (\delta^2\mathbf{r}/\delta t^2)\cdot\Delta t^2/2 + \dots \quad (15)$$

La solución numérica de las ecuaciones del movimiento depende de dos grupos de factores: *a)* el conocimiento de la posición $\mathbf{r}(t)$, la velocidad $\delta\mathbf{r}/\delta t$ y la aceleración $\delta^2\mathbf{r}/\delta t^2$ y *b)* la aproximación que omite a los términos de orden superior. Las coordenadas iniciales se obtienen generalmente de estructuras cristalográficas, mientras que las velocidades iniciales son asignadas al azar siguiendo la distribución de Maxwell-Boltzmann que corresponda a la temperatura de interés en la simulación (normalmente 300 K). Una vez resuelta la ecuación (15), las coordenadas originales se reemplazan por nuevas y las velocidades se actualizan dividiendo el cambio de posición de cada átomo por el incremento de tiempo Δt ; la aceleración se corrige calculando los gradientes $\delta E/\delta \mathbf{r}_i$ con las nuevas coordenadas. Un ciclo se completa al realizar estos cálculos para todos los átomos constituyentes de la molécula, y este proceso se repite miles o quizá, millones de veces para efectuar la simulación dinámica del sistema.

Método de Verlet

El algoritmo usado por BIOGRAF para resolver las ecuaciones diferenciales (15), es el propuesto originalmente por Verlet (1967), y que se describirá brevemente a continuación.

Si \mathbf{V}_{prom} es la velocidad promedio durante el intervalo de tiempo transcurrido entre t y $(t+\Delta t)$, entonces la posición al final del periodo será:

$$\mathbf{r}(t+\Delta t) = \mathbf{r}(t) + \mathbf{V}_{\text{prom}}\Delta t \quad (16)$$

Si se supone que la velocidad cambia linealmente durante Δt , entonces la velocidad promedio puede igualarse a la velocidad instantánea en $\Delta t/2$:

$$\mathbf{V}_{\text{prom}} = \mathbf{V}(t + \Delta t/2) \quad (17)$$

Por otra parte, la velocidad instantánea puede ser calculada de la aceleración promedio desde $(t - \Delta t/2)$ hasta $(t + \Delta t/2)$:

$$\mathbf{V}(t + \Delta t/2) = \mathbf{V}(t - \Delta t/2) + \mathbf{a}_{\text{prom}} \Delta t \quad (18)$$

Si nuevamente se supone que la aceleración es aproximadamente lineal en el intervalo de tiempo descrito, entonces la aceleración promedio puede estimarse como su valor instantáneo en el tiempo t :

$$\mathbf{a}_{\text{prom}} = \mathbf{a}(t) \quad (19)$$

Lo anterior conduce a la siguiente expresión para la velocidad:

$$\mathbf{V}(t + \Delta t/2) = \mathbf{V}(t - \Delta t/2) + \mathbf{a}(t) \cdot \Delta t \quad (20)$$

Al combinar las ecuaciones (16) y (17) se obtiene una ecuación similar para la actualización de las coordenadas, donde $\mathbf{V}(t + \Delta t/2)$ se conoce a través de la ecuación (20):

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{V}(t + \Delta t/2) \cdot \Delta t \quad (21)$$

Puede observarse que las posiciones atómicas se calculan al final de cada incremento de tiempo, mientras que las velocidades se actualizan a la mitad del intervalo ($\Delta t/2$). Cada ciclo que estima las nuevas coordenadas en el tiempo consta de: *a)* el cálculo de la aceleración, $(-1/m_i) \cdot \delta E / \delta \mathbf{r}_i(t)$, al tiempo t , *b)* la actualización de la velocidad al tiempo $(t + \Delta t/2)$, a partir de su valor en el instante $(t - \Delta t/2)$ con la ecuación (20) y *c)* la estimación de las coordenadas para el tiempo $(t + \Delta t)$ usando su valor al tiempo t y la ecuación (21).

Un parámetro clave en el algoritmo de Verlet es el valor asignado al Δt . Obviamente para optimizar los recursos de cómputo conviene emplear el valor más alto del incremento de tiempo. Sin embargo, existen límites fundamentales para el uso de intervalos de tiempo grandes. La principal limitación son los movimientos moleculares de la mayor frecuencia a ser simulados. Las vibraciones rápidas implican cambios abruptos de velocidad y aceleración. Un periodo vibracional debe ser dividido en al menos 5 ó 10 segmentos para satisfacer la suposición de Verlet respecto a que las velocidades y aceleraciones sean lineales a lo largo del periodo de integración. Se conoce que las frecuencias de vibración más altas corresponden a la distorsión de la longitud de enlaces que involucran átomos ligeros; en nuestro caso la vibración más alta se espera para los enlaces entre átomos de carbono e hidrógeno (C-H), cuyo periodo es del orden de 10^{-14} segundos. Por lo tanto en este trabajo se utilizó un Δt de 10^{-15} segundos (1 fs).

Otro concepto fundamental en las simulaciones de dinámica molecular es el de la temperatura. Básicamente, la temperatura es proporcional a la energía cinética de la molécula, lo cual puede expresarse más convenientemente en términos de las velocidades atómicas. La relación entre temperatura y velocidad de un sistema se encuentra a través de la Teoría Cinética de los Gases, con la ecuación de Maxwell-Boltzmann:

$$f(v)\delta v = (m/2\pi kT)^{3/2} \cdot \exp(-mv^2/2kT) \cdot 4\pi v^2 \delta v \quad (22)$$

Esta relación expresa la probabilidad $f(v)$ de que una partícula de masa m posea la velocidad v cuando se encuentra a la temperatura T . En nuestro caso se asignó una temperatura de 300 K para todas las simulaciones de la dinámica molecular. Una sola velocidad no puede asignarse a una temperatura dada, ya que ésta es en realidad, una medida de la distribución de velocidades atómicas establecida por la ecuación de Maxwell-Boltzmann, y que tiene interpretación termodinámica solamente en condiciones de equilibrio.

También de la Teoría Cinética de los Gases, se encuentra que la energía cinética, E_{cin} , de un gas ideal es directamente proporcional a la temperatura termodinámica:

$$E_{\text{cin}} = (3/2) \cdot kT \quad (23)$$

donde k es una constante de proporcionalidad (constante de Boltzmann). Por otro lado, la definición de la energía cinética para un átomo se expresa como:

$$E_{\text{cin}} = (1/2) \cdot mv^2 \quad (24)$$

Así, una expresión para la temperatura en términos de la velocidad para un conjunto de N átomos, puede escribirse como:

$$3kT = \sum_{i=1}^N (mv_i^2) / N \quad (25)$$

Esta última ecuación hace posible calcular la temperatura del sistema para un grupo de átomos en movimiento partiendo directamente de sus velocidades.

Existen diferentes formas termodinámicas en las que una simulación molecular puede ser conducida. Una de ellas es el aplicar la mecánica estadística a través de la metodología del conjunto microcanónico, para así mantener fijas la energía total, el volumen y el número de partículas del sistema. En estos casos, el programa de simulación permite variaciones de temperatura por interconversión espontánea de los componentes cinético y potencial de la energía total. Asimismo, cuando es preciso realizar una simulación en la que tanto la temperatura como el número de partículas y el volumen deban permanecer constantes, los programas de simulación de dinámica molecular aplican la aproximación del conjunto canónico para calcular la evolución molecular en el tiempo. En esta segunda modalidad, las velocidades atómicas que determinan directamente la temperatura del sistema, deben ser reescaladas o ajustadas a intervalos regulares de tiempo que el usuario determina. En este trabajo, se empleó una estrategia combinada de simulación molecular, en la que se calcularon trayectorias de 100 fs manteniendo la energía constante, seguidas de reescalamiento de las velocidades atómicas para igualar la temperatura del sistema con la de un baño térmico a 300 K. Esto es, se utilizó una serie de 100 etapas microcanónicas alternadas con ajustes de la temperatura, que dieron como resultado la simulación de un proceso canónico o isotérmico.

Cálculo de áreas superficiales y volúmenes moleculares

La estimación de las áreas moleculares empleadas en la aplicación del modelo de la quimopapaína a la interpretación de algunos resultados experimentales, fue realizada según el método propuesto por Connolly (1983) que se encuentra implementado en el paquete de cómputo BIOGRAF. En estos cálculos todos los átomos fueron considerados como esferas con base en sus radios van der Waals, donde los átomos de hidrógeno se declararon de manera implícita. Para simular al solvente en contacto con la superficie de la proteína se utilizó una esfera de 1.4 Å de radio. Por su parte, el área hidrofóbica se definió como la contribución de los átomos de carbono de la molécula a la superficie total.

La densidad de las proteínas fue estimada del cociente de la masa molecular y el volumen de la estructura. Este último se calculó con el algoritmo propuesto por Stouch & Jurs (1986) con una densidad de 10 celdas por Å, sin incluir a las moléculas de agua. En el presente trabajo se consideró a las **PSV** constituidas por dos dominios estructurales que, de acuerdo a la numeración de quimopapaína, comprenden los residuos 1 a 107 (primer dominio) y 108 a 218 (segundo dominio).

Superposición de estructuras

La superposición de estructuras moleculares de papaína, actinidina, proteinasa Ω y quimopapaína, se realizó según el criterio de minimización del cuadrado de la distancia entre los carbonos α topológicamente equivalentes de las proteasas mencionadas.

RESULTADOS Y DISCUSIÓN

En esta sección se hará una descripción y discusión de los resultados obtenidos durante el modelado de la quimopapaína que involucró, la selección de la estructura base, la sustitución de los residuos de la quimopapaína sobre la proteína base y posteriormente, la validación del modelo con la información estructural disponible. Hacia el final del capítulo se presenta la aplicación de este modelo a la interpretación de algunos datos experimentales.

Selección de la estructura base

a) Homología en composición.

La **figura 3-1** muestra los resultados de la comparación en composición de aminoácidos de la secuencia de la quimopapaína con las correspondientes a las tres proteasas sulfhidrúlicas vegetales de estructura tridimensional conocida. Debe tenerse presente que cuando se obtiene un mayor índice de diferencia en composición de aminoácidos, existe también una mayor divergencia en su homología. Como puede observarse de la **figura 3-1**, la proteinasa Ω es claramente la proteasa sulfhidrúlica con más cercanía a la composición de la quimopapaína, en lo que al número de aminoácidos de cada tipo se refiere. Este resultado se mantiene invariante aún cuando la composición de aminoácidos se ajuste por la longitud de las cadenas, en cuyo caso se obtienen índices de 0.0039, 0.0061 y 0.0106 para la proteinasa Ω , papaína y actinidina, respectivamente.

b) Homología en secuencia.

En la comparación de secuencias alineadas por parejas entre quimopapaína y cada una de las tres proteasas caracterizadas, la proteinasa Ω muestra el mayor número de residuos conservados y la menor cantidad de inserciones o vacancias de aminoácidos (lugares vacíos en la secuencia), **Tabla 3-1**. Este hecho señala a la proteinasa Ω como la estructura más adecuada para ser usada como base en el modelado de la quimopapaína bajo este criterio de selección, ya que su empleo requerirá el menor número de reemplazos de cadenas laterales (73 mutaciones), y evitará la supresión de residuos de la cadena principal. Únicamente precisará de la

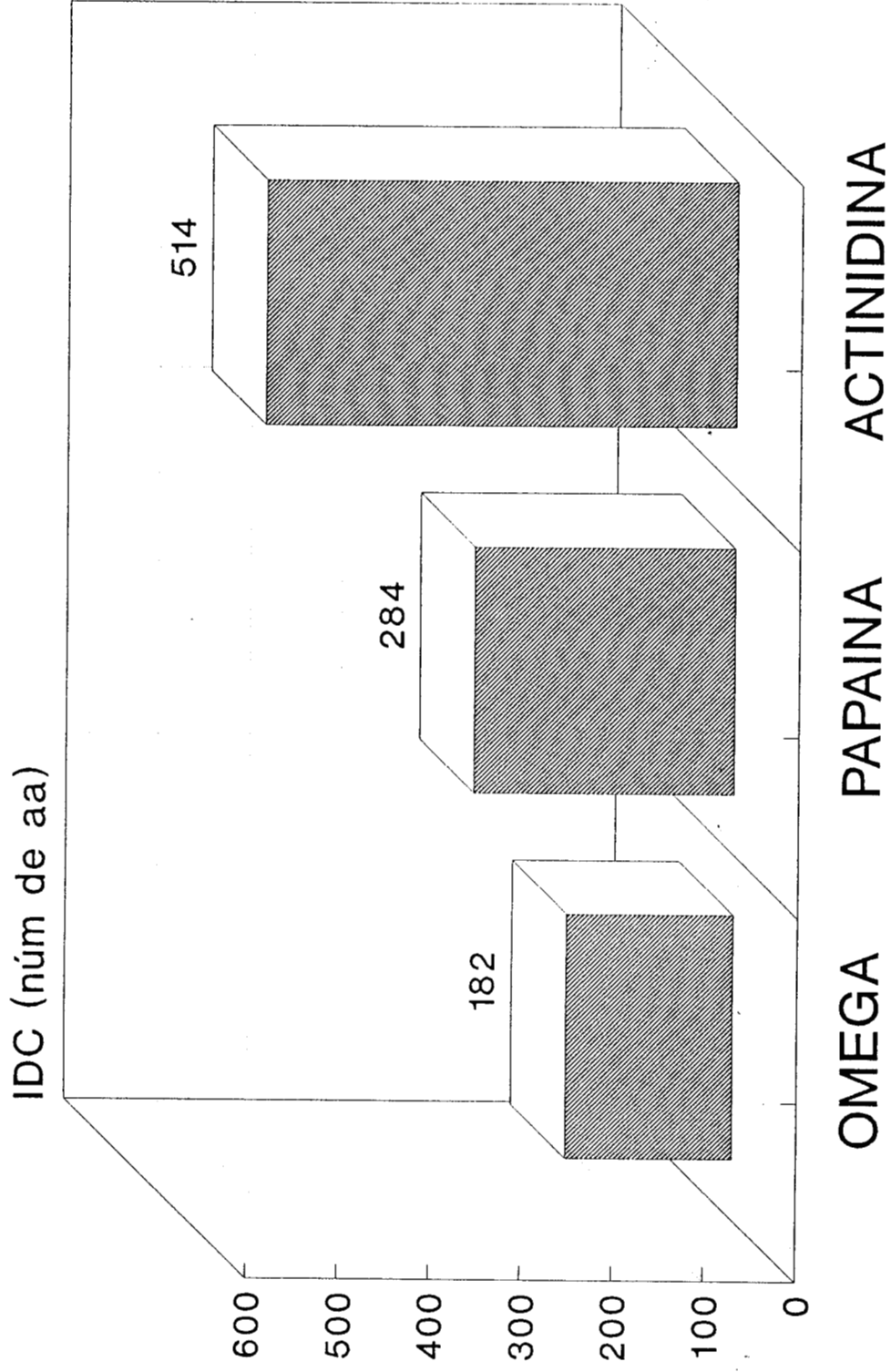


Figura 3-1. Diferencia en composición de aminoácidos entre proteinasa Ω , papaina y actinidina respecto a quimopapaina.

adición de los dos últimos residuos a la cadena, los cuales no tienen análogo en la estructura de base.

TABLA 3-1
**Alineamiento óptimo de secuencias por parejas.
 Comparación con quimopapaína.**

	No. de residuos idénticos	No. de residuos similares	Inserciones*
Proteinasa Ω	143 (66.2%)	23 (10.6%)	0-0
Papaína	126 (59.4%)	30 (14.2%)	1-0
Actinidina	107 (49.1%)	27 (12.4%)	1-2

* La primera cifra corresponde al número de inserciones en la secuencia de la quimopapaína, necesarias para realizar el mejor ajuste. La segunda cifra se refiere a inserciones en la secuencia de la proteína con la que se compara la quimopapaína.

Aunque todos los alineamientos de esta sección se realizaron por parejas, en la **figura 3-2** se muestra el alineamiento simultáneo de las secuencias de quimopapaína y las tres **PSV** cristalizadas para hacer evidentes las zonas de aminoácidos conservados (idénticos y similares), así como para presentar la conectividad de los puentes disulfuro completamente conservados en esta familia según las secuencias reportadas hasta ahora.

c) Significancia de la homología en secuencia.

La dependencia de la significancia en homología con el número de alineamientos de secuencias aleatorias entre las parejas formadas por la quimopapaína y cada una de las **PSV** caracterizadas, se muestra en la **figura 3-3**. Como puede apreciarse, en todas las curvas se presentan grandes fluctuaciones estadísticas en los valores de significancia, cuando se realiza un número bajo de comparaciones (menos de 60). Y por otra parte, se observa que el valor de la significancia resulta notablemente disminuido si este número de comparaciones es demasiado grande. De lo anterior puede concluirse, como ya se había anticipado, que al usar este método estadístico, el número de comparaciones adecuado debería encontrarse entre 80 y 150, ya que a menor muestreo, las fluctuaciones aleatorias pueden modificar sensiblemente el resultado; mientras que por otro lado, un número muy alto de comparaciones enmascara la relación de homología, incluso hasta el

ACTINIDINA	LPSYVDWRSAGAVVDIKSQGECGGCWAFSAIATVEGINKITSGSLISLSEQELID	55
PAPAÍNA	IPEYVDWRQKGAVTPVKNQGSCGSCWAFSAVVTEGIIKIRTGNLNEYSEQELLD	55
PROT.OMEGA	LPENVDRKKGAVTPVRHQGSCGSCWAFSAVATVEGINKIRTGKLVELSEQELVD	55
QUIMOPAPAÍNA	YQSIDWRAKGAVTPVKNQGACGSCWAFSTIATVEGINKIVTGNLLELSEQELVD	55
	* .*** ** . . ** ** ***** . . * .*** ** . * * ***** .*	
ACTINIDINA	CGRTQNTRGCDGGYITDGFQFIINDGGINTQENYPYTAQDGDQVALQDQKYVTI	110
PAPAÍNA	CDR--RSYGCNGGYPWSALQLV-AQYGIHYRNTYPYEGVQRYCRSREKGPYAAKT	107
PROT.OMEGA	CER--RSHGCKGGYPPYALEYV-AKNGIHLRSKYPYKAKQGTCTRAKQVGGPIVKT	107
QUIMOPAPAÍNA	CDK--HSYGCKGGYQTTSLQYV-ANNGVHTSKVYPYQAKQYKCRATDKPGPKVKI	107
	* ** .*** * *** . . *	
ACTINIDINA	DTYDNVPYNNEWALQTAVTYQPVSVVALDAAGDAFKQYASGIFTGPCGTAVDHAIV	165
PAPAÍNA	DGVRQVQPYNEGALLYSIANQPVSVVLEAAGKDFQLYRGGIFVGPCKNKVDHAVA	162
PROT.OMEGA	SGVGRVQPNNEGNLLNAIAKQPVSVVVESKGRPFQLYKGGIFEGPCGTKVDHAVT	162
QUIMOPAPAÍNA	TGYKRVPNCETSFLGALANQPLSVLVEAGGKPFQLYKSGVFDGPGTKLDHAVT	162
	. . . * . * ** .** . . . * . * . * . * .*** . .*** . .	
ACTINIDINA	IVGYGTEGGVDYWIVKNSWDTTWGEEGYMRI LRNVGGA-GTCGIATMPSYPVKYNN	220
PAPAÍNA	AVGYGPN---YILIKNSWGTGWGENGYIRIKRGTGNSYGVCGLYTSSFYPVKN--	212
PROT.OMEGA	AVGYGKSGGKGYILIKNSWGTAWGEGYIRIKRAPGNSPGVCGLYKSSYPTKN--	216
QUIMOPAPAÍNA	AVGYGTS DGKNYIIIKNSWGPNWGEKGYMRLKRQSGNSQGTGCVYKSSYYPFKGFA	218
	**** . * . .**** . .*** .** . * * * . . * .** . . . ** *	

Figura 3-2. Alineamiento de cuatro PSV donde se muestra la posición de los enlaces disulfuro completamente conservados en la familia. Los asteriscos representan zonas de residuos idénticos y los puntos residuos similares.

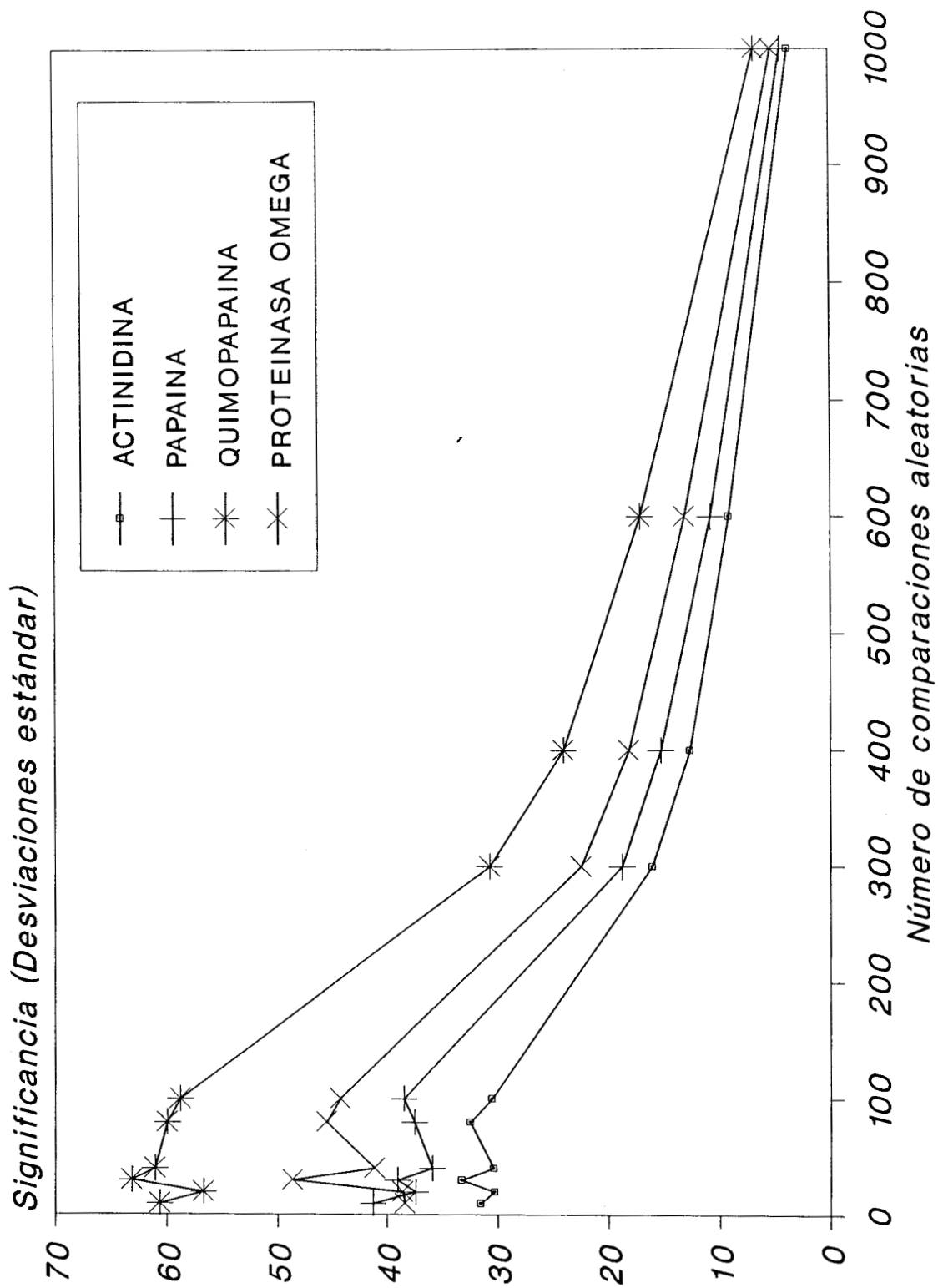


Figura 3-3. Significancia de la homología en secuencia entre quimopapaina y cada una de las PSV cristalizadas. Se incluye la significancia de la homología de quimopapaina consigo misma.

grado en que el marcador de la comparación de la secuencia de quimopapaína consigo misma, tiende a valores muy pequeños (**Fig. 3-3**). Independientemente del efecto descrito, el resultado principal de la aplicación de la significancia en homología como tercer criterio de selección es que éste señala a la proteinasa Ω como la enzima con mayor cercanía topológica a la quimopapaína. Por lo tanto, bajo los tres criterios de comparación estadística empleados, la proteinasa Ω parece ser la mejor de las tres estructuras de **PSV** caracterizadas para utilizarse como molde de la quimopapaína, ya que así, la construcción del modelo involucrará tanto el menor número de mutaciones y la ausencia de modificaciones internas en la cadena principal (sólo su elongación en dos residuos en su extremo amino terminal), como la mayor relación entre las secuencias de la estructura de base y la proteína por modelarse.

Modelado molecular de la quimopapaína

a) Construcción del modelo inicial.

Una vez seleccionada la proteinasa Ω como la estructura tridimensional de partida en el trabajo de modelado, se procedió a efectuar las sustituciones de residuos de aminoácido pertinentes. Según el alineamiento de secuencias de la proteinasa Ω con la quimopapaína de la **figura 3-4**, se requirió la sustitución de 73 residuos de aminoácido (señalados con |) y como ya se había mencionado, la adición de dos residuos en el extremo carboxilo de la cadena polipeptídica de la proteinasa Ω para completar la secuencia de la quimopapaína. Todos los cambios se realizaron por modificación de las cadenas laterales de los residuos que resultaban diferentes en proteinasa Ω respecto a quimopapaína. Para llevar a cabo estas mutaciones, se emplearon las coordenadas cristalográficas de la proteinasa Ω , generosamente proporcionadas por el autor de la determinación de la estructura tridimensional de esta proteasa (Pickersgill *et al.*, 1991). Cada nueva cadena lateral fue construida de tal manera que mantuviera sus ángulos diedros cercanos a los observados en el aminoácido libre, pues se sabe que, en general, estos ángulos no se modifican sensiblemente cuando forman parte de una proteína plegada (Ponder & Richards, 1987; Levitt, 1992). La mayor parte de las mutaciones involucraron cadenas laterales accesibles al disolvente o colocadas en regiones de cadena aleatoria de la proteína (85%), y dado que no se requirió la inserción o remoción de segmentos de la cadena principal, como se había mencionado antes, el proceso de construcción de la estructura inicial sólo precisó de la reorientación interactiva de un

PROT. OMEGA	LPENVDWRKKGAVTPVRHQGSCGSCWAFSAVATVEGINKIRTKGLKLVSEQLVD	55
QUIMOPAPAÍNA	YQSIDWRAKGA VTPVKNQGACGSCWAFSTIATVEGINKI VTGNLLELSEQLVD	55
PROT. OMEGA	CERRSHGCKGGYPPYALEYVAKNGIHLRSKYPYKAKQGTCTRAKQVGGPIVKTSGV	110
QUIMOPAPAÍNA	CDKHSYGCKGGYQTTSLQYVANNGVHTSKVYPYQAKQYKCRATDKPGPKVKITGY	110
PROT. OMEGA	GRVQPNEGILLNAIAKQPVSVVVEKGRPFQLYKGGIFEGPCGTKVDHAVTAVG	165
QUIMOPAPAÍNA	KRVPSNCETSFLGALANQLSVLVEAGGKPFQLYKSGVFDGPGCGTKLDHAVTAVG	165
PROT. OMEGA	YGKSGGKGYILIKNSWGTAWGEKGYIRIKRAPGNPVGGLYKSSYYPTKN	216
QUIMOPAPAÍNA	YGTSDGKNYIIIKNSWGNWGEKGYMRLKRQSGNSQGTGCVYKSSYYPFKGFA	218

Figura 3-4. Alineamiento de secuencias de proteinasa Ω y quimopapaína donde se indican los residuos no conservados entre ambas proteasas.

número limitado de residuos. En esta reorientación se buscó disminuir los contactos van der Waals entre la nueva cadena lateral y el resto de la estructura, como se ha realizado en otros trabajos de modelado molecular (Blundell, 1991; Srinivasan, 1993), a fin de evitar tensiones excesivas antes de intentar la regularización global de la geometría por minimización de energía. Los sitios que presentaron los mayores problemas de compatibilidad entre la proteinasa Ω y la nueva estructura generada durante el proceso de sustitución de las cadenas laterales fueron: Ser84→Lys, Val100→Lys, Ile104→Lys y Val110→Tyr. En todos estos casos los residuos se encontraban en zonas de cadena aleatoria y sólo fue necesario reorientar las nuevas cadenas laterales al exterior de la proteína.

b) Minimización de Energía.

Con el propósito de disminuir la tensión provocada por interacciones de alta energía en la molécula, se procedió a minimizar la energía potencial de la misma a través de una serie de etapas de mecánica molecular. Se sabe que, en general, la sustitución de residuos sobre una estructura tridimensional dada genera tensiones locales debido a problemas de estericidad, y que en estas condiciones una minimización global sólo propagaría las distorsiones de los sitios de mayor tensión hacia el resto de la molécula. Por esto, resulta recomendable identificar y minimizar esas regiones previamente a cualquier proceso de relajamiento masivo. Cuando se emplea BIOGRAF para realizar mutaciones sobre cadenas polipeptídicas, los residuos de prolina involucrados (tanto para la sustitución Pro→Xaa como para Xaa→Pro) son los más notables en cuanto a la creación de tensiones puntuales. De esta manera se efectuaron minimizaciones individuales para los residuos Pro68→Gln, Pro69→Thr, Gly101→Pro, Gln114→Pro, Pro115→Ser, Thr183→Pro, Pro197→Ser y Pro201→Gln, mediante la relajación de sus cadenas laterales, seguida de la minimización de todos los átomos del residuo y, finalmente, de la regularización del residuo completo junto con cuatro vecinos en secuencia (dos previos y dos posteriores). En cada caso se requirieron entre 100 y 200 iteraciones para alcanzar un mínimo de energía potencial.

La siguiente etapa consistió en la minimización simultánea de todas las cadenas laterales, manteniendo fijas las posiciones de los átomos de la cadena principal y de las 134 moléculas de agua que pertenecían a la proteinasa Ω . Esto permitió que las cadenas laterales pudieran reorganizarse energéticamente en el

espacio, sin modificar la topología de la estructura de base. Previos a esta etapa ocurrieron algunos imprevistos: las coordenadas cristalográficas de la proteinasa Ω no cubrían los requisitos del formato **PDB** y fue necesario editar el archivo de 2198 átomos para establecer todos los dobles enlaces presentes, así como cambiar el tipo de átomo de los nitrógenos en los grupos amino de la cadena principal y modificar el residuo carboxilo terminal de la proteína. Además se presentaron traslapes entre las cadenas laterales de la Tyr93 y del Asn 173 con dos moléculas de agua, por lo que éstas tuvieron que ser removidas del sistema. Una vez realizadas todas estas modificaciones, se sometió la estructura a 100 etapas de minimización, con lo que la energía potencial disminuyó de 6861 a 838.5 *kcal/mol*.

El paso siguiente en la minimización de la estructura fue el incluir a las moléculas de agua en el conjunto de átomos móviles que comprendía a las cadenas laterales. Este proceso llevó a la energía potencial de un valor inicial de 1358 hasta -364.8 *kcal/mol* en 210 etapas de minimización. Finalmente, se procedió a la minimización global, permitiendo libertad de movimiento a la totalidad de los átomos de la molécula. En este caso 210 etapas de minimización condujeron a la energía potencial de 396.4 hasta un valor de -1429 *kcal/mol*.

c) Convergencia hacia la conformación de equilibrio.

La primera etapa en la simulación de la dinámica molecular de la quimopapaína involucró el modelado de los dos últimos residuos de esta proteína, debido a que como se mencionó, fueron adicionados en ausencia de aminoácidos topológicamente equivalentes en la estructura de base. En este proceso se permitió el movimiento a los cinco últimos residuos de la cadena polipeptídica (Phe214, Lys215, Gly216, Phe217 y Ala218) y también a sus átomos vecinos en un radio de 5 Å. Esta dinámica se realizó con una primera simulación de 150 *ps* a la temperatura de 300 *K*, seguida de la remoción de 8 de las 10 moléculas de agua incluidas en la región de átomos móviles (lo cual permitió la interacción de los residuos 217 y 218 con el resto de la proteína), y de una minimización de energía. Finalmente, se simuló una nueva dinámica molecular de esta zona, también de 150 *ps* a 300 *K*.

Un proceso de minimización de la quimopapaína precedió a la dinámica molecular global de la estructura, simulada en condiciones de temperatura de 300 *K* durante 60 *ps*. Al confórmero de menor energía en la trayectoria calculada en esta

etapa, se le aplicó una nueva minimización de energía de 60 etapas, lográndose las energías potenciales más bajas hasta este punto del modelado. Sin embargo, al analizar la estereoquímica de este conformero, se detectaron los carbonos α de seis residuos en configuración *D* y dos enlaces peptídicos en conformación *cis*. Antes de pasar a la siguiente etapa de dinámica molecular, se corrigieron todas estas anomalías geométricas, con excepción del enlace peptídico previo a la Pro151 que se mantuvo en conformación *cis*, por ser ésta una característica común observada en las proteasas sulfhidríticas vegetales cristalizadas. Además, se removieron seis moléculas de agua debido a que espontáneamente se separaron del resto del sistema (evaporación) durante el proceso de simulación dinámica. La estructura resultante fue sometida a un nuevo cálculo de minimización de energía, y después a una dinámica molecular de 40 *ps*. De los conformeros obtenidos en este cálculo, se seleccionó el de menor energía potencial, y sobre él se definieron de manera explícita la totalidad de los átomos de hidrógeno (1230) del sistema molecular en estudio. La estructura resultante mostró todos sus residuos en conformación *L* y fue minimizada para regularizar la geometría en presencia de los nuevos átomos agregados (hidrógenos). Posteriormente, se realizaron dos procesos de dinámica molecular, de 35 y 105 *ps*, separados por una nueva minimización de energía. De la dinámica de 105 *ps* se eligieron a los tres conformeros que mostraron la menor energía potencial del conjunto; éstos fueron minimizados y, finalmente, de entre ellos se seleccionó a la geometría de más baja energía para representar a la estructura tridimensional de la quimopapaína. En la **figura 3-5** se esquematiza la naturaleza flexible del modelo obtenido, a través de la superposición de las cadenas α de los diferentes conformeros obtenidas de la simulación de su dinámica molecular de 105 *ps*. Es conveniente señalar que para generar cada una de las estructuras que aparecen en esta **figura 3-5**, el programa calculó en realidad 6000 conformaciones intermedias, para las que fue necesario evaluar cada vez las interacciones entre los 3645 átomos del sistema. Las coordenadas del modelo de la quimopapaína obtenidas después de esta larga serie de etapas, fueron interpretadas por el sistema de despliegue gráfico del paquete de cómputo BIOGRAF, como se ilustra en la **figura 3-6**.

Evaluación de la calidad estereoquímica del modelo

En un estudio sobre más de 400 proteínas con estructura tridimensional conocida, Morris *et al.*, (1992) obtuvieron las tendencias conformacionales generales

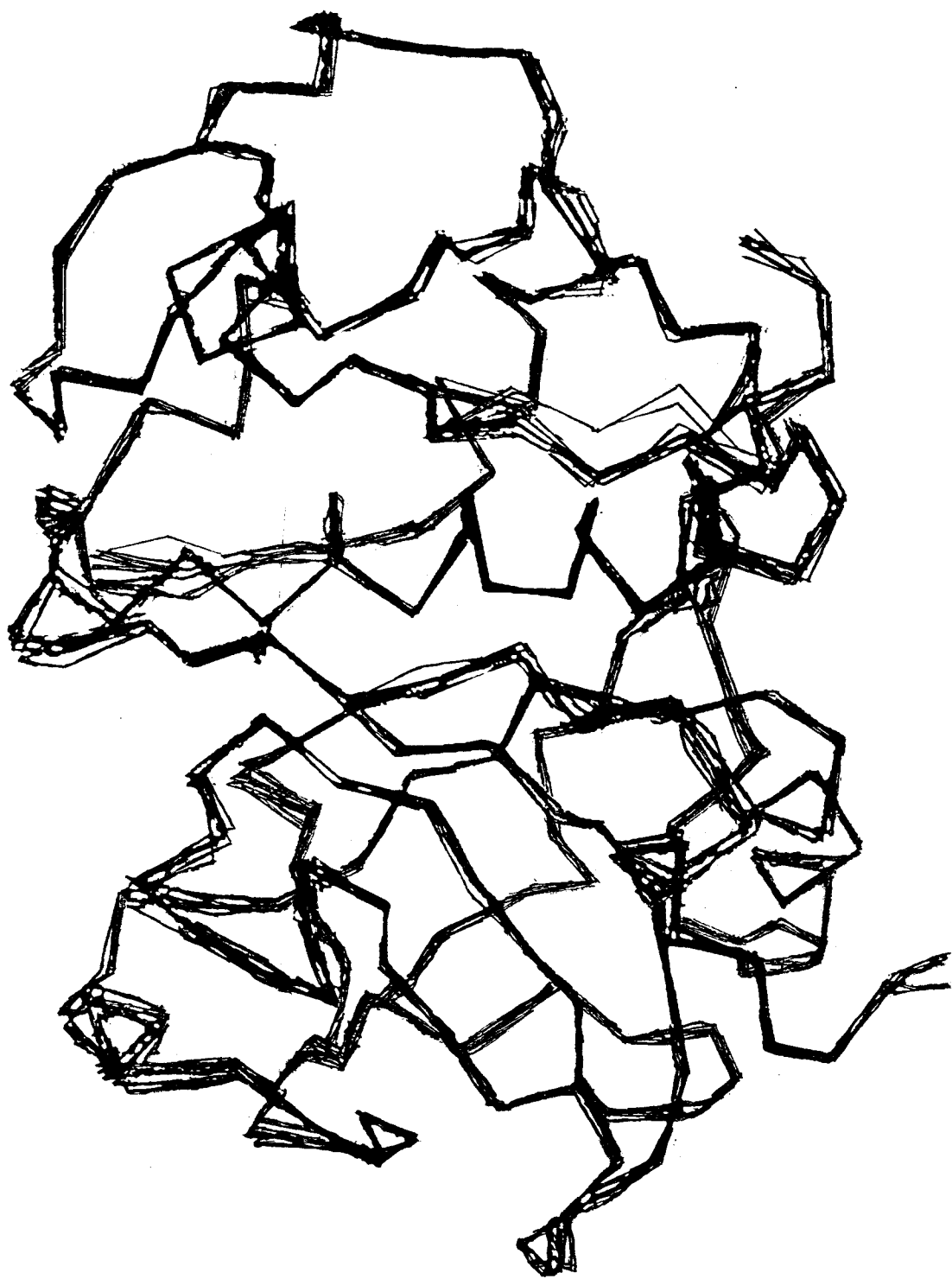


Figura 3-5. Superposición de quince conformeros obtenidos en la simulación de la dinámica molecular del modelo de la quimopapaína.

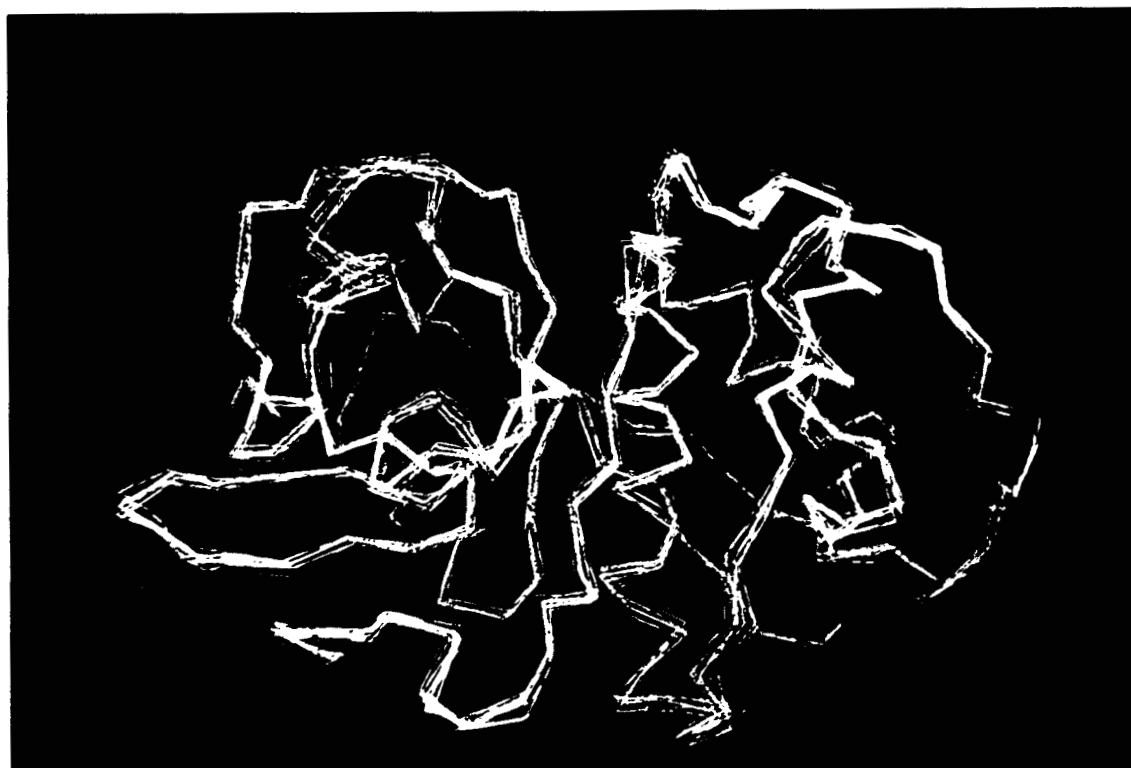
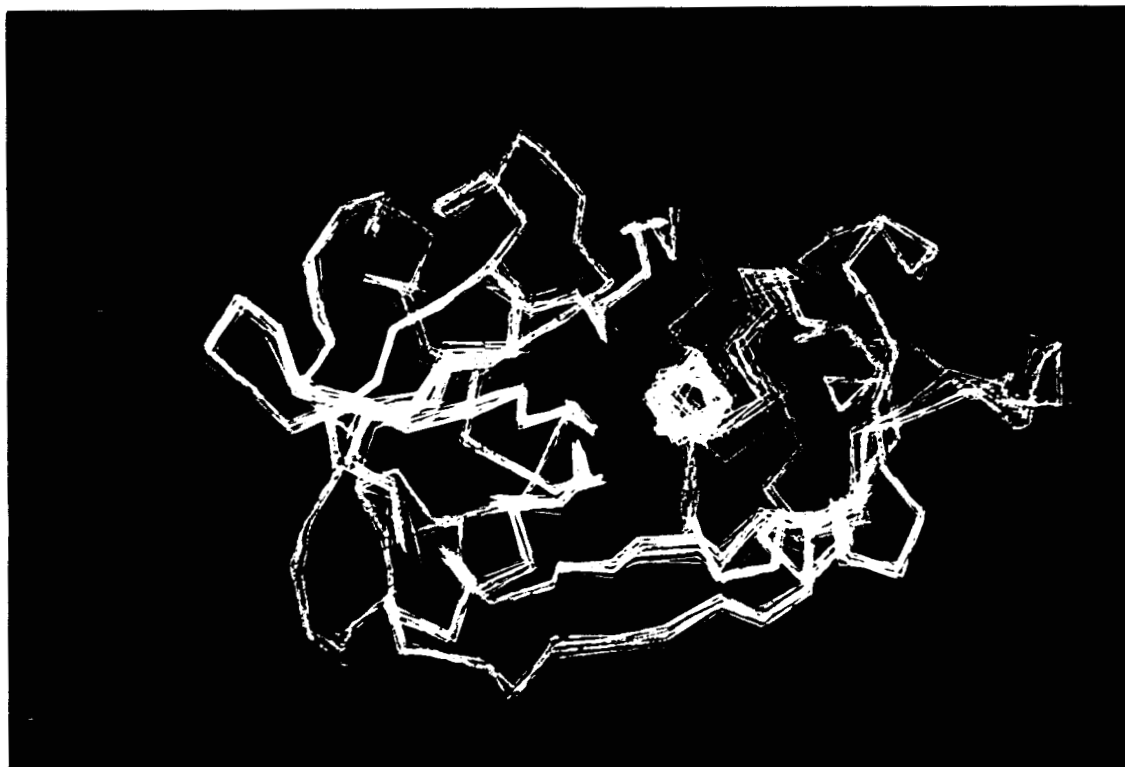


Figura 3-5 (continuación). Dos aspectos de la naturaleza dinámica del modelo de la quimopapaina.

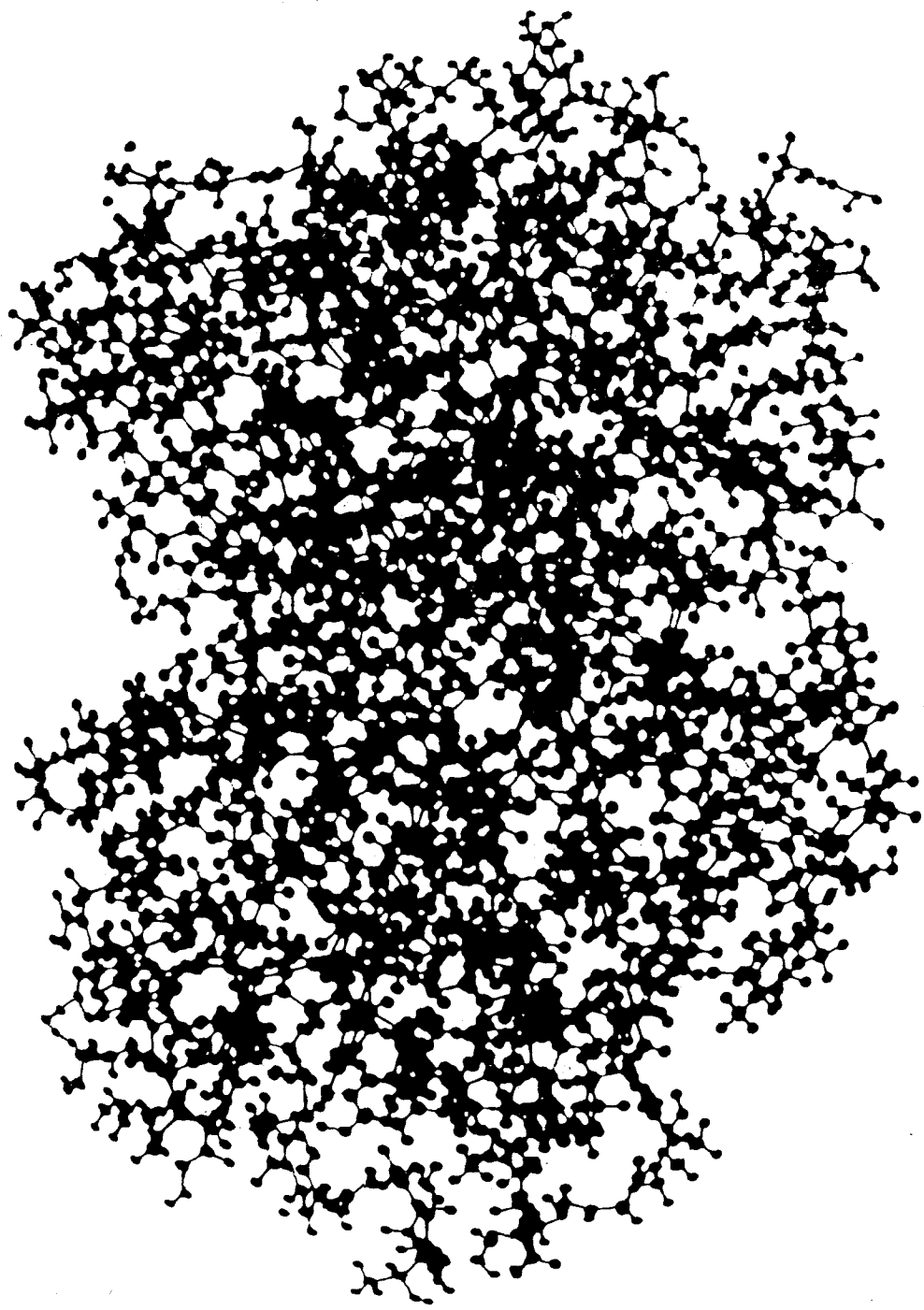


Figura 3-6. Esquema de esferas y bastones del modelo de la quimopapaina.

que seguía la geometría de las proteínas consideradas y la correlación que podía establecerse entre esta geometría, la resolución y el factor R cristalográfico para cada proteína del conjunto estudiado. Entre sus resultados más importantes, estos investigadores definieron tres zonas de incidencia sobre un mapa de Ramachandran que incluye a los 121,870 residuos de aminoácido de 462 proteínas y donde se han obviado a los residuos de prolina y glicina. Las zonas se distinguen entre sí por la densidad de residuos registrada en cada una de ellas y son clasificadas como regiones nucleares (**C**), observadas (**G**) y permitidas (**A**) o también de alta, media y escasa densidad de población, respectivamente (**Fig. 3-7**). Respecto a la correlación entre la geometría, la resolución y el factor R de las proteínas analizadas, Morris *et al.* (1992), encontraron que aquellas estructuras determinadas con buena resolución y con factores R pequeños, presentaban un alto porcentaje de sus residuos dentro de las zonas nucleares en el mapa de Ramachandran (**Fig. 3-8**). El trabajo citado también incluye la determinación de intervalos de confianza para los valores de algunos elementos geométricos, que deben ser observados por las estructuras tridimensionales de alta resolución. Estos elementos comprenden: ángulos diedros Φ y Ψ de residuos que forman parte de hélices en la proteína; ángulos χ_3 de puentes disulfuro; ángulos Φ en prolinas; ángulos ω de todos los residuos del polipéptido; la quiralidad de los carbonos α y de las cadenas laterales de residuos de Thr e Ile; la conformación de los enlaces peptídicos e incluso, los valores de energía promedio que pueden adquirir las interacciones de puentes de hidrógeno en el afinamiento de estructuras cristalográficas (**Tabla 3-II**). El trabajo de Morris *et al.* (1992) es tan completo, que actualmente se utilizan sus criterios para la aceptación de estructuras tridimensionales en las bases del Banco de Datos Cristalográficos de Proteínas (Brookhaven National Laboratory, 1992).

Para la evaluación de la calidad estereoquímica del modelo de la quimopapaína propuesto en el presente trabajo, se sometió la estructura de esta proteína al juicio de los criterios del **PDB** mencionados en el párrafo anterior, y los resultados obtenidos se detallan a continuación.

Del total de los 178 residuos que contiene la quimopapaína al excluir prolinas y glicinas, el 70% de las parejas de ángulos Φ y Ψ se encontraron dentro de las zonas o regiones nucleares según se muestra en el mapa de Ramachandran construido para la quimopapaína en la **figura 3-9**. De acuerdo con Morris *et al.*

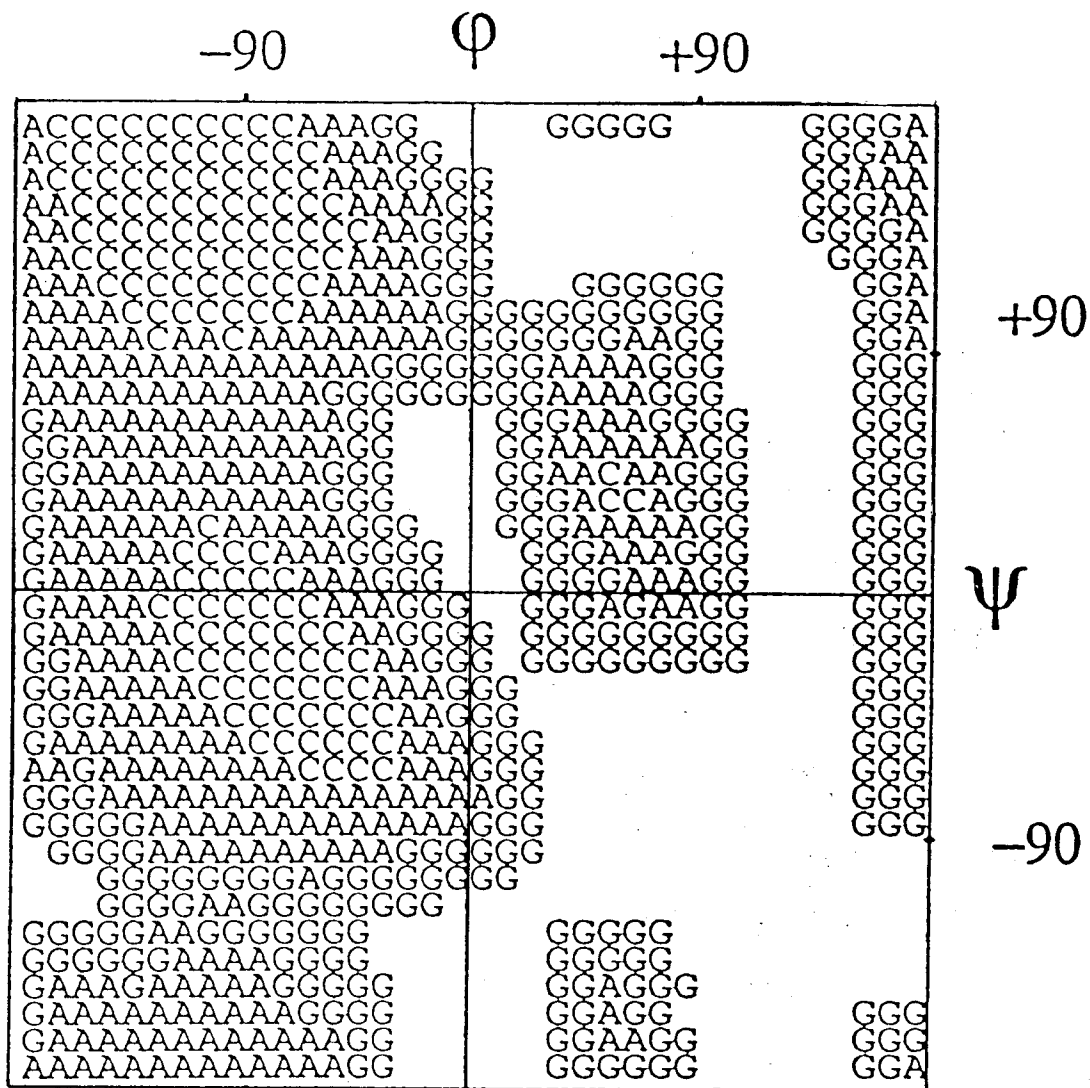


Figura 3-7. Regiones nuclear (C), observada (G) y permitida (A) sobre la densidad de puntos en el mapa de Ramachandran de 121,870 residuos, de acuerdo con Morris *et al.* (1992).

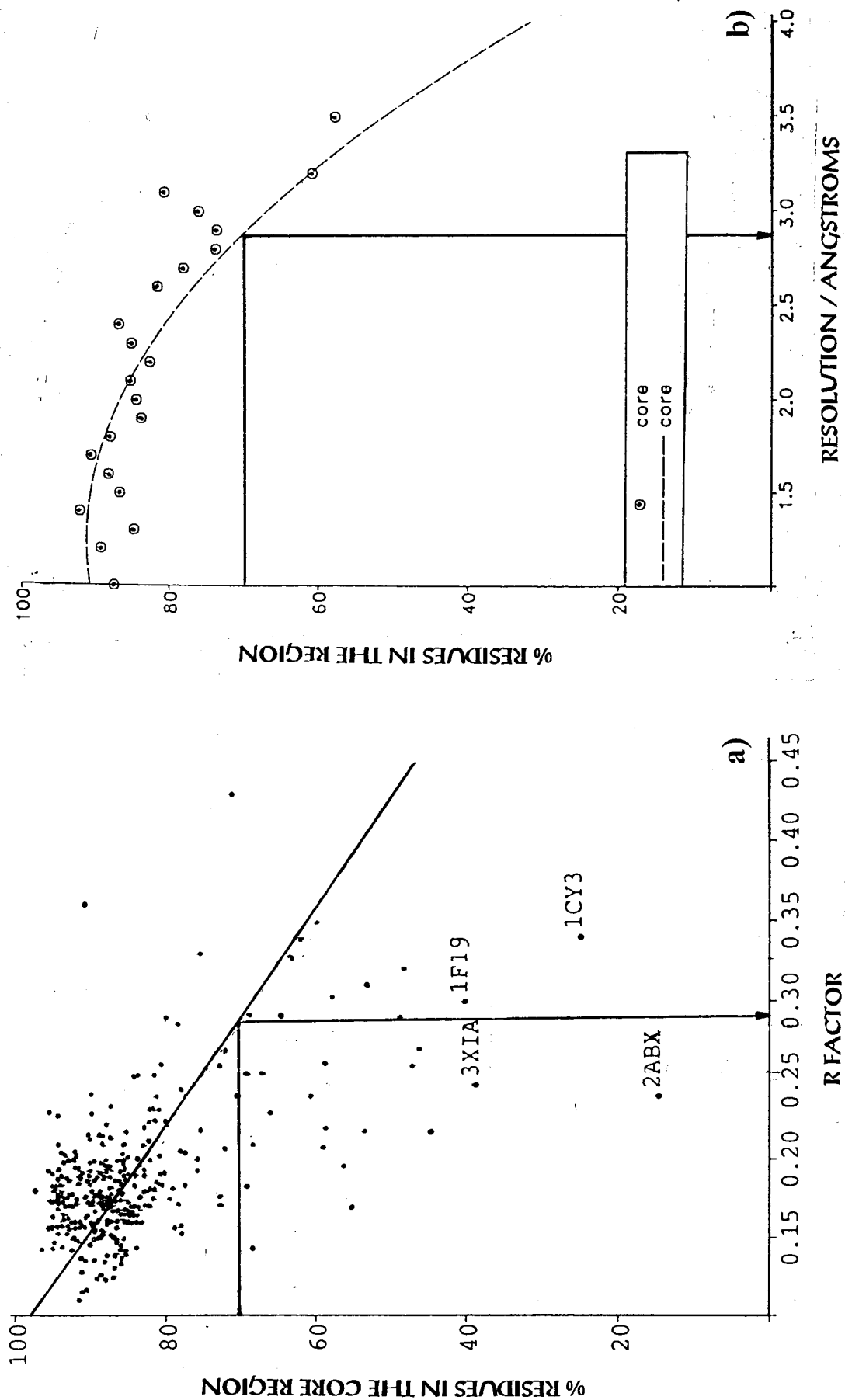


Figura 3-8. Correlación entre el porcentaje de residuos en las regiones nucleares del mapa de Ramachandran y: (a) el factor R para 360 estructuras, (b) la resolución para 462 proteínas; de acuerdo con Morris *et al.* (1992). En ambos casos se indica la interpolación del 70% de residuos en la zona nuclear encontrado para el modelo de la quimopapaina.

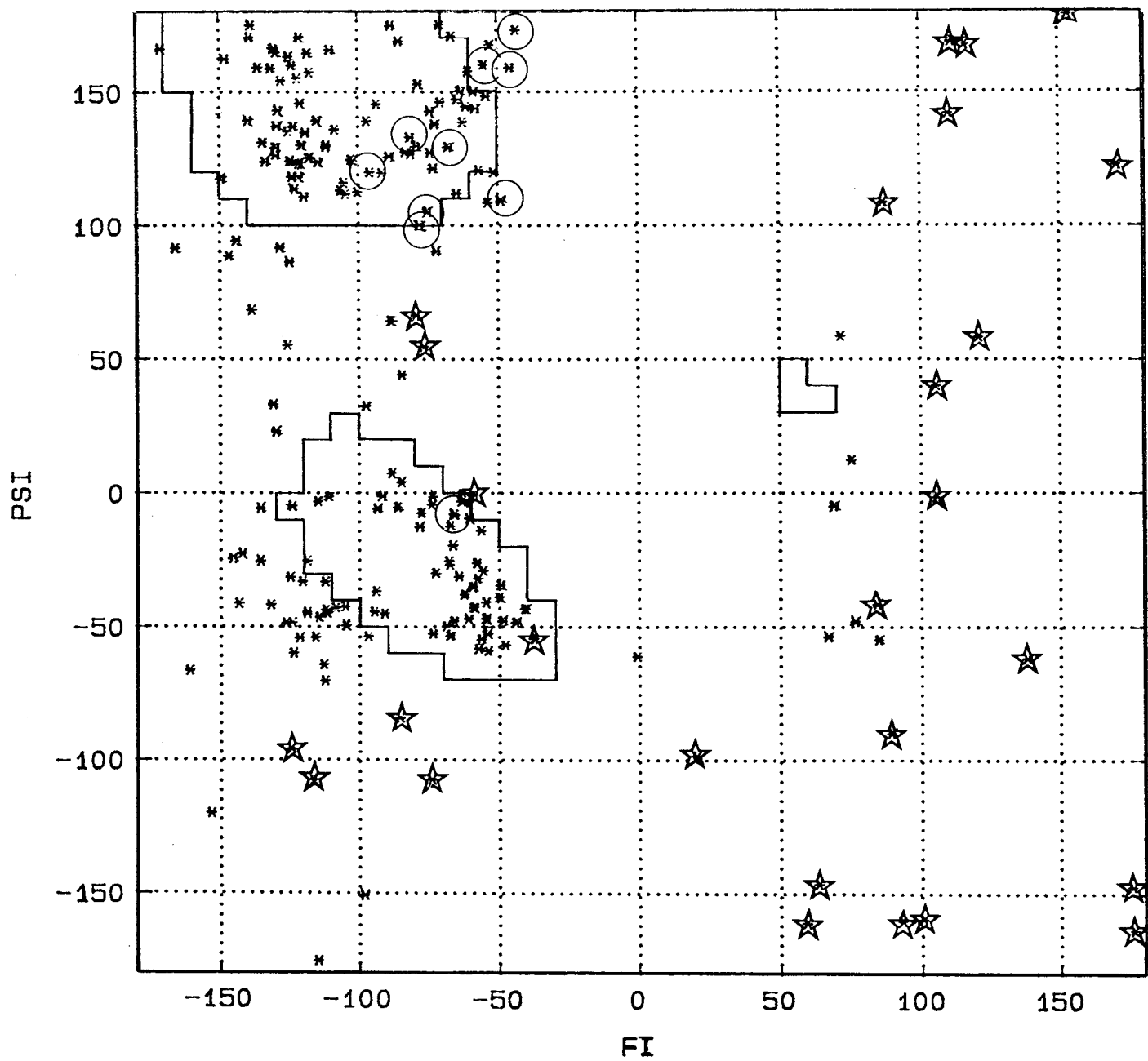


Figura 3-9. Mapa de Ramachandran para el modelo de la chimopapaína. Las líneas continuas delimitan las regiones nucleares de la figura 3-7. Los residuos de prolina y glicina están identificados por ○ y ☆, respectivamente.

(1992), este porcentaje de residuos corresponde a una resolución de 2.9 Å y a un factor R de 0.28 en estructuras de proteínas obtenidas cristalográficamente (Fig. 3-8). Para evaluar la posición del 30% de residuos externos a las zonas nucleares, se representaron las regiones observadas y permitidas de Morris *et al.* sobre el mapa de Ramachandran del modelo de la quimopapaína y se encontró que excepto por cuatro aminoácidos, las regiones observadas cubrían el total de los residuos de la proteína (Fig. 3-10). Es importante señalar que los cuatro aminoácidos no considerados en las regiones observadas fueron la Ala90, la Asn116 y la Ser200 que se encuentran en cadena aleatoria, mientras que la Ala97 aparece al inicio de una hélice pequeña en el primer dominio de la quimopapaína.

TABLA 3-II
Valores de los parámetros estereoquímicos para estructuras cristalográficas de alta resolución.

• Ángulos Φ y Ψ en las zonas de mayor densidad	> 90%
• Ángulo χ_3 en puentes S-S Izquierdos	$-85.8 \pm 10.7^\circ$
Derechos	$96.8 \pm 14.8^\circ$
• Ángulo Φ en prolinas	$-65.8 \pm 11.2^\circ$
• Ángulos helicoidales	
Φ	$-65.3 \pm 11.9^\circ$
Ψ	$-39.4 \pm 11.3^\circ$
• Energía de puentes de hidrógeno	$-2.03 \pm 0.75 \text{ kcal/mol}$
• Enlaces peptídicos	Conformación <i>trans</i>
• Residuos	Isómeros <i>L</i>

Los ángulos diedros χ_3 de los tres puentes disulfuro en la molécula de quimopapaína se encontraron dentro de los límites fijados por el criterio de evaluación estereoquímico. Los valores numéricos de estos ángulos diedros son: -100.06° para el enlace disulfuro formado entre los residuos de cisteína 22 y 63; 91.04° para el que se encuentra entre las cisteínas 56 y 95; y finalmente -77.80° entre los residuos 153 y 204.

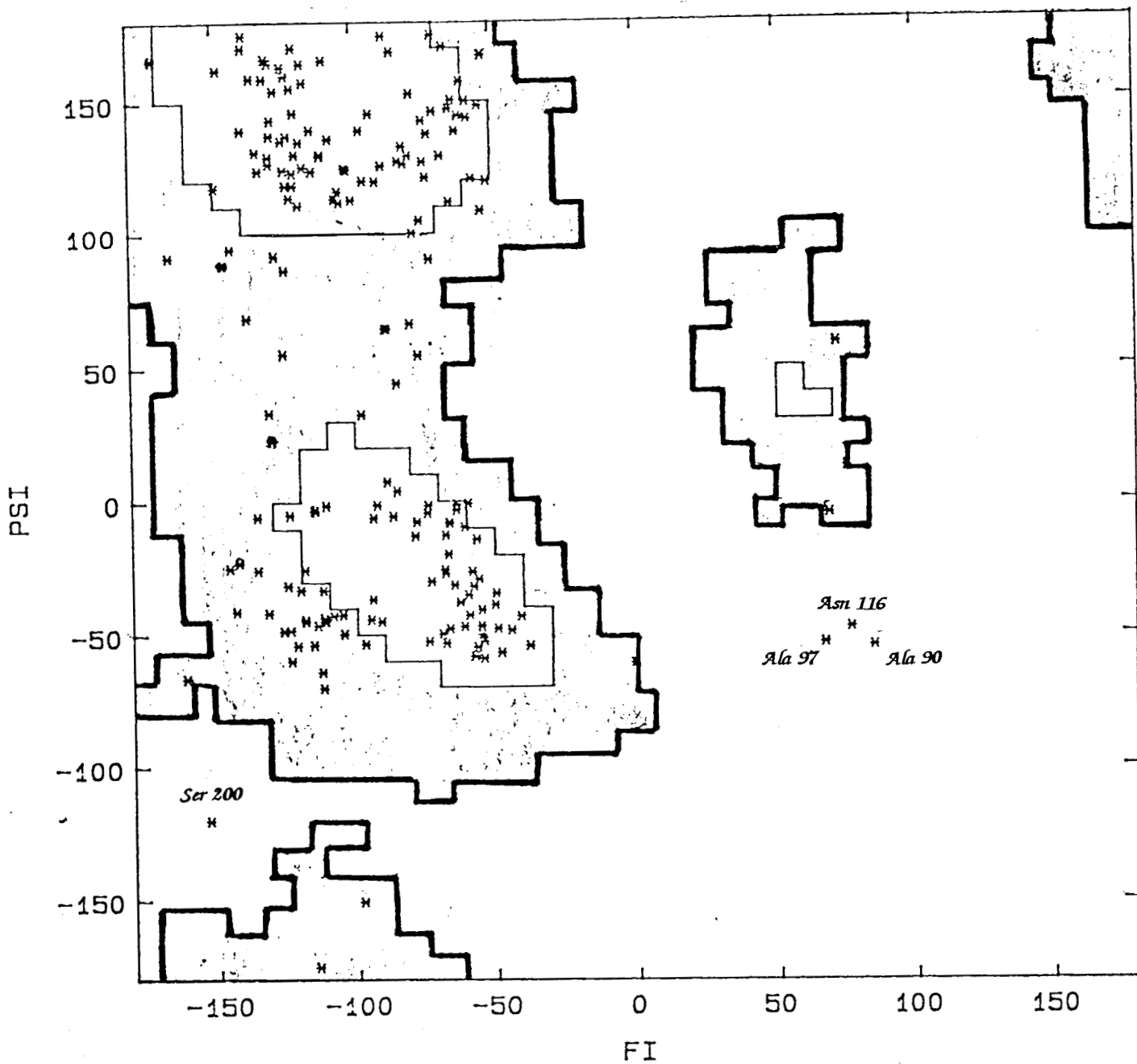


Figura 3-10. Regiones nucleares y observadas de Morris *et al.* (1992) sobre el mapa de Ramachandran del modelo de la quimopapaina, donde se han obviado los residuos de glicina y prolina, y se han identificado los cuatro residuos que aparecen fuera de las zonas observadas.

Respecto a los ángulos diedros Φ de los residuos de prolina en el modelo de la quimopapaína, el 80% incide dentro de los intervalos establecidos por el criterio de evaluación y el resto se encuentra desviado en todos los miembros de la familia de las **PSV** con estructura tridimensional conocida.

Del análisis geométrico del modelo de la quimopapaína se determinó que todos los ángulos ω se encuentran en conformación *trans* a excepción del enlace peptídico previo a la Pro151, que es un residuo completamente conservado en la familia a la que pertenece la quimopapaína y que se encuentra en conformación *cis* en todos los casos conocidos, como ya se había mencionado. También del análisis geométrico, se conoce que todos los carbonos α de la cadena principal de la proteína presentan el isómero *L* y que las cadenas laterales de las treoninas e isoleucinas de la cadena polipeptídica cumplen con la quiralidad observada en los aminoácidos naturales (Morris *et al.*, 1992). El hecho de que todos los residuos de la quimopapaína hayan conservado la configuración *L* durante las últimas etapas de la dinámica molecular, es de la mayor importancia en el caso de una estructura modelada, ya que cualquier inconsistencia geométrica en el interior de la proteína podría haber ocasionado la inversión de algunos centros quirales durante la simulación.

Debido a que cualquier cálculo de energía depende del programa específico que se emplee para afinar una estructura cristalográfica, el valor numérico de la interacción que se propone en el criterio estereoquímico para los puentes de hidrógeno, no resulta comparable directamente con la energía que se calcula para la estructura modelada de la quimopapaína. De manera que fue preciso realizar un proceso de minimización de energía que involucrara por una parte, el mismo programa de mecánica molecular empleado en el caso de la quimopapaína, y por otra, a las proteasas sulfhidríticas vegetales cuyas estructuras tridimensionales se encuentran almacenadas en el Banco de Datos Cristalográficos de Proteínas. Se decidió así trabajar con las estructuras de papaína, actinidina y proteinasa Ω , después de agregar a estas estructuras todos los átomos necesarios (átomos de hidrógeno que no están incluidos en los archivos cristalográficos del **PDB**) para los cálculos energéticos de puentes de hidrógeno. El resultado numérico de la energía promedio para estas interacciones, en las tres proteasas arriba mencionadas, presenta una

desviación menor al 1.5% respecto a su valor medio (-6.45 kcal/mol). Para el caso de la quimopapaína, la energía por puentes de hidrógeno se encuentra a sólo 1.1% del valor medio observado en **PSV**, lo cual permite considerar satisfactoria la geometría de los puentes de hidrógeno presentes en el modelo generado, pues la desviación del valor medio para estas interacciones, encontrada por Morris *et al.*, es de 37% en estructuras cristalográficas de alta resolución.

Todos los resultados de la evaluación de las características geométricas del modelo de la quimopapaína con los estándares empleados para juzgar la calidad estereoquímica en proteínas caracterizadas por difracción de rayos X, que se presentan en la **Tabla 3-III**, permiten considerar al modelo propuesto en este trabajo como una estructura que cubre satisfactoriamente los requerimientos tridimensionales correspondientes a los de una proteína de buena resolución cristalográfica.

TABLA 3-III
Análisis estereoquímico del modelo de la quimopapaína

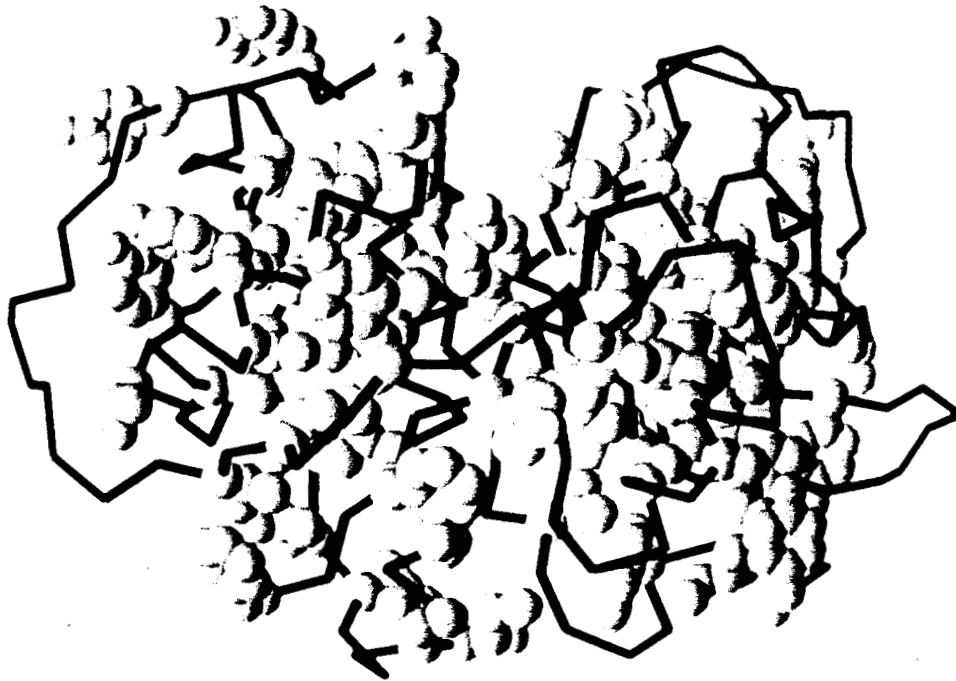
	Cumple	No cumple	%
Residuos con ángulos Φ y Ψ en las zonas nucleares	124	54	70
Ángulos χ_3 en disulfuros	3	0	100
Ángulo Φ en prolina	11	0	100
Ángulos helicoidales	54	11	80
Enlaces peptídicos <i>trans</i>	216	1 (prolina151)	99.5
Residuos en configuración <i>L</i>	218	218	100
Energía de puentes de hidrógeno	<i>(kcal/mol)</i>		
<ul style="list-style-type: none"> • Quimopapaína • Papaína • Actinidina • Proteínasa Ω 	<ul style="list-style-type: none"> -6.38 -6.37 -6.47 -6.52 		
Resolución	2.9 Å		
Factor <i>R</i>	0.28		

Análisis de la estructura del modelo de la quimopapaína

De la revisión del modelo propuesto para la quimopapaína se puede apreciar que la arquitectura cumple las características generales observadas en proteínas. Esto es, la mayor parte de sus cadenas laterales hidrofóbicas se encuentran dirigidas al interior de la molécula, mientras las hidrofílicas se localizan fundamentalmente en la superficie (**Fig. 3-11**). Adicionalmente, el modelo resultó ser consistente con algunos resultados experimentales que reflejan directamente ciertas características de la estructura tridimensional de esta proteína, como son: el ambiente polar del triptofano 181 (Dubois *et al.*, 1988b); la exposición al disolvente de la cisteína 117 (Watson *et al.*, 1990) que se observa en la **figura 3-12**; la distancia mayor de 20 Å entre ese residuo y la otra cisteína libre en la molécula (Topham *et al.*, 1990a); y la semejanza de regiones interdominio (Baker & Drenth, 1987; **Fig. 3-13**) y de la geometría del sitio activo (Carey *et al.*, 1983) con los otros miembros de las **PSV**. Respecto a este último punto, la **Tabla 3-IV** muestra comparativamente algunos valores de parámetros geométricos entre las cadenas laterales de los residuos del sitio activo, determinados en el modelo de la quimopapaína y en las estructuras cristalográficas de las tres proteasas caracterizadas.

En la parte superior de la **figura 3-14**, se puede observar la representación del subsitio **S2** de las **PSV** (según la numeración de papaína), los residuos del sitio catalítico y también un fragmento del sustrato en el que se indica el enlace a hidrolizar (\curvearrowright). En la parte inferior de la misma figura se aprecia el alineamiento múltiple de las secuencias de seis de estas proteasas. Se han señalado con asteriscos las posiciones completamente conservadas en la familia y con puntos aquellos aminoácidos que presentan características similares en sus cadenas laterales. También se han marcado los residuos correspondientes al subsitio **S2** (•) y los de la triada catalítica (#). Se aprecia que, a excepción de la lisina que aparece en bromelaína de tallo en la posición 175, todos los miembros de la familia mantienen conservados los residuos del sitio activo (Cys, His y Asn). Es de tal manera sorprendente el caso del aminoácido 175 en la bromelaína, que incluso se ha propuesto la existencia de un error (Topham *et al.*, 1990b) en la secuencia de aminoácidos de esta enzima que ha sido reportada por Ritonja *et al.* (1989a).

a)



b)



Figura 3-11. Residuos hidrofóbicos (a) e hidrofílicos (b) en el modelo de la quimopapaína. La línea continua representa la cadena α de la proteína.

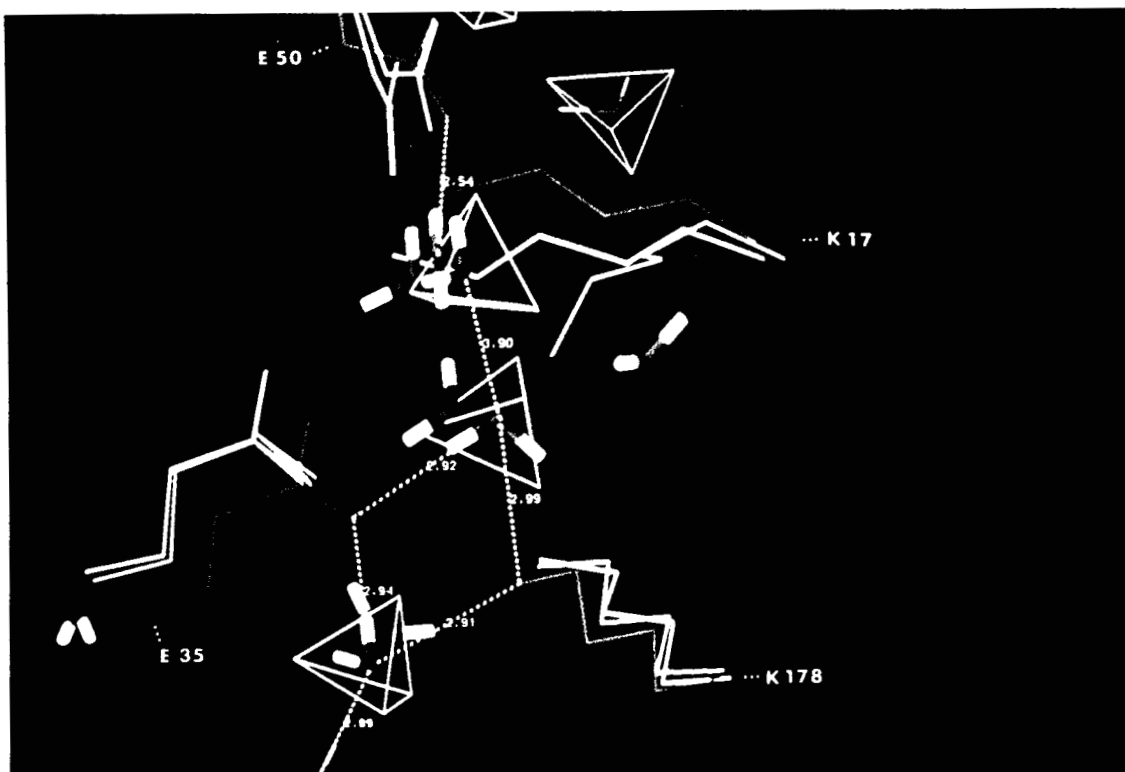


Figura 3-11 (continuación). Residuos hidrofóbicos e hidrofílicos del modelo de la quimopapaina.



Figura 3-12. Exposición de la cisteína 117 en el modelo de la quimopapaína. Se presentan también los tres enlaces disulfuro y la cisteína 25 del sitio activo.

a)



b)

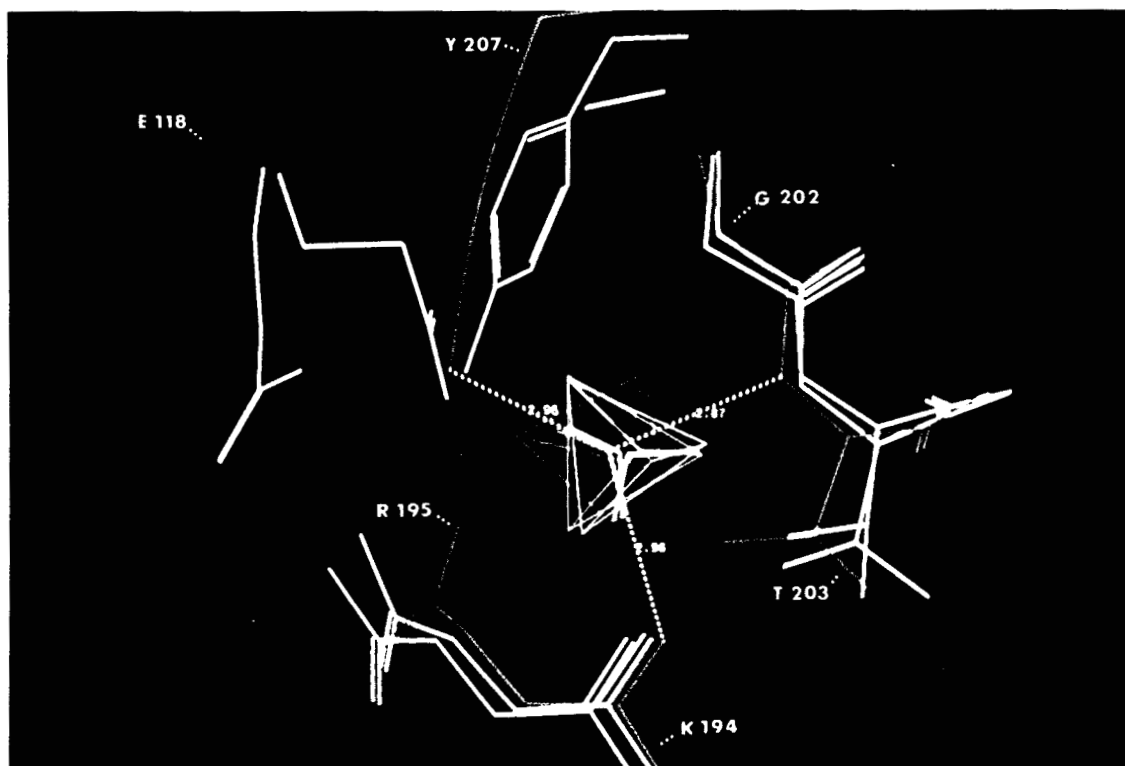
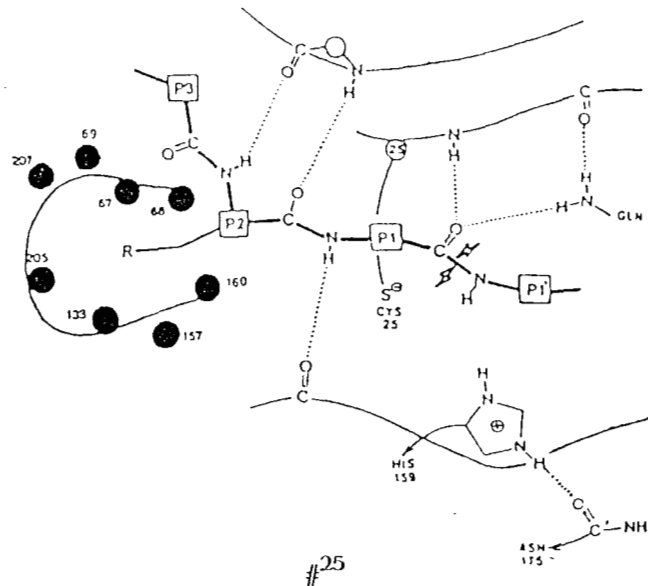


Figura 3-13. Dos sitios conservados en las estructuras de papaina (verde), actinidina (amarillo) y el modelo de la quimopapaina (rojo). a) Zona interdominio. b) Porción del segundo dominio.



ACTINIDINA	-LPSYVDWRSAGAVVDIKSQGECGGCWFSAIATVEGINKITSGSLISLSEQELI	54
PAPAÍNA	-IPEYVDWRQKGAVTPVKNQGSCGSCWAFSAVVTIEGIIKIRTGNLNEYSEQELL	54
PROT. OMEGA	-LPENVDRKKGAVTPVRHQGSCGSCWAFSAVATVEGINKIRTGKLVELSEQELV	54
PAP. PROT. IV	-LPESVDWRAKGAVTPVKHQGYCESWAFSTVATVEGINKIKTGNLVELSEQELV	54
QUIMOPAPAÍNA	-YPQSIDWRAKGAVTPVKNQGACGSCWAFSTIATVEGINKIVTGNLLELSEQELV	54
BROMELAÍNA	AVPQSIDWRDYGAVTSVKQNQPCGACWAFAAIATVESIYKIKKGIPLSEQQVL	55
	* . *** . . . * . * . * . * * . * . * . * . * . * . * . * . . .	

67 ● ● ● 69

ACTINIDINA	DCGRTQNTRGCDGGYITDGFQFIINDGGINTQENYPYTAQDGDVALQDQKYVT	109
PAPAÍNA	DCDR--RSYGCNGGYPSALQLVAQYG-IHYRNTYPYEGVQRYCRSREKGPYAAK	106
PROT. OMEGA	DCER--RSHGCKGGYPYALEYVAKNG-IHLRSKYPPYKAKQGTCAKQVGGPIVK	106
PAP. PROT. IV	DCDL--QSYGCNRYQSTSLQYVAQNG-IHLRAKYPYIAKQOTCRANQVGGPKVK	106
QUIMOPAPAÍNA	DCDK--HSYGCKGGYQTTSLQYVANNG-VHTSKVYPYQAKQYKCRATDKPGPKVK	106
BROMELAÍNA	DCAK--GYGCKGGWEFRAFEFIISNKGVASGAIYPYKAAKGTCKTDGVPNSAY-	106
	** . . . * . * *** . . . *	

● 133

157 ● # ● 160

ACTINIDINA	IDTYDNVPYNNEWALQTAVTYQPVSVVALDAAGDAFKQYASGIFTGPCGTAVDHAI	164
PAPAÍNA	TDGVVRVQPYNEGALLYSIANQPVSVVLEAAGKDFQLYRGGIFVGPCGNKVDHAV	161
PROT. OMEGA	TSGVGRVQPNNEGNLLNAIAKQPVSVVLESKGRPFQLYKGGIFEGPCGTKVDHAV	161
PAP. PROT. IV	TNGVGRVQSNNEGSLNIAHQPVSVVLESAGRDFQNYKGGIFEGSCGTKVDHAV	161
QUIMOPAPAÍNA	ITGYKRVPSNCETSFLGALANQPLSVLVEAGGKPFQLYKSGVFDGPGCTKLDHAV	161
BROMELAÍNA	ITGYARVPRNNESSMMYAVSKQPIVAVDANAN-FQYKSGVFNGPCGTSLNHAV	160
	... * . * . . . * . * . * . * . * . * . * . . . * . * . * . * . * . . .	

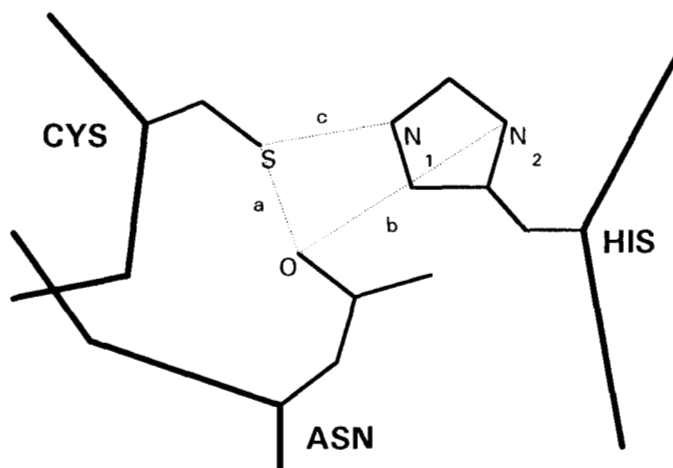
175

205 ● ● 207

ACTINIDINA	VIVGYGTEGGVDYWIVKNSWDTTWGEEGYMRILRNVGGA-GTCGIATMPSYPVKYNN	220
PAPAÍNA	AAVGYGPN----YILIKNSWGTGWGENGYIRIKRGTGNSYGVCGLYTSSFYPVKN--	212
PROT. OMEGA	TAVGYGKSGGKGYILIKNSWGTAWGEGYIRIKRAPGNSPGVCGLYKSSYYPTKN--	216
PAP. PROT. IV	TAVGYGKSGGKGYILIKNSWGPWGGENGYIRIRRASGNSPGCVYRSSYYPIKN--	216
QUIMOPAPAÍNA	TAVGYGTSQKNYIILIKNSWGPNWGEGYMRILKRQSGNSQGTGVYKSSYYPFKGFA	218
BROMELAÍNA	TAIGYGQDS----IIYPKKWGWGEGYIRMARDVSSSSGICGIAIDPLYPTLEE-	212
	. . *** * . * . * . * . * . * . * . . . * . * . * . * . * . .	

Figura 3-14. Residuos del sitio activo (#) y del subsitio S2 (●) en PSV, según la numeración de papaína, sobre el alineamiento múltiple de las proteasas completamente secuenciadas de la familia. También se indica el enlace que sufre hidrólisis en el sustrato (✓).

TABLA 3-IV
Geometría del sitio activo en las PSV.



	PAPAÍNA	ACTINIDINA	PROTEINASA Ω	QUIMOPAPAÍNA
a (Å)	3.65	3.30	3.91	3.91
b (Å)	2.77	2.79	2.72	4.64
c (Å)	7.25	6.99	7.36	6.96
$\angle SN_1O$ (°)	116.01	71.22	117.44	87.05
$\angle SN_2O$ (°)	134.19	60.03	124.15	97.68

Comparación estructural del modelo con otros miembros de las PSV

De la **figura 3-14**, y siguiendo con la numeración de papaína, puede apreciarse también que la tirosina 67 se conserva en cinco de las proteasas, con la bromelaína nuevamente como la única excepción. La posición 68 muestra gran variabilidad en la familia siendo un residuo no polar el que ocupa esta posición en los casos de actinidina, papaína y proteinasa Ω ; polar en proteinasa IV y quimopapaína e incluso; cargado eléctricamente en la bromelaína. Por su parte, la posición 69 muestra también la sustitución de cadenas laterales con características muy diferentes; en tamaño, desde la pequeña serina hasta el triptofano, y en polaridad, desde la fenilalanina hasta la treonina. Las posiciones 133, 157 y 160 conservan la

naturaleza hidrofóbica de sus residuos, manteniendo completamente invariante a la alanina en esta última posición. Finalmente, los residuos 205 y 207 en las proteasas analizadas presentan un gran carácter polar e hidrofóbico, respectivamente, excepto en el caso de la actinidina. A juzgar por la naturaleza de los residuos del subsitio **S2**, esta cavidad parece tener el mayor volumen en esta última enzima, mientras que en la bromelaína se distingue por la presencia de dos cargas negativas.

Mediante el uso del modelo de la quimopapaína y de las estructuras tridimensionales de actinidina, papaína y proteinasa Ω , se determinaron las áreas de las cadenas laterales de los residuos que conforman a los subsitios **S2** en cada una de estas proteínas y que se anotan en la **tabla 3-IV**. La superficie del subsitio **S2** representada con puntos, se muestra en la **figura 3-15** para el modelo de la quimopapaína.

TABLA 3-V

	Area de la superficie del subsitio S2 (Å^2)
Quimopapaína	332.0
Actinidina	302.2
Papaína	341.2
Proteinasa Ω	266.4

Puede apreciarse que los valores obtenidos en la tabla anterior, son muy similares entre sí y que el mayor volumen, esperado para la actinidina, no es evidente de este análisis de áreas superficiales. De aquí se hace patente la necesidad de estimar los volúmenes libres del subsitio **S2**; un trabajo que no resulta trivial, pues requiere entre otras cosas, de la definición precisa y sistemática de los límites de estas regiones.

El análisis del modelo continuó con la comparación de las características topológicas de los miembros de la familia de las proteasas sulfhidríticas vegetales caracterizadas. La quimopapaína muestra la presencia, típica en la familia, de dos dominios estructurales, y su cadena principal es muy semejante a la encontrada por Baker y Drenth (1987) para papaína, arquetipo de las **PSV** (**Fig. 3-16**). Las estructuras de estas dos enzimas pueden superponerse fácilmente, aunque se observan pequeñas porciones de la cadena principal que divergen. Para ubicar esas zonas, se

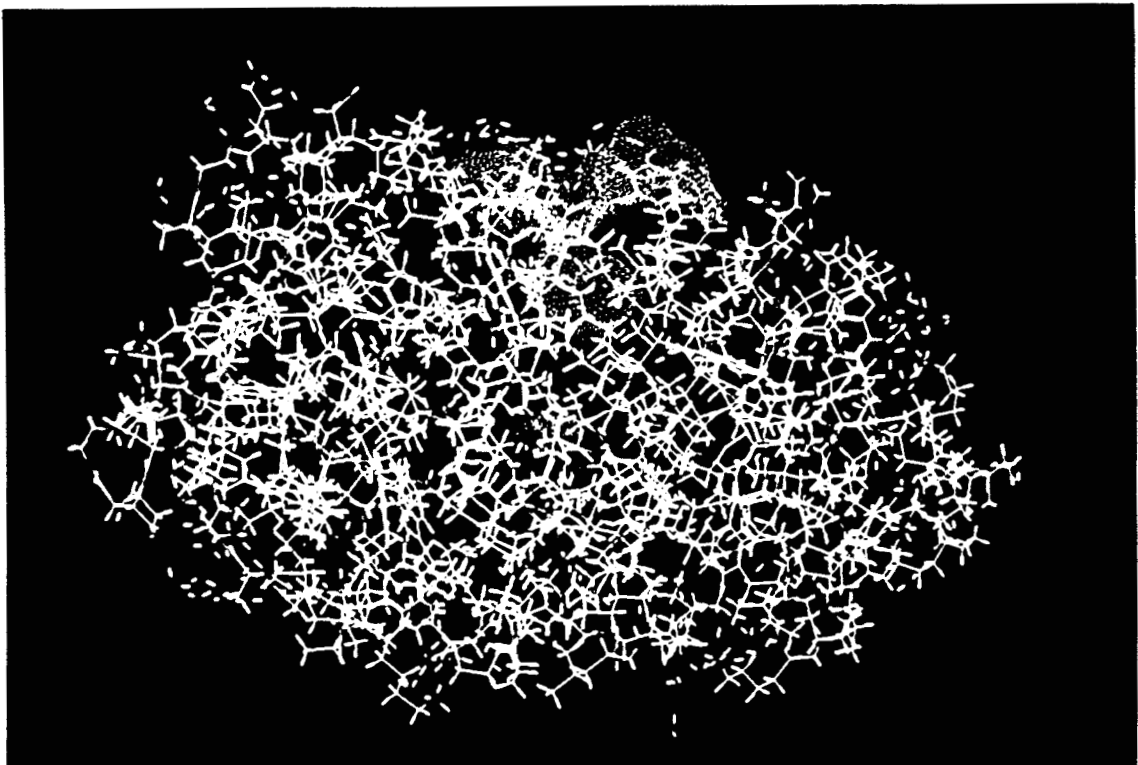
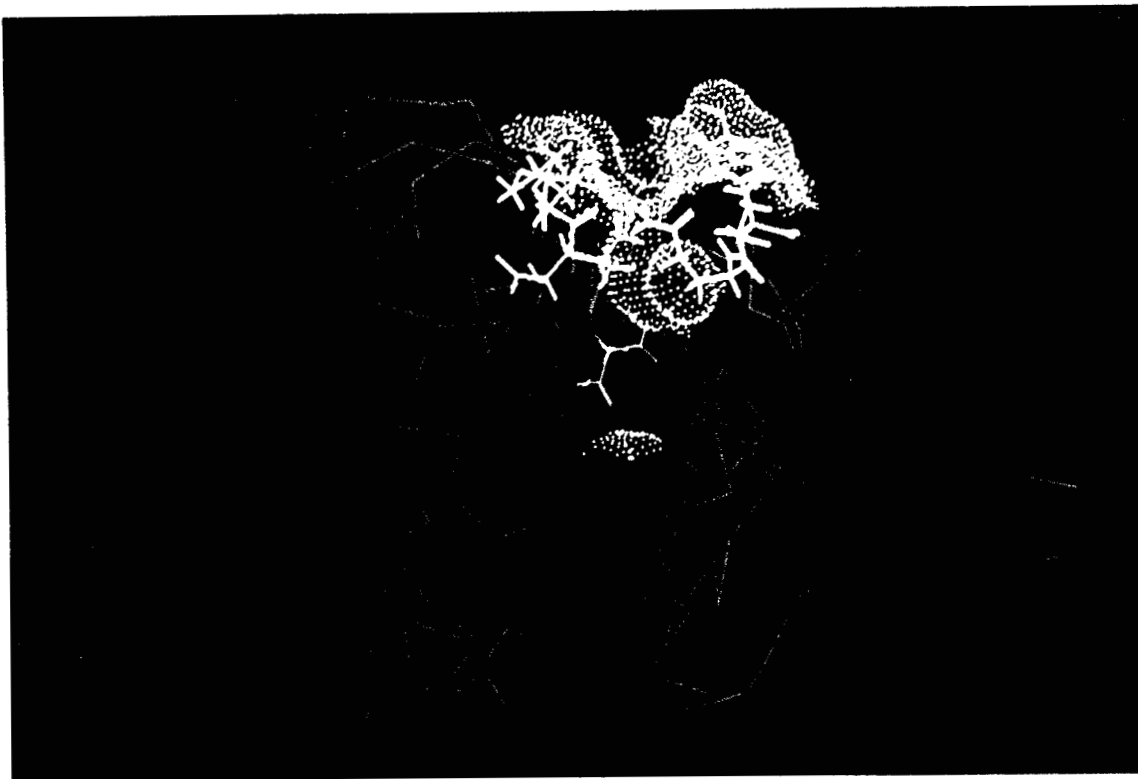


Figura 3-15. Representación con puntos de la superficie expuesta al disolvente del subsitio S2 en el modelo de la quimopapaina.

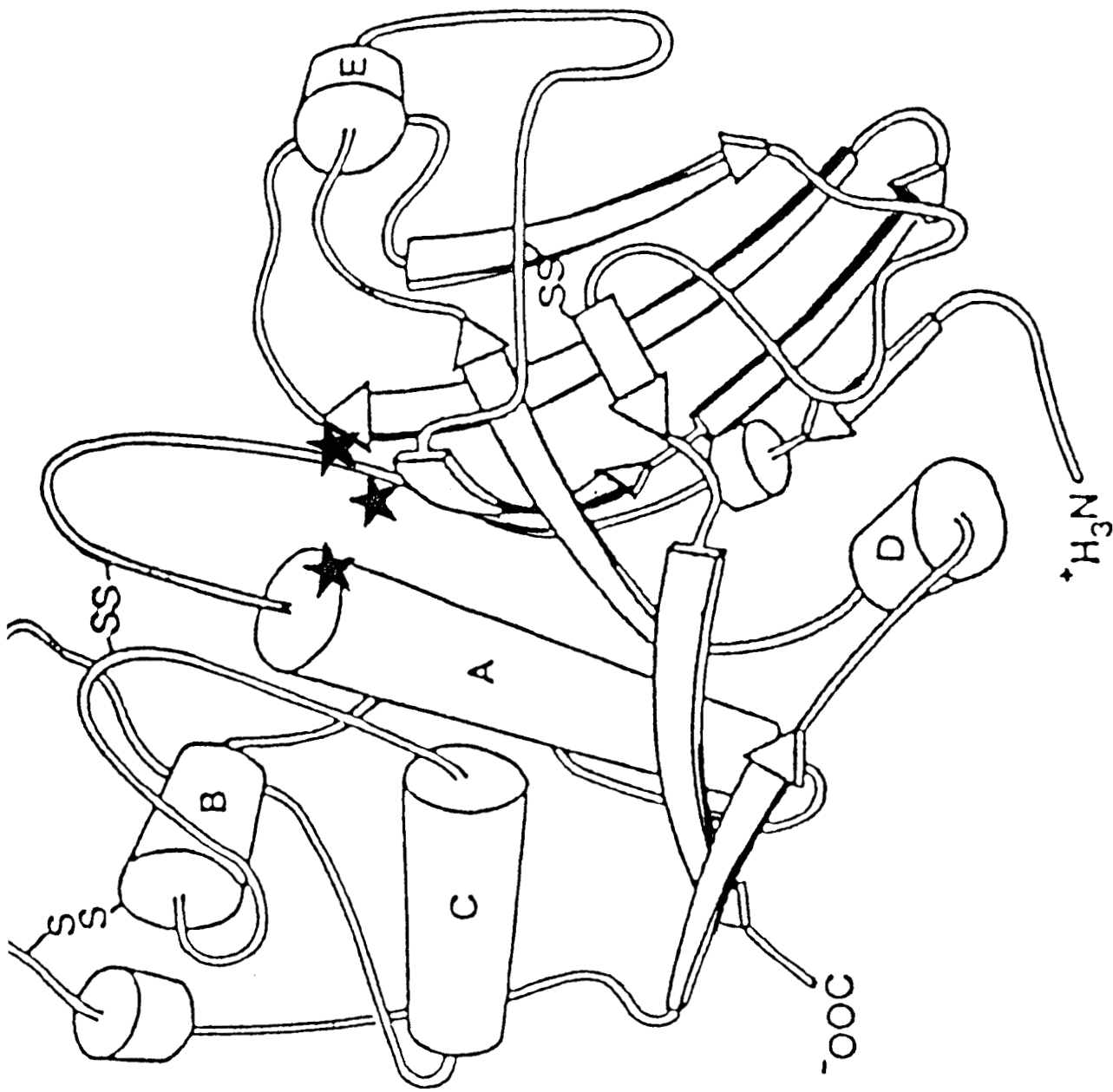


Figura 3-16. Esquema de la estructura tridimensional de la quimopapaína de acuerdo con Baker & Drenth (1987). Las hélices α se representan con cilindros, y las hojas β con flechas. Las estrellas marcan la posición de los residuos del sitio activo.

calculó la diferencia en posición de los carbonos α de la cadena principal de la quimopapaína respecto a la papaína, y también respecto a la proteinasa Ω . Esto último se realizó con el objeto de identificar cualquier posible influencia de la estructura de base sobre la conformación final del modelo. Los resultados señalan (**Fig. 3-17a**) que la estructura obtenida para la quimopapaína es incluso, en algunas porciones de la molécula, más cercana a la papaína que a la propia proteinasa Ω ; esto es, no se observan tendencias geométricas en el modelo que indiquen la permanencia de la conformación de la estructura de base en la quimopapaína. Por su parte, es notable la gran similitud en ubicación y magnitudes relativas de las zonas de mayor variabilidad presentes en la **figura 3-17a**, con las desviaciones encontradas por Baker & Drenth (1987) en la posición de la cadena principal de papaína y actinidina (**Fig. 3-17b**). Este último resultado indica que el modelado logró llevar a la molécula, desde el espacio conformacional correspondiente a la estructura de la proteinasa Ω , hasta una posición que refleja razonablemente el compromiso de las interacciones que experimentan los grupos químicos en la quimopapaína. Un criterio adicional de comparación topológica consistió en el cálculo de las diferencias en los ángulos Φ y Ψ , entre el modelo, la papaína y la proteinasa Ω (**Figs. 3-18 y 3-19**). En este caso se observó la concordancia existente entre las regiones de la cadena principal que ven alterados sus ángulos diedros y los que divergen más en la posición de sus carbonos α . De esta manera, se tienen ya dos evidencias independientes que identifican, en el modelo de la quimopapaína, características geométricas particulares que permiten su consideración como una estructura tridimensional más de la familia de las proteasas sulfhidríticas vegetales. La **figura 3-20** muestra la superposición de las cadenas principales de todas las estructuras conocidas de miembros de las **PSV**.

Aplicaciones del modelo propuesto

En un estudio experimental reciente donde se aplica la espectroscopía de dicroísmo circular sobre papaína, quimopapaína y proteinasa Ω (Solís-Mendiola *et al.*, 1992), se advierte la existencia aparente de un diferente contenido de estructuras secundarias entre estas tres proteasas, y se encuentra un contenido mayor de residuos en conformación de hélice α en quimopapaína. Además, estos mismos autores indican que las variaciones en los espectros de dicroísmo circular sugieren la presencia de un patrón de doblado distinto para la quimopapaína respecto al que se ha encontrado para las otras dos enzimas estudiadas e incluso, la posibilidad de que

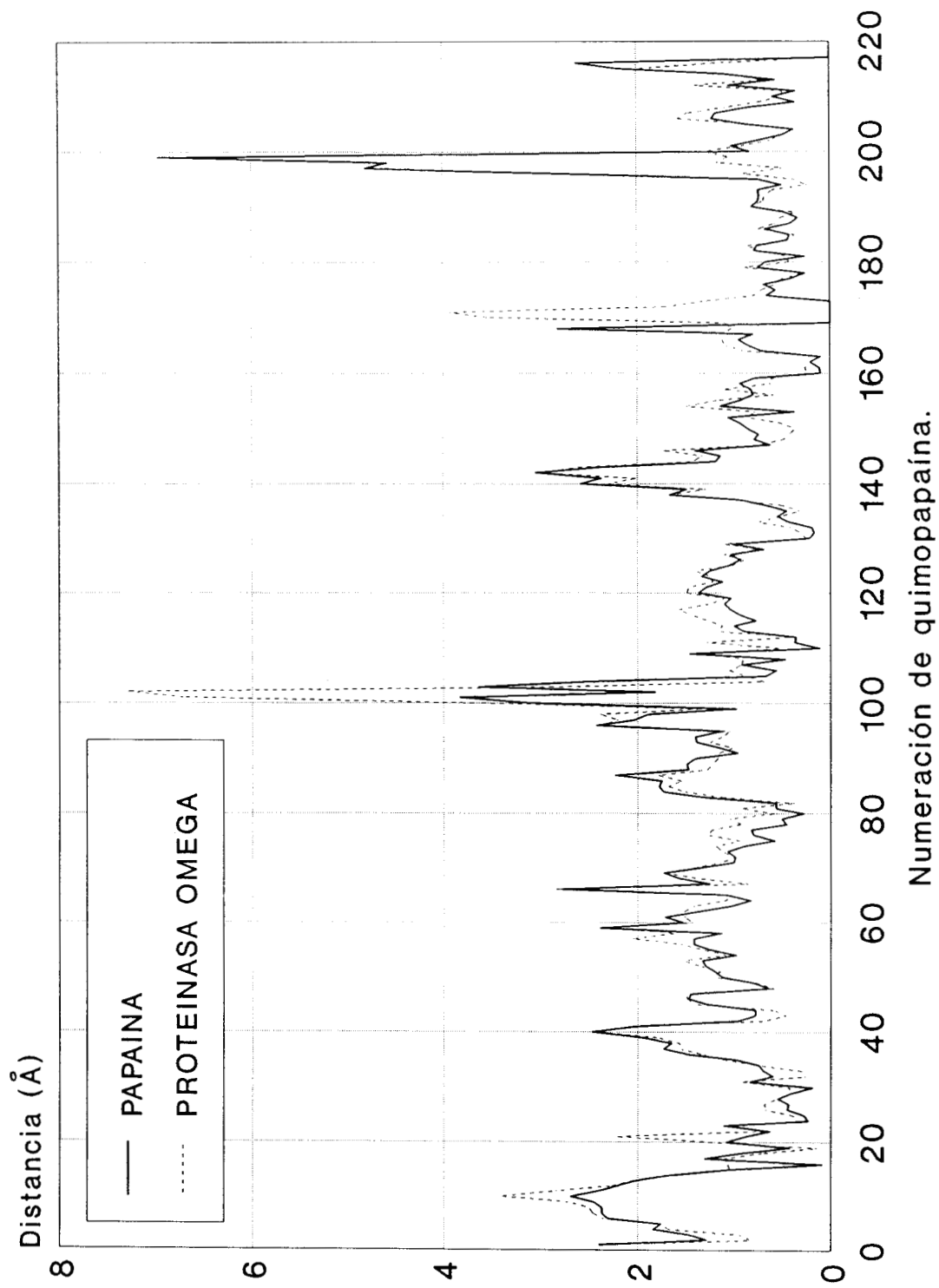


Figura 3-17a. Diferencia en la posición de los carbonos α en papaina (—) y proteinasa Ω (···) respecto a quimopapaina.

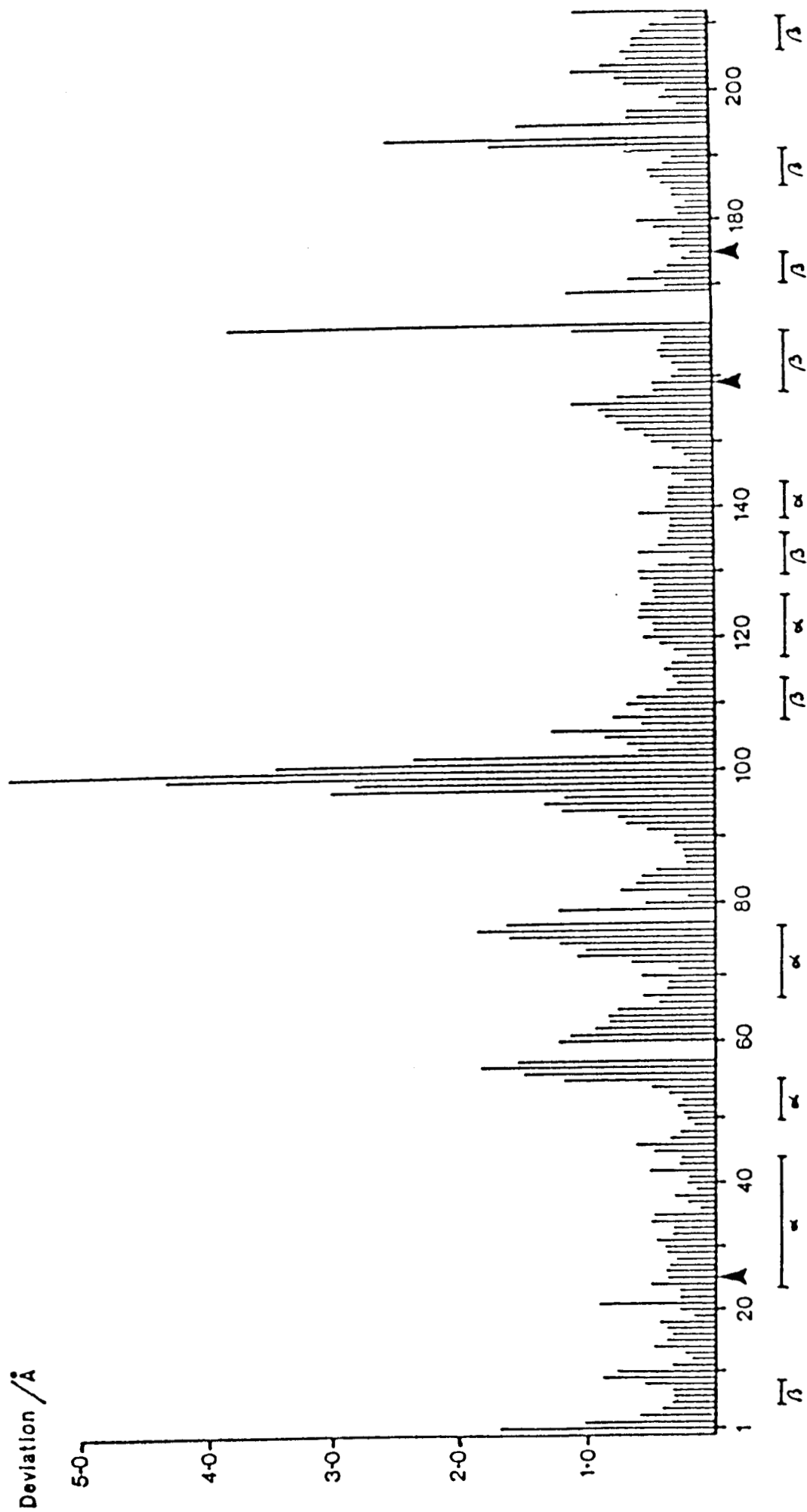


Figura 3-17b. Promedio de las desviaciones por residuo para la posición de los átomos de la cadena principal entre papaina y actinidina, de acuerdo con Baker & Drenth (1987).

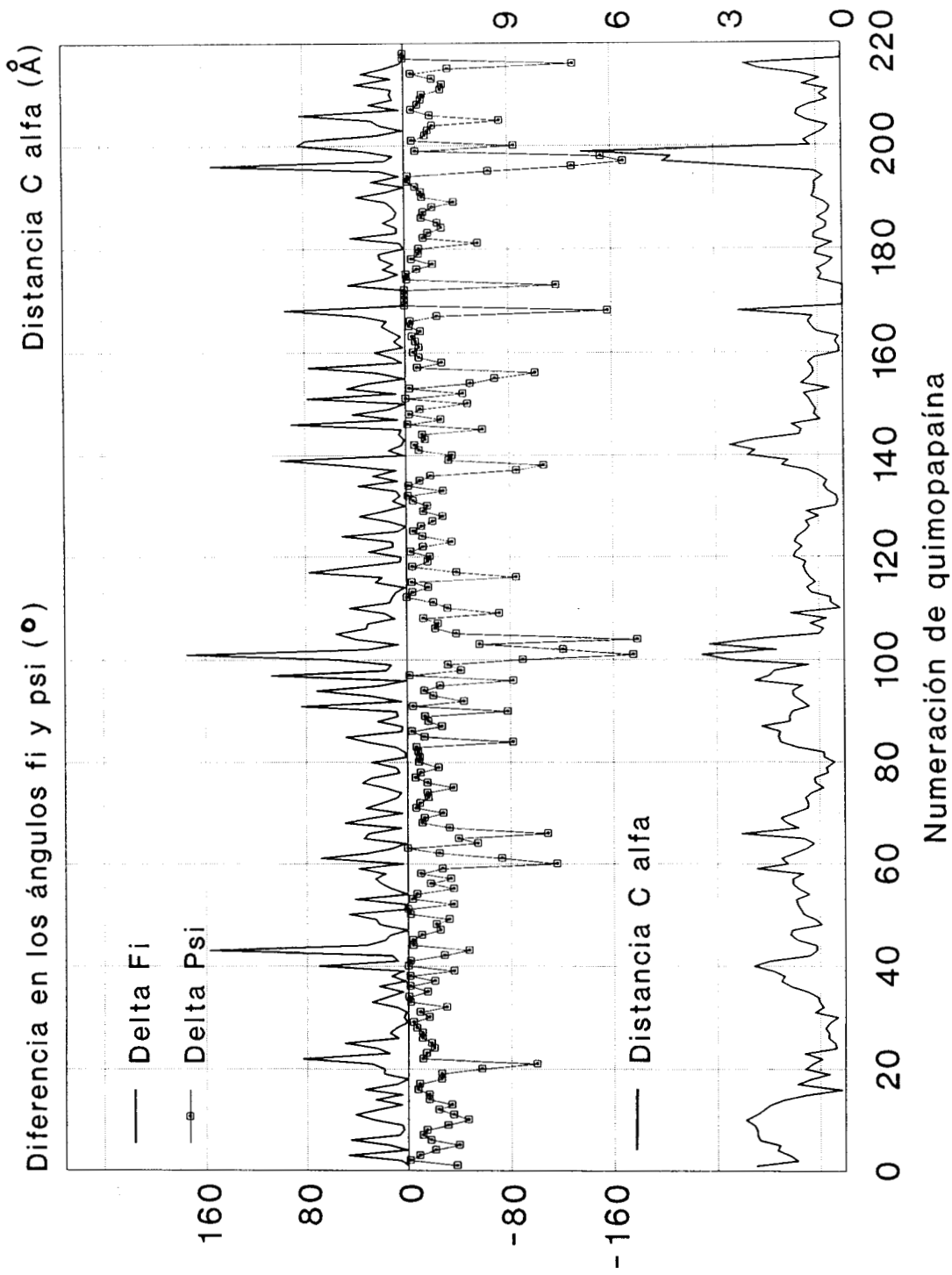


Figura 3-18. Desviación de los valores de los ángulos Φ y Ψ , y diferencia en la posición de los carbonos α entre papaina y el modelo de la quimopapaína.

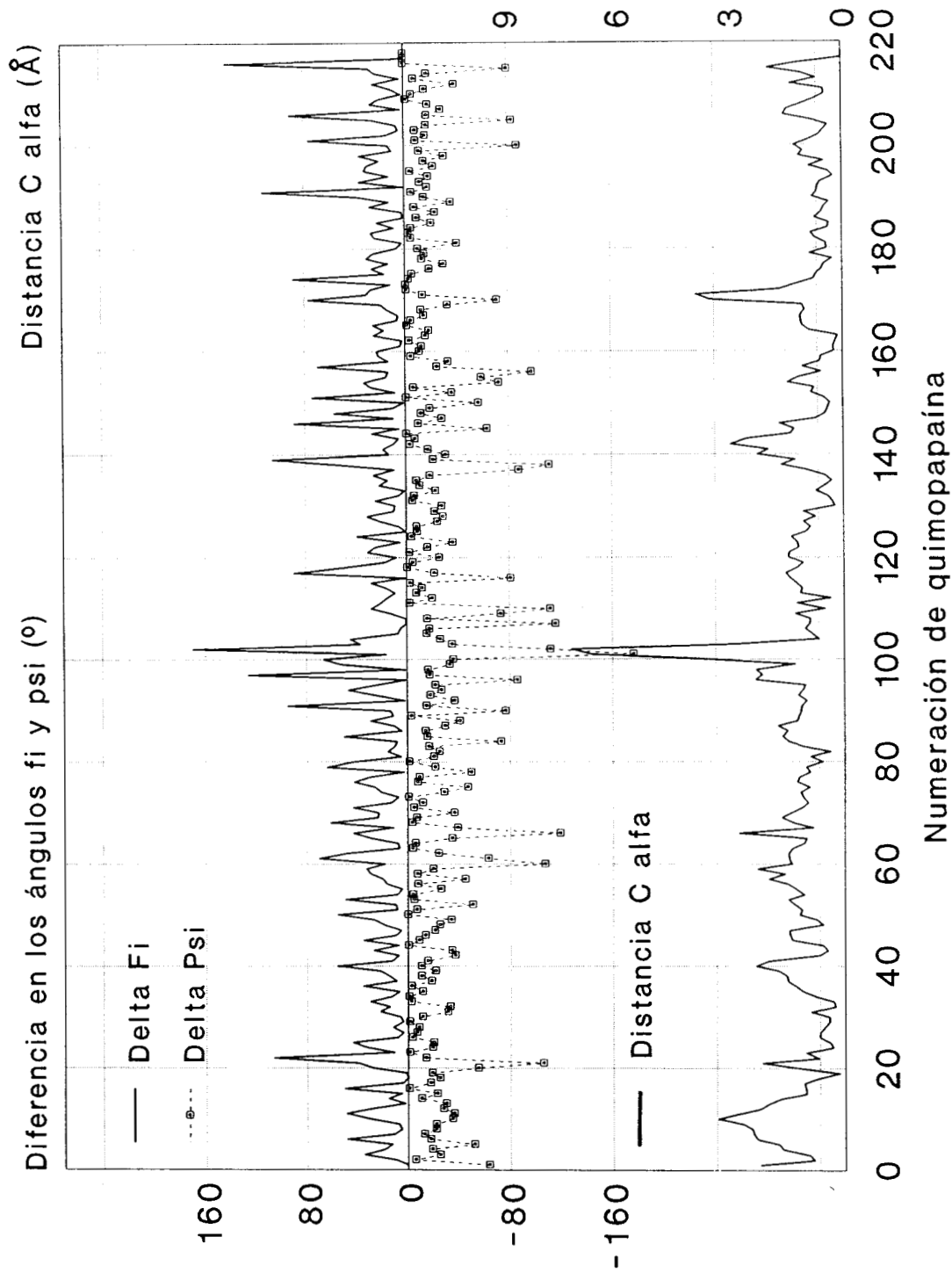
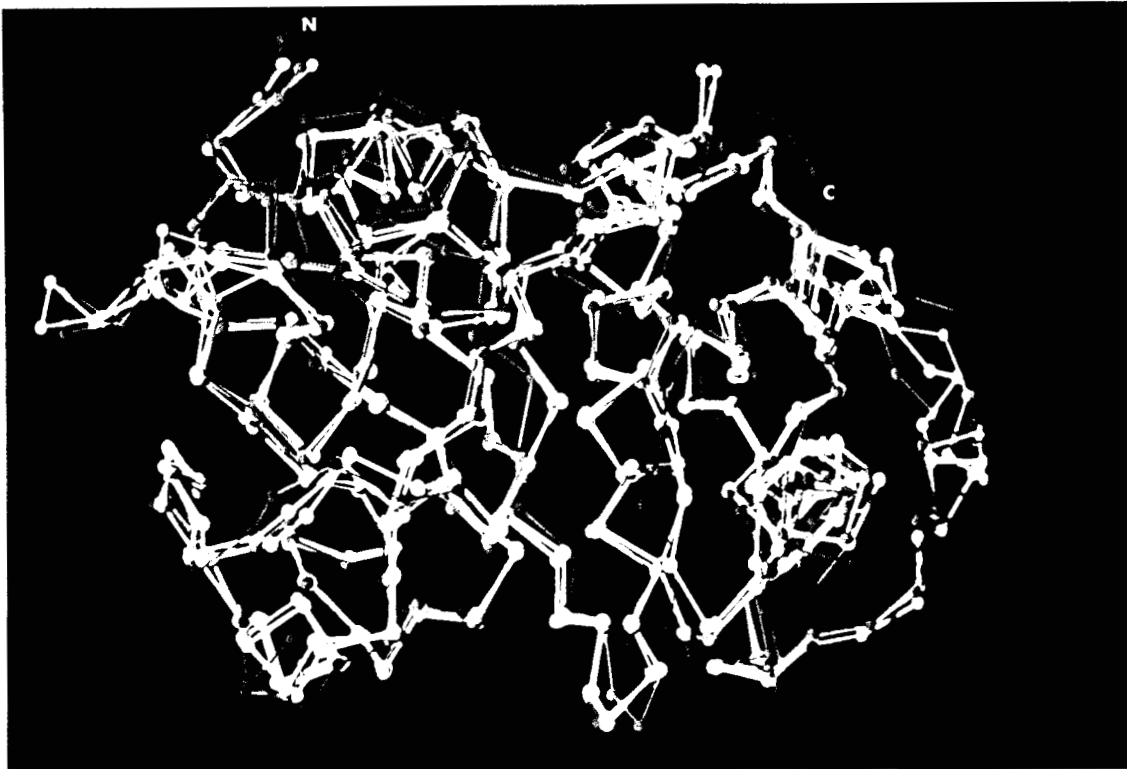


Figura 3-19. Desviación de los valores de los ángulos Φ y Ψ y diferencia en la posición de los carbonos α entre proteinasa Ω y el modelo de la quimopapaina.

a)



b)

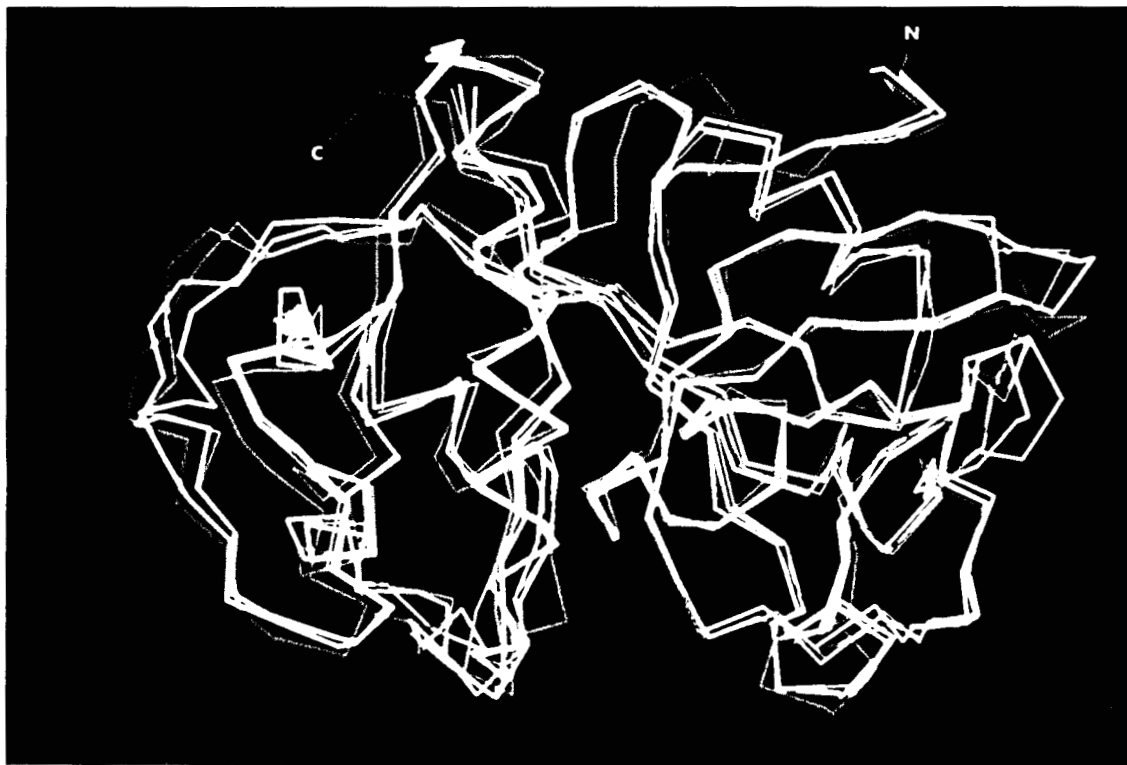
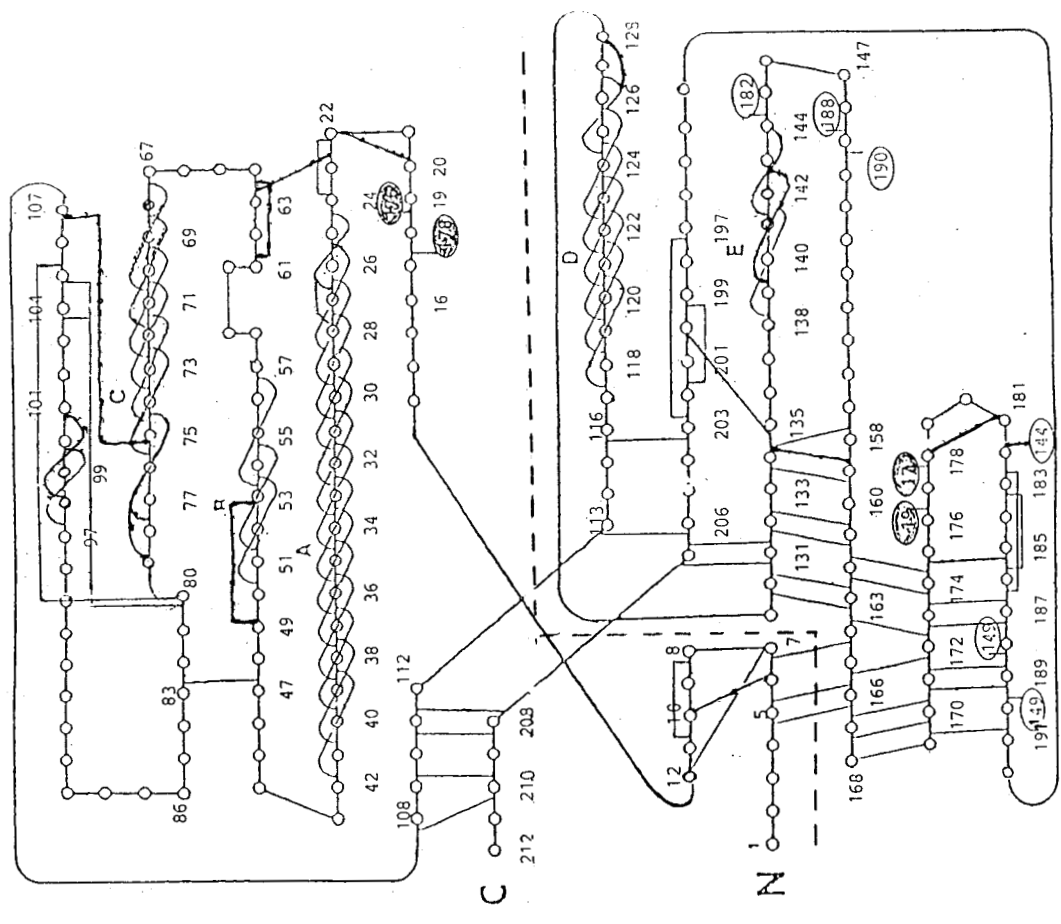


Figura 3-20. Superposición de las cadenas α de papaína (verde), actinidina (amarillo), proteinasa Ω (azul) y quimopapaína (rojo) en representación de esferas y bastones (a) y barras (b).

esta enzima pertenezca a una clase estructural diferente (α/β , segmentos alternados de hélices α y hojas- β). El hecho de que la quimopapaína pudiera tener un patrón de doblado distinto o pertenecer a una clase estructural diferente a las otras dos **PSV** estudiadas, resulta inesperado dada la alta homología en secuencia existente entre estas proteínas (más del 50%) como se mencionó en ese trabajo. En un intento por explicar esta situación, se construyó la red de puentes de hidrógeno entre los átomos de las cadenas principales de las tres **PSV** consideradas, cuyos esquemas de conectividad se muestran en la **figura 3-21**. Estas representaciones se basan en un estudio similar practicado sobre actinidina (Baker & Drenth, 1987) el cual se incluye también en la **figura 3-21** citada, para su comparación. Los esquemas de puentes de hidrógeno definen claramente la extensión de cada una de las estructuras secundarias presentes en las moléculas, y pueden dar una indicación de la clase estructural a la que pertenece cada proteína. La construcción de esta red de interacciones fue posible en quimopapaína sólo debido a que se contaba con un modelo de su estructura tridimensional, en el que pueden definirse los puentes de hidrógeno de la misma manera en que se efectúa sobre una estructura determinada cristalográficamente. Del análisis de esta información se hace evidente que la quimopapaína mantiene el patrón general de doblado presente en las **PSV**, y además conserva la clasificación estructural de la familia ($\alpha+\beta$, regiones predominantemente helicoidales y zonas ricas en hojas- β en porciones separadas de la molécula). Es interesante notar aquí, que la papaína posee un par de cadenas en conformación de hoja- β de menor longitud respecto a las otras **PSV** (que corresponde a la inserción de cuatro residuos entre las posiciones 169 y 170 de papaína) y que todas las hojas β presentes en la familia son antiparalelas.

Respecto al punto del diferente contenido en estructura secundaria y particularmente a la mayor incidencia de residuos en conformación helicoidal en quimopapaína, en la **figura 3-22** se presenta el alineamiento de las secuencias de esta enzima junto con las de la papaína y la proteinasa Ω y se señalan las diferentes estructuras secundarias presentes en cada proteína. Estas estructuras secundarias fueron determinadas tanto por el criterio de puentes de hidrógeno como por el de ángulos Φ y Ψ . Es evidente la gran semejanza en localización y extensión de hélices α y de hojas- β en estas tres proteasas, siendo la única diferencia detectable la longitud de las dos primeras hélices (A y B). De aquí puede proponerse que el contenido de estructura secundaria en estas enzimas, no es capaz de explicar

PAPAINA



QUIMOPAPAINA

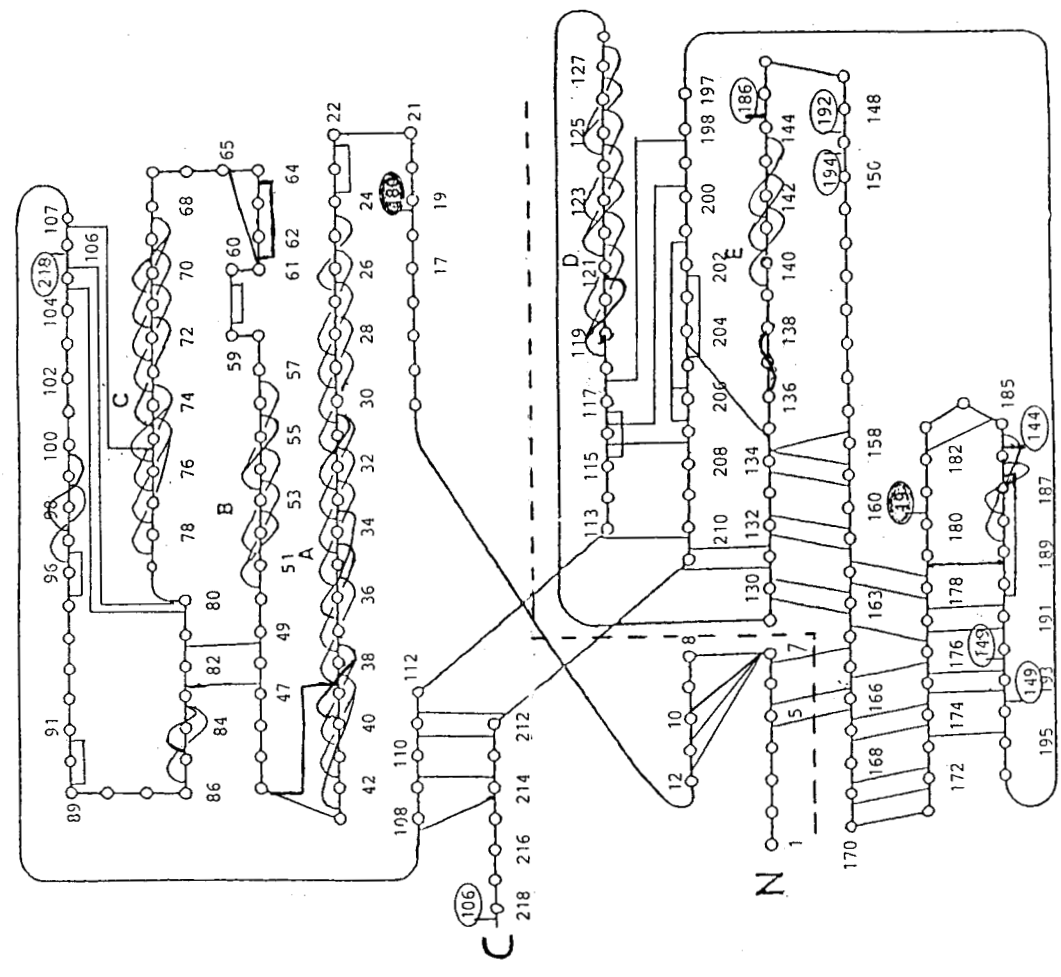
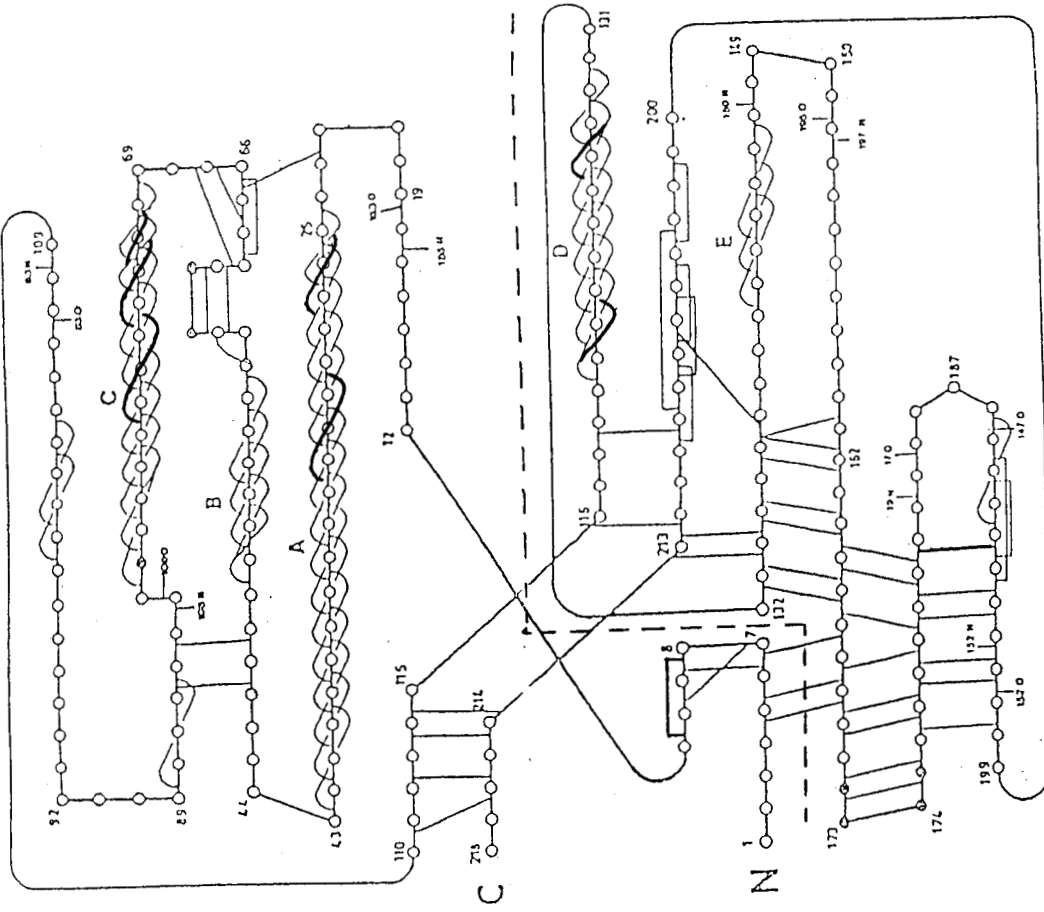


Figura 3-21. Esquemas de conectividad de puentes de hidrógeno en las cadenas principales de quimopapaina y papaina. La línea punteada separa a los dominios estructurales de acuerdo con la convención empleada en este trabajo.

ACTINIDINA



PROTEINASA OMEGA

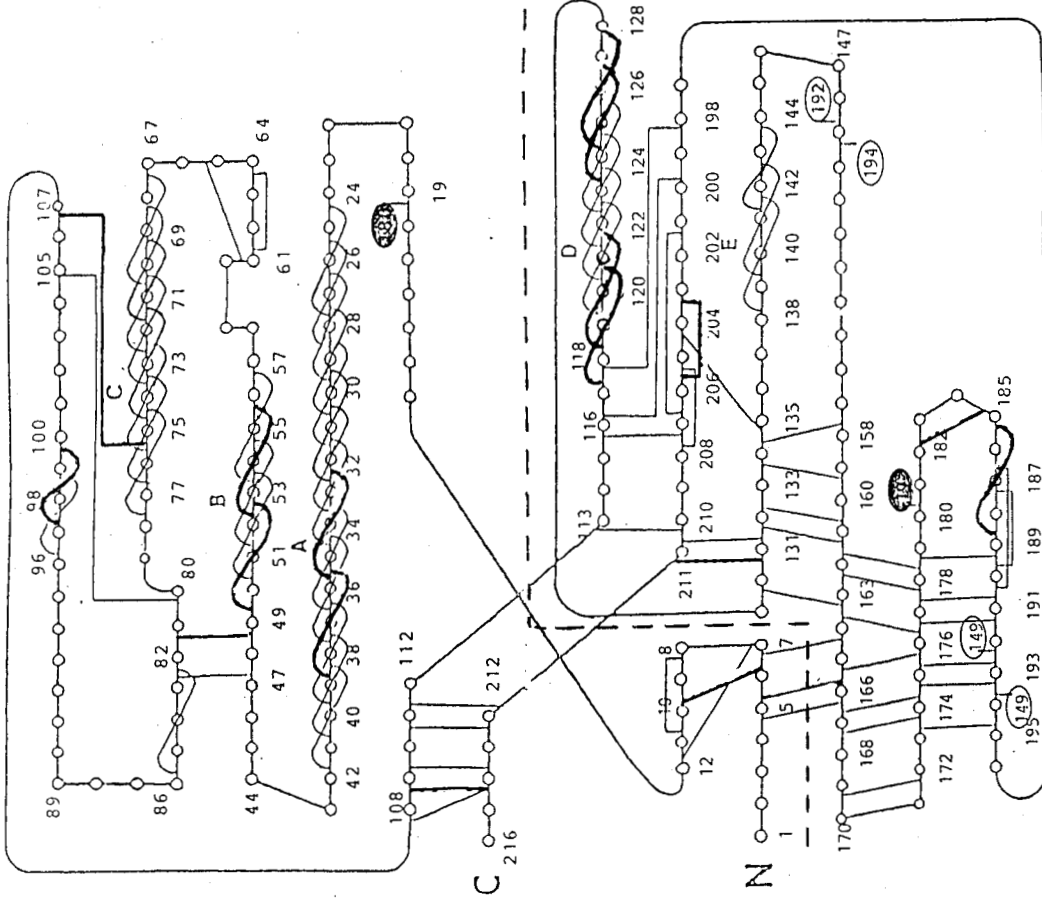


Figura 3-21 (continuación). Esquemas de conectividad de puentes de hidrógeno en las cadenas principales de proteinasa Ω y actinidina. Esta última según el esquema de Baker & Drenth (1987). La línea punteada separa a los dominios estructurales de acuerdo con la convención empleada.

	1		A		B							
PAPAÍNA	IPEY	<u>VDW</u>	RQKGAVTPVKNQGSCGS	<u>CWAFSAVV</u>	<u>TIEGLIK</u>	IRTGNLNEYS	<u>EQELLD</u>	55				
PROT.OMEGA	LPEN	<u>VDW</u>	RKKGAVTPVRHQGSCGS	<u>CWAFSAVAT</u>	<u>VEGINKI</u>	RTGKLVLS	<u>EQELVD</u>	55				
QUIMOPAPAÍNA	YPOS	<u>IDW</u>	RAKGAVTPVKNQGACGS	<u>SCWAFSTIAT</u>	<u>VEGINKIV</u>	TGNLLEL	<u>SEQELVD</u>	55				
		C				2						
PAPAÍNA	CDRRSYGCNGGY	<u>PWSALQLVAQ</u>	YGIHYRNTYPYEGVQRYCRSREKGPYAAKT	<u>DGV</u>				110				
PROT.OMEGA	CERRSHGCKGGY	<u>PPYALEYVAK</u>	NGIHLRSKYPYKAKQGT	CRAKQVGGPIVKT	<u>SGV</u>			110				
QUIMOPAPAÍNA	CDKHSYGCKGGY	<u>QTTSLQYVAN</u>	NGVHTSKVYPYQAKQYKCRATDKPGPKVKI	<u>TGY</u>				110				
	2	D	3	E		4						
PAPAÍNA	<u>ROVQ</u>	PYNE	<u>EGALLYSIANQ</u>	<u>PVSVVLE</u>	EAAG	<u>KDFQL</u>	YRGGIFVGP	CGNKV	<u>DHAVA</u>	<u>AVG</u>	165	
PROT.OMEGA	<u>GRVQ</u>	PNN	<u>EGNLLNAIAKQ</u>	<u>PVSVVV</u>	ESKGR	<u>PFQL</u>	YKGGIFEG	PCG	TKV	<u>DHA</u>	<u>VTAVG</u>	165
QUIMOPAPAÍNA	<u>KRV</u>	PSNC	<u>ETSFLGALANQ</u>	<u>LSVLV</u>	EAGGK	<u>PFQLY</u>	KSGVFDG	PCG	TKL	<u>DHA</u>	<u>VTAVG</u>	165
	4	5		6		7						
PAPAÍNA	<u>YG</u>	PN---	<u>YILIKN</u>	SWG	TG	WGEN	<u>GYIRIKR</u>	GTGNSY	GVCGLYTSS	<u>FYPVKN</u>	212	
PROT.OMEGA	<u>YGKS</u>	GG	<u>KGYILIK</u>	NSW	TAW	GEK	<u>GYIRIKR</u>	APGNSP	GVCGLYKSS	<u>YYPTKN</u>	216	
QUIMOPAPAÍNA	<u>YGTS</u>	DG	<u>KNYIIIK</u>	NSW	GNW	GEK	<u>GYMRLKR</u>	QSGNSQ	GTCGVYKSS	<u>YYPFKGFA</u>	218	

Figura 3-22. Estructuras secundarias marcadas sobre las secuencias de aminoácidos de papaína, proteinasa Ω y quimopapaína. Las hélices α se denotan con letras y las hojas β con números.

completamente las diferencias observadas en los espectros de dicroísmo circular de las tres proteínas estudiadas. Tales diferencias se presentan principalmente en la cercanía de los 200 nm, donde una señal que resulta muy evidente en papaína, desaparece en quimopapaína y además, para ésta última, la magnitud relativa de las señales cercanas a 210 y 220 nm se encuentra invertida (**Fig. 3-23**). Por otra parte, al observar los espectros de dicroísmo circular de la bromelaína de tallo (Arroyo-Reyna, 1993), la actinidina (Valle-Guadarrama, 1993) y la ficina (López y Celis, 1993), se encuentra que todas estas proteasas sulfhidrúlicas parecen poseer espectros con características intermedias a las que se presentan en los correspondientes a papaína y quimopapaína. Estos espectros se incluyen en la **figura 3-24**, con fines comparativos.

Al intentar dar una interpretación a lo que aparentemente se trata de una desaparición gradual de la señal de 200 nm en los espectros de dicroísmo circular obtenidos para las **PSV** en la región de enlaces peptídicos (180-250 nm), que se observa en la **figura 3-24**, y considerando que las estructuras secundarias de esas enzimas se encuentran ampliamente conservadas (al menos en las estructuras tridimensionales conocidas), se pensó en atribuir tal desvanecimiento de la señal de 200 nm a la presencia de algún cromóforo no peptídico que hubiese sido obviado en trabajos previos. Debido a que la cantidad de cadenas laterales con grupos químicos aromáticos, específicamente triptofanos, varía sin seguir una tendencia particular al pasar de una proteína a otra de la familia, se decidió realizar un análisis comparativo de la geometría de los tres puentes disulfuro que se encuentran completamente conservados en las **PSV**. Esta decisión se tomó en base a lo reportado por Perczel *et al.* (1992) acerca de la contribución de estos elementos estructurales a las señales de dicroísmo circular en la región del ultravioleta lejano, y a los reportes de Casey & Martin (1972) y Ottnad *et al.* (1975) quienes encuentran una señal de signo negativo alrededor de los 200 nm para compuestos modelo que contienen enlaces disulfuro. Los resultados de este análisis se presentan en la **figura 3-25**, donde se aprecian los valores de los ángulos diedros C β -S-S-C β de los puentes disulfuro para papaína, proteinasa Ω , actinidina y quimopapaína. Puede notarse que en los cuatro casos, los valores de los ángulos muestran una tendencia suave a cambiar en el mismo orden en que disminuye la señal de 200 nm de sus espectros (proteínasa omega seguida de papaína, actinidina y por último quimopapaína), lo que hace muy razonable la asignación de esa banda en dicroísmo circular a la geometría de las tres interacciones covalentes entre residuos de cisteína en las **PSV**. Los resultados del análisis anterior

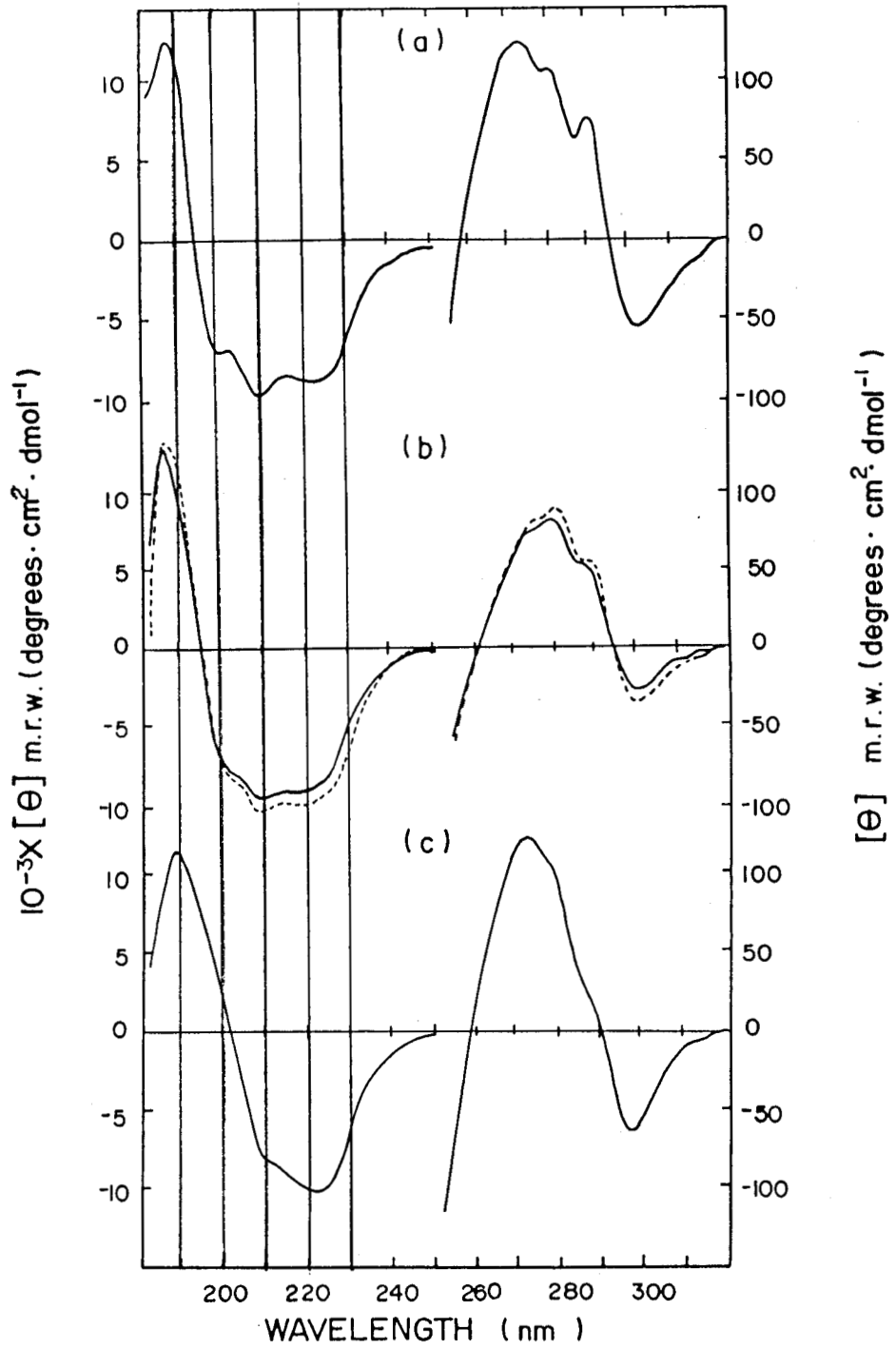


Figura 3-23. Espectros de dicroísmo circular de papaina (a) proteinasa Ω (b) y quimopapaina (c), de acuerdo con Solís-Mendiola *et al.* (1990).

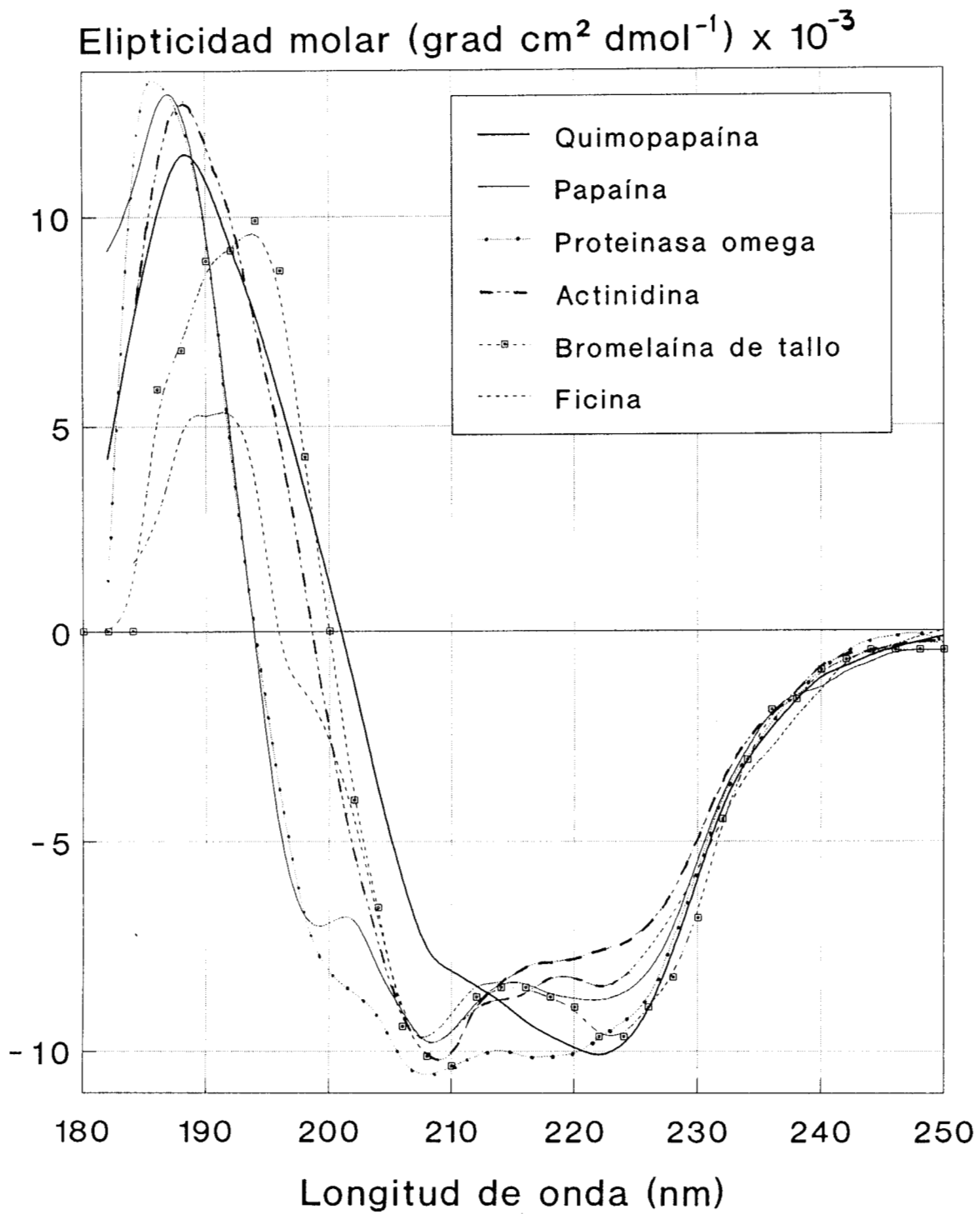


Figura 3-24. Espectros de dicroísmo circular para seis PSV en la zona del ultravioleta lejano.

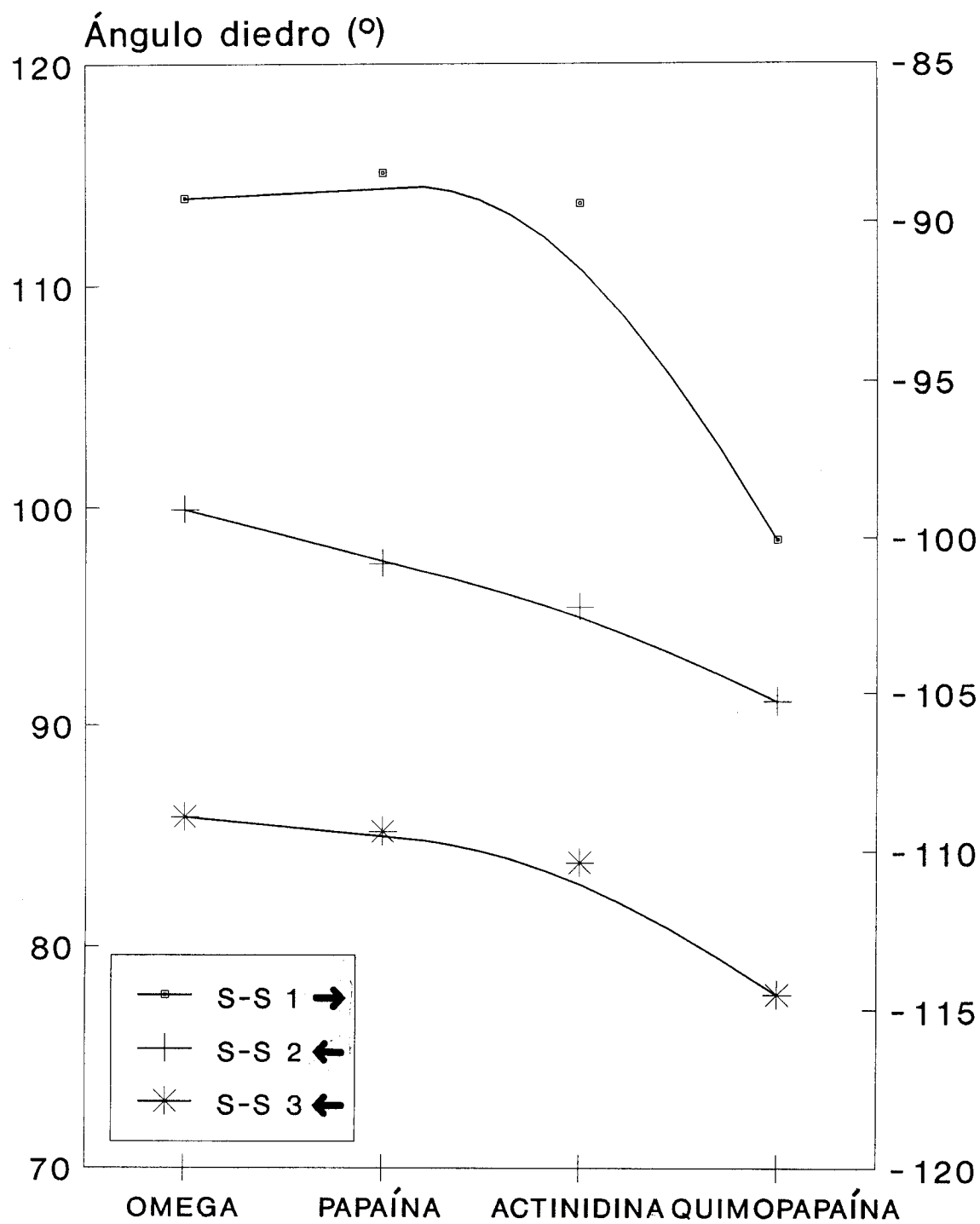


Figura 3-25. Ángulos diedros C β -S-S-C β de los tres enlaces disulfuro para proteinasa Ω , papaína, actinidina y quimopapaína.

sugieren a esta familia como un sistema que permitiría estudiar por primera vez, la influencia de la conformación de los puentes disulfuro en la espectroscopía de dicroísmo circular en proteínas, en la región del ultravioleta lejano. Es interesante señalar que la orientación relativa de algunas estructuras secundarias cambia de una **PSV** a otra, siguiendo la misma tendencia de las señales de dicroísmo circular a 200 nm. Esto significa que las posiciones relativas de estructuras secundarias podrían contribuir significativamente a la forma de los espectros (**Fig. 3-26**).

También se han realizado estudios de estabilidad sobre papaína y quimopapaína (Solís-Mendiola *et al.*, 1993) en los que se han identificado variaciones de los parámetros termodinámicos para el proceso de desnaturalización térmica en ambas enzimas. Esto resulta muy llamativo, pues constituye el segundo grupo de resultados experimentales que apuntan en la dirección de la existencia de diferencias estructurales entre estas enzimas, a pesar de su cercanía filogenética. La interpretación de las diferencias observadas, como se mencionó en el primer artículo citado de estos autores (Solís-Mendiola *et al.*, 1992), sólo sería posible a través de un estudio estructural topológico detallado de cada proteína involucrada, y el contar ahora con un modelo tridimensional de la quimopapaína nos permitirá intentar esta interpretación. De acuerdo con los resultados calorimétricos comparativos entre quimopapaína y papaína que se obtuvieron en el trabajo de Solís-Mendiola *et al.* en el año de 1993, la quimopapaína presenta una mayor cooperatividad en su proceso de desnaturalización térmica, un menor cambio en su capacidad calorífica durante la transición, y una mayor entalpía de interacción entre sus dominios estructurales.

En relación a la mayor cooperatividad en el proceso de desnaturalización térmica de quimopapaína respecto a papaína, Solís-Mendiola *et al.*, (1993) proponen la existencia de una mayor interacción entre los dominios estructurales de la primera de estas enzimas; de acuerdo con un mecanismo de desnaturalización planteado y desarrollado por ellos mismos. La interpretación estructural de esta conclusión, derivada de los datos experimentales, fue hecha en el presente trabajo a través de la estimación de la cantidad de interacciones interdominio y de su energía correspondiente, mediante el empleo de las coordenadas espaciales atómicas de la estructura cristalográfica de la papaína y del modelo obtenido para la quimopapaína. Las interacciones interdominio consideradas aquí, se refirieron a los puentes de hidrógeno entre diferentes grupos de la cadena polipeptídica y entre moléculas de

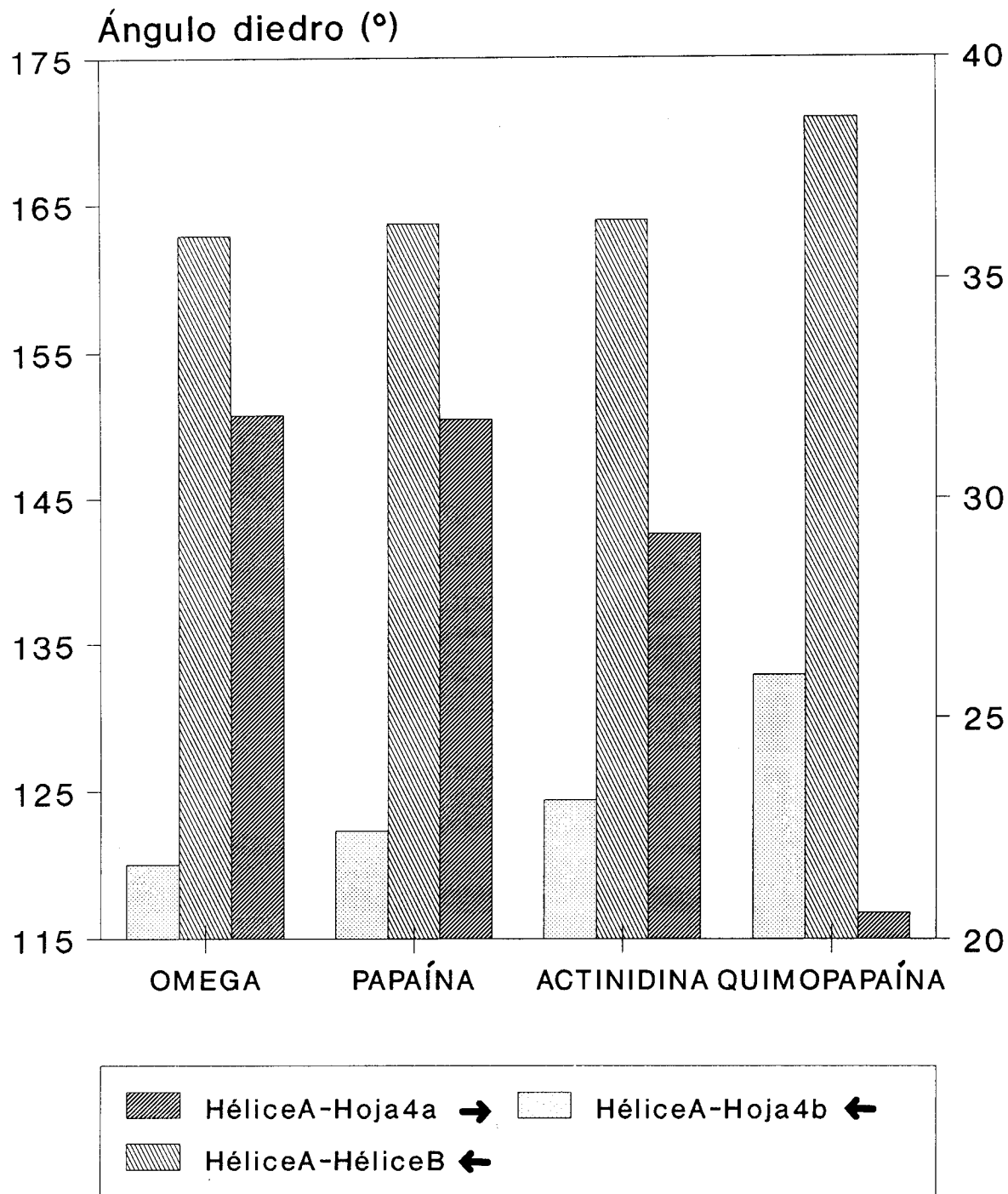


Figura 3-26. Ángulos diedros entre estructuras secundarias en proteinasa Ω , papaína, actinidina y quimopapaína; 4a y 4b se refieren a la primera y segunda porciones de la cuarta hoja β .

agua que se enlazaban directamente con átomos de ambos dominios en la proteína. Otras interacciones interdominio consideradas fueron los contactos van der Waals y las fuerzas electrostáticas (que incluyen puentes salinos). La **Tabla 3-VI** muestra comparativamente el número de puentes de hidrógeno y puentes salinos interdominio para quimopapaína y papaína. Se observa que existe un número mayor de interacciones de ambos tipos para el caso del modelo de la quimopapaína, donde la mayor contribución es debida a los puentes de hidrógeno entre cadenas laterales. Es interesante notar que en el caso de esta última proteasa se encontraron dos puentes salinos entre grupos químicos de dominios diferentes, mientras que para la papaína no se detecta este tipo de interacciones.

TABLA 3-VI
Número de interacciones interdominio.

	PAPAÍNA	QUIMOPAPAÍNA
Puentes de hidrógeno		
Entre átomos de la cadena principal	5	5
Entre átomos de cadenas laterales	2	7
Entre cadena lateral y principal	5	4
TOTALES	12	16
Puentes salinos	0	2

Cabe resaltar que el número de interacciones considerado en la **Tabla 3-VI**, no es necesariamente un reflejo de la energía de atracción entre los dominios estructurales, pero aún en caso afirmativo, existiría la participación adicional de las interacciones van der Waals, cuyo número es difícil estimar. Por esta razón se presenta en la **Tabla 3-VII** la contribución de cada tipo de interacción a la energía potencial total de la región interdominio. Es importante señalar que en estos cálculos se utilizó el mismo campo de fuerzas empleado para el modelado de la quimopapaína, y puesto que cada campo tiene su parametrización particular, fue necesario relajar la estructura cristalográfica de la papaína mediante el programa BIOGRAF para que los resultados fuesen comparables. En esta tabla se anotan las estimaciones energéticas realizadas sobre la papaína antes y después de la relajación.

TABLA 3-VII
Energía de Interacción Interdominio.

	Papaína Cristalográfica	Papaína Minimizada	Quimopapaína
Tipo de Interacción (kcal/mol)			
Van der Waals	-101.71	-132.81	-108.82
Electrostática	-54.42	+12.40	-94.23
Puentes de Hidrógeno	-83.99	-15.39	-107.73
Energía Potencial Interdominio Total	-239.28	-75.80	-309.37

Es muy interesante notar en esta tabla que los dominios de la quimopapaína se encuentran más firmemente unidos entre sí que en el caso de la papaína (tanto cristalográfica como minimizada), a juzgar por la menor energía potencial interdominio total encontrada aquí (mayor valor absoluto).

En la **Fig. 3-27** se ha señalado, sobre las secuencias de papaína y quimopapaína, la posición de los residuos que intervienen en la formación de los puentes de hidrógeno interdominio, a que se refiere la **Tabla 3-VI**. Es notable el papel protagónico que desempeñan el W7 en ambas enzimas, así como la S29 en papaína y la Q19, Y166, y S180 en quimopapaína, ya que intervienen en más de un tipo de interacción interdominio a la vez. Por otra parte, en la **Tabla 3-VIII** se enlistan las moléculas de agua que enlazan a los dominios estructurales en las enzimas estudiadas. Estas moléculas de agua son tanto las determinadas cristalográficamente en el caso de la papaína, como las presentes en la estructura que sirvió de base en la construcción del modelo de la quimopapaína y que permanecieron después de las múltiples etapas de minimización y dinámica molecular. Dado que en solución, ambas proteínas podrían tener un mayor número de moléculas de agua que interaccionaran simultáneamente con los dos dominios estructurales, debe tomarse con cautela el hecho de que exista un número mayor de

QUIMOPAPAÍNA	Y <u>P</u> <u>Q</u> <u>S</u> <u>I</u> <u>D</u> <u>W</u> R AKGAVTPVKN <u>Q</u> GACGSCWAF S TIATVEGINKI	40
PAPAÍNA	I <u>P</u> <u>E</u> <u>Y</u> <u>V</u> <u>D</u> <u>W</u> R QKGAVTPVKN <u>Q</u> GSCGSCWAF S AVVTIEGIKI	40
QUIMOPAPAÍNA	V T GNNLELSEQELVDCDKHSYGCKGGYQTTSL <u>Q</u> YVANNGV	80
PAPAÍNA	R T GNNLNEYSEQELLDCDRRSYGCNGGYPWSALQLVAQYGI	80
QUIMOPAPAÍNA	HTSKVYPYQAKQYKCRATDKPGPKV <u>K</u> ITG Y KRVPSNC E T S	120
PAPAÍNA	HYRNTYPYEGVQRYCRSREKGPYAAKTDGVRQVQPYNEGA	120
QUIMOPAPAÍNA	FLGALANQPL S VLVEAGGKPFQLYKSGVFDGPCGTKLD H A	160
PAPAÍNA	LLYSIANQPVS V VLEAAGKDFQLYRGGIFVGPCGNKVDHA	160
QUIMOPAPAÍNA	V T AVG <u>Y</u> GTSDGKNYIIIKN <u>S</u> WGPNWGE K GYMRLKRQSGNS	200
PAPAÍNA	VAAVG <u>Y</u> GPN----YILIKN <u>S</u> WGTGWGENGYIRIKRGTGNS	196
QUIMOPAPAÍNA	QGT C GVYKSSYYPF K GFA	218
PAPAÍNA	YGV C GLYTSSFY P VKN	212

Figura 3-27. Comparación de los residuos involucrados en la formación de puentes de hidrógeno interdominio en quimopapaína y papaína. Se utilizaron las siguientes convenciones para distinguir los diferentes tipos de puentes de hidrógeno: entre cadena principal, subrayado; entre cadenas laterales, **negritas**; y mixtos, *itálicas*.

estos puentes de agua en quimopapaína, como era de esperarse según los resultados experimentales.

TABLA 3-VIII
Moléculas de agua del interdominio en papaína y quimopapaína.

PAPAÍNA			QUIMOPAPAÍNA		
Dominio 1	Agua	Dominio 2	Dominio 1	Agua	Dominio2
E50	303→307	K174	S4	218	T168, Y174
T33	305	Q128, P129	S4	219	Y166, Y174
E35, K17	306→307	K174	V13	264	Y190
E50	306→307	K174	P15, E35	225	K178
E35	307	K174	G66	289→229	K157
K17	308	N175	G66	289→230	D158
F28	328→308	N175	Q68	241	S131
F28	328→313	K174	V16	251→247	Y190
L72	320	G109, T107	V16	251→250	K188
E3	322	R191	R8	253	E187
V35	392→340	S205	R8	253→254	E187
G20	400→343	W177	Q68	258→274	Y211
R41	339→353	Q128	Q68	258→326	S209
R41	446→353	Q128	V13	264→221	Y190
V16	359→361	N184	V13, S4	265→219	Y166, Y174
D6	384→360	Y170	E35, A32	267	K178
R8	370→364	E183	T69	276	Y110
R8	370	E183	T69	275→276	Y110
R8	482→370	E183	T33	284	P129, N128
K106	372	K211	N18	286	W181
W69	412→390	S205	T42	292	G216
R41	339→453	K211, N212	G20, N18	302→286	W181
V35	392→335	Y186	N18	312	N184
			Q3	341	T168
			Y1	342	E118

Los resultados para la región interdominio de papaína y quimopapaína que son mostrados en las **Tablas 3-VI a 3-VIII** (número de interacciones, energía potencial y puentes de agua), indican que los dominios estructurales tienen una

mayor interacción en quimopapaína que en la papaína, como se esperaba de los resultados calorimétricos mencionados previamente.

Para continuar con la utilización del modelo de quimopapaína en la interpretación estructural de la información termodinámica obtenida experimentalmente por Solís-Mendiola *et al.*, (1993), se emplearon las ideas propuestas por Privalov (1979) respecto a la correlación lineal positiva entre el cambio en la capacidad calorífica (ΔC_p) de una solución de proteína, inducido por su desnaturalización, y la cantidad de contactos hidrofóbicos en la estructura nativa. Para ello se determinaron las densidades moleculares en papaína y quimopapaína, las cuales se presentan en la **Tabla 3-IX**, y donde se aprecia una densidad mayor en cada uno de los dominios estructurales para el caso de la papaína, como sería de esperar para un número más alto de contactos hidrofóbicos en esa enzima, responsables de un mayor ΔC_p .

TABLA 3-IX
Propiedades físicas por dominio.

	Volumen (Å^3)	Superficie (Å^2)	Masa molar (g/mol)	Densidad (g/cm ³)
PAPAÍNA				
Dominio1 (1-107)	10098.6	10737.0	12050	1.9821
Dominio 2 (108-212)	9425.9	10077.6	11295	1.9905
QUIMOPAPAÍNA				
Dominio 1 (1-107)	9826.0	10762.2	11703	1.9784
Dominio2 (108-218)	10115.9	11092.7	11972	1.9659

Una nueva interpretación del cambio en la capacidad calorífica debido al proceso de desnaturalización térmica en proteínas, ha sido propuesta por Murphy *et al.* (1992) y Spolar *et al.* (1992), quienes encontraron que el ΔC_p es directamente proporcional al incremento del área hidrofóbica accesible al disolvente (ΔA_{hf}) generada por el desplegamiento de la cadena polipeptídica según la siguiente expresión:

$$\Delta C_p = 0.24 \Delta A_{hf} \quad (26)$$

Al calcular el ΔA_{hf} para papaína y quimopapaína, se obtuvieron valores de 12,393.7 y 11,625.8 Å^2 , respectivamente. A partir de estos datos y con la ecuación (26), se encontraron los valores de ΔC_p para ambas proteínas que se anotan en la **Tabla 3-X**.

TABLA 3-X
Valores de ΔC_p teórico y calorimétrico.

	PAPAÍNA	QUIMOPAPAÍNA
Valor Estimado (kcal/mol K)	3.0 ± 0.37	2.8 ± 0.35
Valor Experimental (kcal/mol K)	3.3 ± 0.45	2.3 ± 0.50

Como se observa, los resultados teóricos y calorimétricos coinciden razonablemente bien, dentro del error experimental estimado. La discrepancia entre estos dos grupos de valores, es quizá debida a la orientación particular de las cadenas laterales superficiales de estas moléculas, en los conformeros utilizados para estimar el ΔA_{hf} .

CAPÍTULO 4

CONCLUSIONES

Uno de los objetivos más importantes en el estudio de proteínas, es el poder asignar una estructura tridimensional precisa a cualquier secuencia de aminoácidos particular. Como se ha mencionado, actualmente se realizan grandes esfuerzos para incrementar las bases de datos de estructuras tridimensionales de proteínas a través de técnicas experimentales como la difracción de rayos X y RMN, pero en muchas ocasiones, la naturaleza de las macromoléculas hace impracticable la utilización de los métodos experimentales mencionados. Una alternativa viable que ha mostrado ser exitosa en la determinación de estructuras tridimensionales es la simulación o modelado molecular, el cual ha tenido un creciente impacto en la comunidad sólo en años recientes debido a la asequibilidad de procesadores más veloces y de mayor capacidad.

Es importante mencionar que cualquier modelo tridimensional de una proteína representa por sí mismo sólo una propuesta estructural y no debe considerarse completamente equivalente a la geometría real de la molécula modelada, a menos que demuestre ampliamente la validez del modelo. En este trabajo se propone un arreglo tridimensional con detalle atómico de un integrante de la familia de las proteasas sulfhidrúlicas vegetales, la quimopapaína. Este modelo fue sometido a dos criterios de evaluación geométrica independientes: la aplicación de las pautas utilizadas para aceptar estructuras cristalográficas de alta resolución en el Banco de Datos Cristalográficos de Proteínas, y la comparación con información experimental publicada por diferentes autores que se relaciona directa o indirectamente con la estructura tridimensional de esta proteasa. De acuerdo con los resultados del análisis geométrico practicado, el modelo de la quimopapaína obtenido en este trabajo cumplió satisfactoriamente los requerimientos estructurales impuestos por ambos criterios. Cabe señalar que el criterio de calidad estereoquímica empleado en este trabajo, es muy estricto y, aún ciertas estructuras cristalinas de proteínas que se encuentran accesibles en el Banco de Datos, no logran cumplir los estándares. Por otra parte, es importante mencionar que en este primer modelado molecular de la quimopapaína nunca se incluyeron restricciones sobre la geometría inicial, que condujeran a la estructura a cumplir con las observaciones experimentales que sobre la arquitectura de las proteasas sulfhidrúlicas vegetales se dispone.

Los resultados obtenidos en el presente trabajo permiten afirmar que el objetivo de proponer un modelo tridimensional para la quimopapaína, que contara con la calidad suficiente para ser empleado como una estructura más en la familia de las PSV, fue convenientemente alcanzado. En una segunda etapa, el modelo de la quimopapaína fue utilizado para realizar una revisión comparativa con los demás miembros de la familia caracterizados y posteriormente se le empleó en la interpretación a nivel molecular de algunos resultados calorimétricos y espectroscópicos.

En la interpretación estructural de los resultados espectroscópicos de dicroísmo circular sobre los miembros de la familia de las PSV, se proponen en este trabajo dos posibles explicaciones a las inesperadas diferencias observadas en los espectros de estas proteasas de gran homología en secuencia de aminoácidos. Una de las interpretaciones involucra el cambio simultáneo y gradual en la geometría de los tres enlaces disulfuro presentes en cada una de estas proteasas. La otra explicación está relacionada con el cambio paulatino en la orientación relativa de las estructuras secundarias prácticamente invariantes en esta familia de enzimas vegetales. Aunque las dos exégesis por separado podrían explicar el efecto en los espectros de dicroísmo circular, también es posible que la combinación de estas dos interpretaciones fuese compatible con una explicación más general, que correlacionara una distorsión global progresiva de las estructuras con los cambios espectroscópicos. Esto significa que las modificaciones en la geometría de los enlaces disulfuro y en las interacciones entre estructuras secundarias, podrían ser sólo consecuencia de tal distorsión generalizada y no la causa *per se* que generara las diferencias en la forma de los espectros de dicroísmo circular. Para poder discriminar de entre los posibles orígenes de las disparidades espectroscópicas se requiere de mayor información, por lo que en lo sucesivo se analizará con particular atención la incidencia de cromóforos no convencionales en las estructuras tridimensionales de estas proteínas, que pudieran ser responsables de la aparición de señales en la región del ultravioleta lejano en dicroísmo circular.

Con relación a la aplicación del modelo propuesto para interpretar la información calorimétrica sobre papaína y quimopapaína, el hecho más notable es, quizá, la explicación molecular sobre la diferente cooperatividad en el proceso de alteración estructural inducida térmicamente en estas proteínas. Según el mecanismo propuesto para la desnaturalización de ambas enzimas, las interacciones entre sus

dominios estructurales juegan un papel determinante en la cooperatividad del fenómeno, debiendo ser mayores para el caso de la quimopapaína. Al analizar las regiones interdominio en la estructura de la papaína y las correspondientes en el modelo, se encuentra que bajo todos los criterios empleados aquí, (puentes de hidrógeno, puentes salinos, energía de interacción no covalente y puentes de agua) el modelo de la quimopapaína resulta ser congruente con el mecanismo de desnaturalización mencionado y que fue sugerido por otros autores (Solís-Mendiola *et al.*, 1992).

Por otro lado, de la estimación del cambio en el área hidrofóbica expuesta al solvente causada por la desnaturalización térmica en papaína y quimopapaína, se logró estimar en este trabajo el ΔC_p para ambas proteasas con resultados comparables, dentro del error experimental, con los valores determinados calorimétricamente. Estos resultados corroboran las ideas sugeridas por Murphy *et al.* (1992) y Spolar *et al.* (1992), relativas a la dependencia lineal entre los cambios de área hidrofóbica expuesta con la capacidad calorífica de las moléculas al sufrir la pérdida de la estructura nativa. Un trabajo futuro interesante podría ser la inclusión del cambio en el área polar o hidrofílica, inducida por la desnaturalización, en la estimación del ΔC_p como lo han sugerido los autores arriba mencionados, y también podría ser de utilidad el emplear diferentes conformeros de las dos proteínas estudiadas y observar la influencia que esto tenga en los valores del parámetro termodinámico mencionado.

Dentro de las perspectivas para continuar con la línea de investigación del modelado molecular de proteínas, en el Área de Biofísicoquímica se pretende emplear estrategias más sofisticadas en la simulación estructural que comprenden entre otras: la inclusión explícita de moléculas de disolvente, el análisis de las trayectorias obtenidas en la dinámica molecular, simulaciones más prolongadas en el tiempo, uso de las técnicas de recocido y templado para un mejor muestreo del espacio conformacional y uso de nuevos algoritmos para construir las conformaciones iniciales que sirven de molde a la molécula de estructura tridimensional desconocida. Todo esto con la finalidad de obtener estructuras que reflejen más fielmente los detalles geométricos que establecen las diferencias entre la estabilidad y la función de moléculas tan complicadamente precisas como son las proteínas.

BIBLIOGRAFÍA

- Abola E., Bernstein F.C., Bryant S.H., Koetzle T.F. y Weng J. (1987) (En) *Crystallographic Databases-Information Content, Software systems, Scientific Applications*. (Allen F.H., Bergerhoff G. y Sievers R., eds). Data Commission of the International Union of Crystallography. pp. 107-132.
- Altamirano M.M. y Calcagno M.L. (1993) *Mensaje Bioquímico* **Vol. XVII**, Fac. de Medicina, UNAM.
- Arroyo-Reyna A. (1993) UAM-Iztapalapa. *Comunicación personal*.
- Baker E.N. (1980) *J. Mol. Biol.* **141**:441-484.
- Baker E.N. y Drenth J. (1987) (En) *Biological Macromolecules and Assemblies* (Jurnak F.A. y McPherson A., eds.), vol. 3, pp. 313-368. John Wiley & Sons, N.Y.
- Bernstein F.C., Koetzle T.F., Williams G.J.B., Meyer E.F., Brice M.D., Rodgers J.R., Kennard O., Shimanouchi T. y Tasumi M. (1977) *J. Mol. Biol.* **112**: 535-542.
- Biosym Technologies (1993) *Discover User Guide*, Versión **3.1**, California.
- Blundell T.L. (1991) (En) *Protein Conformation*. Ciba Foundation Symposium 161, pp. 28-51. John Wiley & Sons, Chichester.
- Blundell T.L., Sibanda B.L., Sternberg M.J.E. y Thornton J.M. (1987) *Nature* **326**: 347-352.
- Blundell T.L., Carney D., Gardner S., Hayes F., Howlin B., Hubbard J.T., Overington J., Singh D.A., Sibanda B.L. y Sutcliffe M. (1988) *Eur. J. Biochem.* **172**: 513-520.
- Born M. y Oppenheimer J.R. (1927) *Ann. Physik* **84**: 457.
- Branden C. y Tooze J. (1991) *Introduction to Protein Structure*. Garland Publishing, Inc., N.Y.
- Brocklehurst K., Baines B.S. y Malthouse J.P.G. (1981) *Biochem. J.* **197**: 739-746.
- Brookhaven National Laboratory, Protein Data Bank, Quarterly Newsletter No. 61 julio 1992, Upton, N.Y.
- Brooks B.R., Bruccoleri R.E., Olafson B.D., States D.J., Swaminathan S. y Karplus M. (1983) *J. Comp. Chem.* **4**:187-217.
- Browne W.J., North A.C.T., Phillips D.C., Drew K., Vanaman T.C. y Hill R.L. (1969) *J. Mol. Biol.* **42**: 65-86.

- Buttle D.J. y Barret A.J. (1984) *Biochem. J.* **223**: 81-88.
- Carey P.R., Ozaki Y. y Storer A.C. (1983) *Biochem. Biophys. Res. Comm.* **117**:725-731.
- Carne A. y Moore C.H. (1978) *Biochem. J.* **173**:73-83.
- Casey J., y Martin R.B. (1972) *J. Am. Chem. Soc.* **94**: 6141-6151.
- Clore G.M. y Gronenborn A.M. (1991) *Science* **252**:1390-1398.
- Connolly, M.L. (1983) *J. Appl. Crystallogr.* **16**: 548-558; Connolly, M.L. (1983) *Science* **221**: 709-713.
- Daggett V. (1993) (En) *Techniques in Protein Chemistry IV*, publicación auspiciada por The Protein Society (Hogue Angeletti R. ed.), Academic Press, Inc., San Diego California, pp. 525-532.
- Dayhoff M.O. (1978) (En) *Atlas of Protein Sequence and Structure*, 5 suppl. 3:1-8. Nat. Biomed. Res. Found., Washington D.C.
- De Vlieg J. y van Gunsteren W.F. (1991) *Methods Enzymol.* **202**:268-300.
- Doolittle R.F. (1985) *Investigación y Ciencia* **111**:54-65.
- Drenth J., Kalk K.H. y Swen H.M. (1976) *Biochemistry* **15**:3731-3738.
- Dubois T., Kleinschmidt T., Schnek A.G., Looze Y. y Braunitzer G. (1988a) *Biol. Chem. Hoppe-Seyler.* **369**: 741-754.
- Dubois T., Jacquet A., Schnek A.G. y Looze Y. (1988b) *Biol. Chem. Hoppe-Seyler.* **369**:733-740.
- Ebata M. y Yasunobu K.T. (1962) *J. Biol. Chem.* **237**: 1086-1094.
- Enzyme Handbook* (1991) (Schomburg D. y Salzman M. eds.), Springer-Verlag, Berlin Heidelberg.
- Ermer O. (1976) *Structure and Bonding* **27**: 161-211.
- Fasman G.D. (1989) *Prediction of Protein Structure and the Principles of Protein Conformation*. Plenum Press, N.Y.
- Feng D.F., Johnson M.S. y Doolittle R.F. (1985) *J. Mol. Evol.* **21**:112-125.
- Glazer A.N. y Shmith E.L. (1971) (En) *The Enzymes Vol 3* (Boyer P.D. ed.), Academic Press N.Y. pp. 501-546.

- Goraj K., Renard A. y Martial J. (1990) *Prot. Eng.* **3**: 259-266.
- Greer J. (1981) *J. Mol. Biol.* **153**:1027-1042.
- Hagler A.T. (1985) (En) *Conformation in Biology and Drug Design, The Peptides 7*: 213-299 (Merenhofer J., ed.), Academic Press, N.Y.
- Hahn K., Klis W.A. y Stewart J.M. (1990) *Science* **248**: 1544-1546.
- Holmgren A. y Branden C.J. (1989) *Nature* **342**: 248-251.
- Holwerda B.C. y Rogers J.C. (1992) *Plant Physiol.* **99**:848-855.
- Jaquet A., Kleinschmidt T., Schnek A.G., Looze Y. y Braunitzer G. (1989) *Biol. Chem. Hoppe-Seyler* **370**: 425-434.
- Jansen E.F., y Balls A.K. (1941) *J. Biol. Chem.* **137**: 459.
- Kamphuis J.G., Kalk K.H., Swarte M.B.A. y Drenth J. (1985) *J. Mol. Biol.* **179**: 233-256.
- Kamphuis I.G., Drenth J. y Baker E.N. (1985a) *J. Mol Biol.* **182**: 317-329.
- Kamphuis I.G., Kalk K.H., Swarte M.B.A. y Drenth J. (1985b) *J. Mol. Biol.* **179**: 233-256.
- Karplus M. y Petsko G.A. (1990) *Nature* **347**:631-639.
- Kortt A.A., Hinds J.A. y Zerner B. (1974) *Biochemistry* **13**:2023-2028.
- Langone J.J. (ed.) (1991) *Molecular Design and Modeling: Concepts and Applications. Part A. Proteins, Peptides and Enzymes. Methods Enzymol. Vol 202* Academic Press, N.Y.
- Levitt M. (1992) *J. Mol. Biol.* **226**: 507-553.
- López y Celis I. (1993) UAM-Iztapalapa. *Comunicación personal*.
- MayoS.L. Olafson B.D. y Goddard III W.A. (1990) *J. Phys. Chem.* **94**: 8897-8909.
- McIntosh L.P. y Dahlquist F.W. (1990) *Quart. Rev. Biophys.* **23**:1-38.
- Mitchel R.E.J., Chaiken I.M. y Smith E.L. (1970) *J. Biol. Chem.* **245**: 3485-3492.
- More S. (1969) (En) *Structure and Function Relationships of Proteolytic Enzymes*, (Desnuelle P., Neurath H., y Ottesen M. eds), Academic Press N.Y., pp. 289-297.

- Morris A.L., McArthur M.W., Hutchinson E.G. y Thornton J.M. (1992) *Proteins: Struct. Funct. Genet.* **12**:345-364.
- Murphy K.P., Bhakuni V., Xie D. y Freire E. (1992) *J. Mol. Biol.* **227**: 293-306.
- Myers E.W. y Miller W. (1988) *CABIOS* **4**:11-17.
- Needleman S.B. y Wunsch C. D. (1970) *J. Mol. Biol.* **48**:443-453.
- Ottnad M., Ottnad C., Hartter P. y Jung G. (1975) *Tetrahedron* **31**: 1155-1161.
- Pal G.P. y Sinha N.K. (1980) *Arch. Biochem. Biophys.* **202**:321-329.
- Perczel A., Park K. y Fasman G.D. (1992) *Proteins: Struct. Funct. Genet.* **13**:57-69.
- Perutz M. (1992) *Protein Structure. New Approaches to Disease and Therapy.* W.H. Freeman and Co. N.Y.
- Pickersgill R.W., Rizkallah P., Harris G.W. y Goodenough P.W. (1991) *Acta Cryst.* **B47**:766-771.
- Ponder J.W. y Richards F.M. (1987) *J. Mol. Biol.* **193**: 775-791.
- Privalov P.L. (1979) *Adv. Protein Chem.* **33**:167-241.
- Richardson J.S. y Richardson D.C. (1987) (En) *Protein Engineering* (Oxender D.L. y Fox C.F. eds.), Alan R. Liss N.Y., p 149.
- Ritonja A., Rowan A.D., Buttle D.J., Rawlings N.D., Turk V. y Barret A.J. (1989a) *FEBS Letters* **247**:419-429.
- Ritonja A., Buttle D.J. Rawlings N.D., Turk V. y Barret A.J. (1989b) *FEBS Letters* **258**: 109-112.
- Rojo-Domínguez A. y Padilla-Zúñiga A.J. (1993) Memorias del VII Simposio de Estudiantes de Posgrado en Química "Fernando Romo", pp. 78-84. Universidad Autónoma Metropolitana-Iztapalapa.
- Ryu S.E., Kwong P.D. Truneh A., Porter T.G., Arthos J., Rosenberg M., Dai X., Xuong N.H., Axel R., Sweet R.W. y Hendrickson W.A. (1990) *Nature* **348**: 419-426.
- Schwartz R.M. y Dayhoff M.O. (1978) (En) *Atlas of Protein Sequence and Structure*, 5 suppl. 3:353-358. Nat. Biomed. Res. Found., Washington D.C.
- Solís-Mendiola S., Rojo-Domínguez A. y Hernández-Arana A. (1989) *Biochem. J.* **257**: 183-186.

- Solís-Mendiola S., Arroyo-Reyna A. y Hernández-Arana A. (1992) *Biochim. Biophys. Acta* **1118**:288-292.
- Solís-Mendiola S., Rojo-Domínguez A. y Hernández-Arana A. (1993) *Biochim. Biophys. Acta* (en prensa).
- Spolar E.S., Livingstone J.R. y Record M.T. (1992) *Biochemistry* **31**: 3947-3955.
- Srinivasan S., March C.J. y Sudarsanam S. (1993) *Protein Sci.* **2**:277-289.
- Stouch T.R. y Jurs P.C. (1986) *J. Chem. Inform. Comput. Sci.* **26**:4-12.
- Tablero M., Arreguín R., Arreguín B., Soriano M., Sánchez R.I., Rodríguez-Romero A. y Hernández-Arana A. (1991) *Plant Sci.* **74**:7-15.
- Takahashi N., Yasuda Y., Goto K., Miyake T. y Murachi T. (1973) *J. Biochem.* **74**: 355-373.
- Topham C.M., Overington J., Kowlessur D., Thomas M., Thomas E.W. y Brocklehurst K. (1990a) *Biochem. Soc. Trans.* **18**:579-580.
- Topham C.M., Overington J., O'Driscoll M., Salih E., Thomas M., Thomas E.W. y Brocklehurst K. (1990b) *Biochem. Soc. Trans.* **18**: 933-934.
- Valle-Guadarrama M.E. (1993) UAM-Iztapalapa. Comunicación personal.
- Verlet L. (1967) *Phys. Rev.* **159**: 98-103.
- Watson D.C., Yaguchi M. y Lynn K.R. (1990) *Biochem. J.* **266**:75-81.
- Whittier R.F., Dean D.A. y Rogers J.C. (1987) *Nucleic Acids Res.* **15**: 2515-2535.
- Wiener S.J., Kollmann P.A., Nguyen D.T. y Case D.A. (1986) *Comp. Chem.* **7**: 230.