



Casa abierta al tiempo

UNIVERSIDAD AUTONOMA METROPOLITANA

✓ ESTUDIO CONFORMACIONAL Y
MODELADO MOLECULAR DE LA
GLICOPROTEINA ACIDA α_1 HUMANA.

TESIS QUE PRESENTA

✓ ARTURO ROJO DOMINGUEZ
PARA OBTENER EL GRADO DE
✓ DOCTOR EN CIENCIAS.

✓ DICIEMBRE 1992. ✓

✓ ASESOR: ANDRES HERNANDEZ ARANA

UNIVERSIDAD AUTONOMA METROPOLITANA-IZTAPALAPA.

✓ DIVISION DE CIENCIAS BASICAS E INGENIERIA.

DEPARTAMENTO DE QUIMICA.

AREA DE BIOFISICOQUIMICA.

UNIDAD IZTAPALAPA

Av. Michoacán y La Purísima, Col. Vicentina, Iztapalapa, D.F. C.P. 09340. Tel.: 686-03-22 TELEFAX: (5) 686-89-99 TELEX: UAMME 176496

AGRADECIMIENTOS.

Deseo expresar mi agradecimiento a todas las personas e instituciones que cooperaron en la realización de este trabajo. Al Consejo Nacional de Ciencia y Tecnología por el apoyo financiero brindado a través de la beca con número de registro 51302 y de los recursos otorgados al Area de Biofísicoquímica a través de una serie de convenios. Al Centro Nacional de la Transfusión Sanguínea y a los Laboratorios Cutter de México por la generosa donación del sobrenadante de la fracción V del método de Cohn, a partir del cual se aisló y purificó la glicoproteína ácida α_1 usada en parte de estos estudios.

Además agradezco la ayuda recibida de los doctores Marcia E. Newcomer (U. Vanderbilt, Tennessee), Lindsay Sawyer (U. de Edinburgo, Reino Unido), Hazel M. Holden (Instituto de Investigación en Enzimas, U. de Wisconsin), Robert Huber (Instituto de Bioquímica Max-Planck, Alemania) y Xavier Soberón (Instituto de Biotecnología, UNAM) quienes amablemente me enviaron archivos de coordenadas cristalográficas y comentarios de gran utilidad. En particular, estoy muy agradecido en el Dr. H. Brian Halsall (U. de Cincinnati), uno de los especialistas mundiales en el estudio de la glicoproteína α_1 , por su apoyo e interés por que este trabajo pudiera realizarse. Reconozco también la hospitalidad de diferentes instituciones: de la Facultad de Medicina de la UNAM y en especial a los doctores Mario Calcagno y Myriam Altamirano quienes hicieron posible gran parte del trabajo sobre el análisis de secuencias; del Instituto de Química de la UNAM en donde los doctores Adela Rodríguez, Roberto Arreguín y Manuel Soriano me enseñaron y permitieron el uso de sus recursos de cómputo; de las Areas de Físicoquímica Teórica y Química Cuántica del Departamento de Química de la UAM-Iztapalapa, quienes han sido excelentes anfitriones, en especial al Dr. Alberto Vela y al Quím. Rodolfo Alvarez por su auxilio y ayuda en múltiples ocasiones, y al Dr. José Luis Gázquez por su dirección y apoyo desde el inicio de este proyecto.

Valoro y agradezco también los comentarios y sugerencias realizadas por los sinodales de este trabajo y la ayuda de mis compañeros del Area de Biofísicoquímica, a quienes invito a cultivar de manera complementaria a la experimental el análisis y modelado de estructuras de proteínas. Un agradecimiento muy especial para la Q. F. B. Irma Rodríguez y el M. en Q. Rafael Zubillaga por permitirme utilizar algunos de sus resultados en esta tesis y al Biól. Enrique García por las observaciones sobre el manuscrito. Incluyo ahora un lugar especial en estas líneas para reconocer el esfuerzo de otros dos de mis compañeros, sin cuyo esfuerzo y apoyo este trabajo nunca hubiera concluido: al Dr. Andrés Hernández a quien manifiesto mi admiración y agradecimiento y para quien la unión de las siguientes palabras resulta insuficiente para describirlo: maestro, director y amigo; y a la próxima M. en Q. Jaqueline Padilla quien cooperó en los cálculos, en la elaboración tanto del texto como de las figuras de este trabajo y quien ha tenido que soportarme como compañero de trabajo y esposo.

A todos ustedes, junto con mis padres y hermano, quienes siempre han apoyado mi formación, dedico esta tesis.

I N D I C E .

CAPITULO.		PAGINA.
1	INTRODUCCION.	1
	GENERALIDADES.	4
2	ESTUDIOS EXPERIMENTALES.	
	MATERIALES Y METODOS.	7
	RESIDUOS AROMATICOS SUPERFICIALES.	18
	ESTIMACION DE ESTRUCTURAS SECUNDARIAS.	24
	DESNATURALIZACION TERMICA.	27
	DESNATURALIZACION INDUCIDA POR GUANIDINA.	35
	CONCLUSIONES DE LOS RESULTADOS EXPERIMENTALES.	40
3	ANALISIS TEORICO DE LA SECUENCIA DE AMINOACIDOS.	
	METODOS.	45
	COMPOSICION Y SECUENCIA.	53
	PREDICCIÓN DE ESTRUCTURA SECUNDARIA.	65
	HOMOLOGIA.	67
	ALINEAMIENTO.	78
	CONCLUSIONES DEL ANALISIS DE SECUENCIA.	84
4	MODELADO MOLECULAR.	
	METODOS.	87
	PREDICCIÓN DE LA ESTRUCTURA TRIDIMENSIONAL.	95
	ANALISIS DEL MODELO.	100
	COMPARACION CON LOS RESULTADOS EXPERIMENTALES.	103
5	CONCLUSIONES GENERALES.	109
	APENDICE I.	
	LISTADO DEL PROGRAMA TRANSICION.	112
	APENDICE II.	
	ABREVIATURAS Y SIMBOLOS.	118
	BIBLIOGRAFIA.	120

CAPITULO 1.

INTRODUCCION.

El objetivo de este trabajo consiste en estudiar las características estructurales de la glicoproteína ácida α_1 de plasma humano (GPA), también llamada orosomucoide, y proponer un modelo de la estructura tridimensional para esta macromolécula.

Actualmente, la técnica experimental más común para determinar la estructura tridimensional de una proteína es la difracción de rayos X. Este método requiere, además del equipo especializado y una gran experiencia en la interpretación de los datos, que la proteína que desea estudiarse sea capaz de formar cristales estables a los rayos X, con dimensiones mínimas de 0.3 mm y con suficiente orden para difractarlos (Cantor y Schimmel, 1980). La GPA contiene un alto porcentaje de carbohidratos, lo que dificulta su cristalización dado que a altas concentraciones tiende a gelificar (Schmid, 1953; Schmid et al., 1973). Aunque se han logrado obtener algunos cristales de GPA, éstos no han permitido obtener información sobre su estructura (McPherson et al., 1984); esto puede deberse a la microheterogeneidad observada en la GPA, la cual puede tener diferentes aminoácidos en su estructura primaria y distinto tipo de cadenas de carbohidratos.

Una técnica alternativa que permite obtener información estructural de una macromolécula es la resonancia magnética nuclear multidimensional; sin embargo, el análisis de los datos obtenidos es muy complejo, limitando el método al estudio de proteínas de menos de 150 residuos de aminoácidos (Bax, 1989; Clure y Gronenborn, 1991). Esto descarta a la GPA como candidato para este tipo de estudios (por lo menos con la tecnología actual), pues además de estar formada por 183 residuos, contiene aproximadamente 40% en peso de carbohidratos (Schmid, 1975).

Como puede observarse, es deseable buscar un camino alternativo para obtener información estructural de la GPA. El conocimiento más profundo de la estructura de esta glicoproteína podría permitir la determinación de su función metabólica, el estudio de su utilidad farmacológica y en general contribuiría con los esfuerzos que realiza un conjunto cada vez más numeroso de grupos de investigación en todo el mundo, encaminados a descifrar el código de doblado de las proteínas.

La síntesis de una proteína (o enzima) con actividad biológica involucra una serie de etapas que transmiten la información de su estructura desde el ácido desoxirribonucleico hasta la macromolécula final. En estas etapas se utiliza un código que convierte un codón (serie de tres bases en el ácido nucleico) en un aminoácido y posteriormente en un polipéptido; este código fue descifrado en la década de los años sesentas, y se conoce como código genético (Crick et al., 1961). Sin embargo, aunque conocemos los mecanismos generales para la síntesis de una proteína, existe una etapa que se ha resistido a revelar la manera en que la estructura tridimensional (que es muy precisa, biológicamente activa y prácticamente única) se encuentra codificada en la secuencia primaria de la molécula (Goldberg, 1985). Se le ha llamado código de doblado a la serie de reglas que, a nivel de interacciones tanto intramoleculares como intermoleculares con el medio que la rodea, dirigen el proceso de estructuración de una proteína desde la síntesis de su secuencia de aminoácidos hasta la molécula nativa (Ghèlis y Yon, 1982). Se ha observado que para una gran cantidad de moléculas estudiadas, basta proporcionar las condiciones adecuadas de temperatura, pH, fuerza iónica, etc., para que se realice *in vitro* este doblado de la cadena hasta adquirir la estructura nativa (Jerne, 1967; Anfinsen, 1973). Es por esta razón que se infiere que toda la información del producto tridimensional final está contenida unidimensionalmente en el orden en que se unen los diferentes aminoácidos en la secuencia de la proteína.

El conocer el código de doblado, en principio, permitiría conocer qué aminoácidos cambiar, agregar o eliminar para modificar la estabilidad, especificidad o características de una enzima natural, o incluso, diseñar una enzima artificial con funciones de interés biológico, ecológico, médico, industrial o electrónico que no existan actualmente en la naturaleza.

Es así como el análisis de las secuencias de las proteínas gana cada día mayor importancia en el estudio de estas moléculas. La cantidad de métodos y parámetros tanto calculados como empíricos es cada vez mayor (Doolittle, 1990). Actualmente, a partir de estos métodos se pueden conocer múltiples características de una proteína, como por ejemplo: el contenido y localización en la cadena polipeptídica de estructuras secundarias, residuos expuestos al solvente, punto isoeléctrico, coeficiente de absorción, localización de posibles sitios de glicosilación, actividad biológica probable, sitios de hidrólisis por diferentes agentes, zonas de separación entre dominios,

flexibilidad de diferentes zonas de la cadena, entre otras. En este trabajo se aplican algunos de estos métodos a la secuencia de la GPA. Cabe señalar que el uso de estos algoritmos para el análisis de secuencias de proteínas no consiste simplemente en la alimentación a la computadora de la secuencia a estudiar y la posterior obtención de un resultado concluyente (Doolittle, 1986). El análisis en general depende de uno o más parámetros que el usuario asigna de acuerdo a las características de la molécula en cuestión, además de que la información que se obtiene es sujeta a interpretación, siendo el criterio del usuario el que decide qué parte de estos resultados es consistente con otras observaciones experimentales. De esta manera, la información química, biológica y fisicoquímica que se disponga sobre la proteína en estudio resulta de un gran valor para el análisis teórico y para su validación. Después de muchos años de esfuerzo en el estudio de estructuras de proteínas, resulta evidente que sin el conocimiento de sus características termodinámicas, cualquier estudio teórico de su geometría es sólo especulación. Las fuentes de información teórica y experimental son complementarias (Privalov, 1989). Puede obtenerse información sobre la conformación molecular de una proteína mediante el empleo de técnicas espectroscópicas, como son la absorción de radiación ultravioleta y el dicroísmo circular. Con estas herramientas puede conocerse la topografía (esto es, las características de la superficie) de la macromolécula, el contenido de estructuras secundarias y el mecanismo por el cual la cadena de aminoácidos pierde (o adquiere) su extraordinaria estructura nativa, empleando un código para interpretar y convertir la serie de aminoácidos en una conformación tridimensional.

Una característica muy interesante del código de doblado es su degeneración; es decir, se ha observado que dos proteínas pueden diferir en un alto porcentaje de su secuencia de aminoácidos manteniendo su patrón general de doblado prácticamente sin cambios (Doolittle, 1981; Petsko, 1991). Así, existen grupos de proteínas con un porcentaje pequeño de homología en sus secuencias pero con estructuras tridimensionales muy parecidas (Feng *et al.*, 1985; Kuriyan *et al.*, 1991). Estos conjuntos de moléculas pueden compartir una misma función biológica o bien diferir en ella. Incluso se observan casos de proteínas con más de un dominio estructural en donde cada uno de éstos puede pertenecer a un grupo estructural diferente (Richardson, 1981).

La estrategia empleada en este trabajo para estudiar a la glicoproteína ácida α_1 se basó en la aplicación interdisciplinaria de técnicas y estudios experimentales que apoyaron y se vieron enriquecidos por estudios teóricos, tanto de composición y secuencia como de cálculos energéticos a nivel molecular en su etapa final. Las etapas para el estudio de la GPA pueden resumirse en el estudio de sus características conformacionales, de su estabilidad termodinámica y la obtención, con técnicas espectroscópicas, de cualquier información relacionada con su estructura secundaria y tridimensional. Adicionalmente, se realizó el análisis teórico de la composición y secuencia de aminoácidos de esta proteína, la búsqueda de homología con una familia de proteínas, en particular con las lipocalinas con las que se ha reportado similitud (Ikenaka et al., 1972; Pervaiz y Brew, 1987), y la justificación respecto a que si esta homología es suficiente para suponer que parte de la estructura tridimensional de la GPA se asemeja a la de los cuatro miembros de esta familia cuya estructura se conoce. Una vez que se estableció la homología, fue necesario establecer las zonas de semejanza y proponer una estructura plausible para las regiones de la GPA que divergen de las lipocalinas conocidas, empleando la información experimental y biológica disponible, además de métodos de mecánica y dinámica molecular. Finalmente, se analizaron las características del modelo propuesto, comparándolo con la información experimental.

GENERALIDADES.

La glicoproteína ácida α_1 es la proteína de plasma humano con mayor porcentaje de carbohidratos en su estructura (Schmid, 1975). Su concentración normal es de alrededor de 0.8 mg por cada ml de plasma, pero esta cifra puede aumentar hasta en un orden de magnitud en algunos padecimientos o estadios (Dente et al., 1985), por lo que se le clasifica como reactivo de fase aguda. Algunos de estos padecimientos o estadios incluyen procesos inflamatorios, de alta proliferación celular (cirugía mayor, embarazo, cáncer, etc.), o la práctica de ejercicio (Wong y Hsia, 1983).

El interés por el estudio de la GPA inició al final de la década de los años cuarentas cuando se creyó que podría emplearse en la detección del cáncer. Se han sugerido una gran cantidad de posibles funciones metabólicas, pero hasta el momento no se ha logrado determinar con precisión cuál o cuáles de ellas son las que esta molécula realiza (Westphal, 1985). Dentro de estas funciones

sugeridas se encuentran: inmunosupresión (Bennett y Schmid, 1980), interacción con colágeno (Franzblau *et al.*, 1971) y trombina (Andersen *et al.*, 1980), inhibición y promoción de agregación de plaquetas (Andersen y Eika, 1980) y unión de una larga serie de ligandos dentro de los cuales destacan una gran variedad de moléculas hidrofóbicas (como los esteroides y, en especial, progesterona; Ganguly *et al.*, 1967) y drogas básicas (como amitriptilina, bupivacaína, clorpromazina, disopiramida, imipramina, lidocaína, meperidina, nortriptilina, propranolol y quinidina; Goolkasian *et al.*, 1983), llegando incluso a establecerse que algunos medicamentos pueden unirse a esta glicoproteína en altos porcentajes (Aki y Yamamoto, 1989).

Dado el interés fisicoquímico, biológico y médico que tiene la GPA, ha captado la atención de un gran número de grupos de investigación los cuales han establecido varias de sus características estructurales; éstas servirán de punto de partida para el presente trabajo.

Respecto a su estructura, Schmid (1975) publicó una revisión extensiva de las características estructurales de la GPA conocidas hasta ese entonces, dentro de las que destacan: el alto contenido de carbohidratos de esta molécula, con un 40-45% de la masa molar total, correspondiendo al 10% del total de carbohidratos unidos a proteínas del plasma humano; una alta proporción de éstos en forma de ácido siálico; microheterogeneidad periférica, debida posiblemente a diferencias en la porción de carbohidratos; una sola cadena polipeptídica con dos puentes disulfuro ya localizados, cinco residuos de asparagina en donde se unen las cadenas de carbohidratos y la ausencia de un grupo amino terminal libre. Este último aminoácido corresponde a un residuo de ácido glutámico con el carboxilo de la cadena lateral condensado con el amino terminal para formar un residuo de ácido piroglutámico, también conocido como ácido pirrolidincarboxílico.

En el trabajo de revisión de Schmid se menciona que se ha sugerido la existencia de homología entre las inmunoglobulinas y la GPA (Ikenaka, 1972; Schmid *et al.*, 1973; Schmid, 1975). También se describe que la GPA posee un alto reemplazo de aminoácidos: 21 de los 181 residuos totales de la cadena polipeptídica; esta característica de la GPA asemeja la variabilidad en estructura primaria de otras proteínas del plasma, las inmunoglobulinas. Estudios posteriores (Dente *et al.*, 1985), en donde se secuenció uno de los genes que codifican a la GPA, mostraron que el número de aminoácidos de esta proteína es en realidad de 183, encontrándose dos aminoácidos que la

secuenciación de la GPA no detectó en la región formada por los últimos 20 residuos del extremo carboxilo.

Dado que esta macromolécula presenta un punto isoeléctrico inusualmente bajo ($pI=3.5$), debido principalmente a su alto contenido de ácido siálico (la GPA desializada tiene un $pI=5.4$), se ha logrado aislar y purificar empleando precipitación selectiva y cromatografía de intercambio iónico (Hellerstein *et al.*, 1985). La GPA es sintetizada principalmente en el hígado, aunque se ha aislado de diversos fluidos y órganos. Incluso se ha sugerido que puede sintetizarse un precursor de la GPA en los leucocitos (Gahmberg y Andersson, 1978). Además, no sólo se han realizado estudios de la glicoproteína ácida α_1 humana; se conoce la estructura del gen que codifica la de rata (Reinke y Feigelson, 1985; Liao *et al.*, 1985) y la de conejo (Ray y Ray, 1991), y se ha aislado de otros mamíferos como chimpancé, bovino, carnero y caballo, siendo la humana la que contiene el mayor porcentaje de ácido siálico, entre 11 y 12% (Jeanloz, 1972).

En los siguientes capítulos se presentan los estudios realizados sobre la estructura de la GPA empleando diferentes aproximaciones: la aplicación de métodos experimentales (capítulo 2), en particular técnicas espectroscópicas; el análisis teórico de la composición y secuencia de aminoácidos de la parte proteica de la GPA (capítulo 3); y el modelado de su estructura molecular (capítulo 4).

CAPITULO 2.

ESTUDIOS EXPERIMENTALES.

MATERIALES Y METODOS.

En esta sección se describen las técnicas empleadas para obtener los resultados contenidos en este capítulo. Los métodos usados en otros capítulos se describirán posteriormente.

Perturbación por solvente. Los reactivos utilizados para este estudio fueron: dimetilsulfóxido (DMSO) como agente perturbante; como compuestos modelo, los derivados de tirosina y triptofano N-acetil-L-tirosinil etil éster (ATyrEE) y N-acetil-L-triptofanil etil éster (ATrpEE). Todas estas sustancias fueron obtenidas de la compañía Merck (Darmstadt, Alemania). El resto de los compuestos usados fueron al menos grado reactivo.

La glicoproteína ácida α_1 (GPA) empleada se aisló en nuestro laboratorio por el método de Bezkorovainy y Winzler (1961) a partir del sobrenadante de la fracción V del método de Cohn (1946) para el fraccionamiento del plasma humano. La homogeneidad de la preparación de GPA fue probada por electroforesis en gel de poliacrilamida en presencia de dodecilsulfato de sodio y por cromatografías de filtración en gel e intercambio iónico de alta resolución. En los tres casos se observó una sola banda o bien un único pico cromatográfico simétrico. Para eliminar cualquier ligando hidrofóbico unido a la GPA, ésta fue deslipidada según el procedimiento descrito por Ganguly *et al.* (1976). El dimetilsulfóxido fue seleccionado como agente perturbante debido a su propiedad de provocar la mayor intensidad de las bandas diferenciales en los espectros ultravioleta reportados (Herskovitz y Sorensen, 1968 a,b). El efecto del DMSO sobre el espectro de absorción diferencial de la GPA se analizó a concentraciones de 5, 10, 15 y 20% en volumen.

Todas las disoluciones empleadas en este estudio fueron preparadas en regulador de fosfatos 0.01M, pH 6.8 con KCl 0.10M. Las muestras de GPA con que se obtuvieron los espectros diferenciales se encontraban a concentraciones entre 27.4 μM (1.12 mg/ml) y 60.1 μM (2.46 mg/ml). Las disoluciones de compuestos modelo fueron empleadas a concentraciones de 9.90×10^{-4} M en el caso de ATyrEE y 2.34×10^{-4} M para el ATrpEE. Todas las concentraciones fueron determinadas por su absorción en el ultravioleta, utilizando los

siguientes coeficientes de extinción: $36610 \text{ M}^{-1}\text{cm}^{-1}$ a 278 nm ($A_{1\text{cm}, 278\text{nm}}^{1\%} = 8.93$) para la GPA (Schmid, 1975); $5550 \text{ M}^{-1}\text{cm}^{-1}$ a 282 nm para ATrpEE y $1340 \text{ M}^{-1}\text{cm}^{-1}$ a 274.5 nm para el ATyrEE (Herskovitz y Sorensen, 1968a).

Para determinar los espectros diferenciales se utilizó un espectrofotómetro VARIAN DMS80 con celdas de doble cámara. El recorrido óptico en cada cámara fue de 1 cm . Como se muestra en la figura 2-1a, en la celda de referencia se colocaron soluciones de GPA y DMSO en cámaras separadas para evitar su interacción. En la celda de muestra, ambas sustancias se mezclaron en una misma cámara llenando con solvente la restante. Cada espectro fue obtenido por triplicado en la región de 240 a 320 nm utilizando un ancho de banda espectral de 0.5 nm y una velocidad de barrido de 10 nm/min . La temperatura a la cual se realizaron los experimentos fue de 25°C en todos los casos.

Estimación de Estructuras Secundarias. Los espectros de dicroísmo circular de la GPA y la β -lactoglobulina (BLG) fueron obtenidos en un espectropolarímetro JASCO J500A calibrado con ácido (+)-10-camforsulfónico (Hennessey y Johnson, 1982), en la región de 180 a 260 nm , a 25°C . La ventana de barrido utilizada (producto de la velocidad de barrido por la constante de tiempo) fue de 0.7 nm , mientras que el voltaje dinódico (que alimenta al fotomultiplicador y amplificador de señal para mantener una corriente constante) siempre fue menor a 900 V . La elipticidad por residuo medio, $[\theta]_{\text{rm}}$, se calculó con base en una masa molecular promedio por residuo de 117.8 Da para la GPA (Dente et al., 1985), y de 112.6 Da en el caso de la BLG (Lee et al., 1987). Ambas proteínas empleadas en estos estudios fueron obtenidas a través de la Compañía Sigma (St. Louis MO, Estados Unidos). A la glicoproteína le fueron removidos los lípidos que originalmente une, por el método de Ganguly et al. (1976), mientras que la BLG se utilizó sin tratamientos posteriores. Los estudios se realizaron en disoluciones preparadas con regulador de fosfatos 0.01 M , a un valor de pH de 7.4 y con concentraciones de proteína de aproximadamente 0.1 mg/ml en celdas de cuarzo de 0.05 cm de recorrido óptico. La concentración precisa de las soluciones se determinó a partir de su absorción a 278 nm , utilizando los coeficientes de extinción de $A_{1\text{cm}, 278\text{nm}}^{1\%} = 8.93$ y 9.60 para la GPA y BLG respectivamente (Schmid, 1975; Lapanje y Polkar, 1989). La estimación de estructuras secundarias a partir de los espectros de dicroísmo circular de ambas proteínas, se realizó utilizando el método de Perczel et al. (1991; 1992).

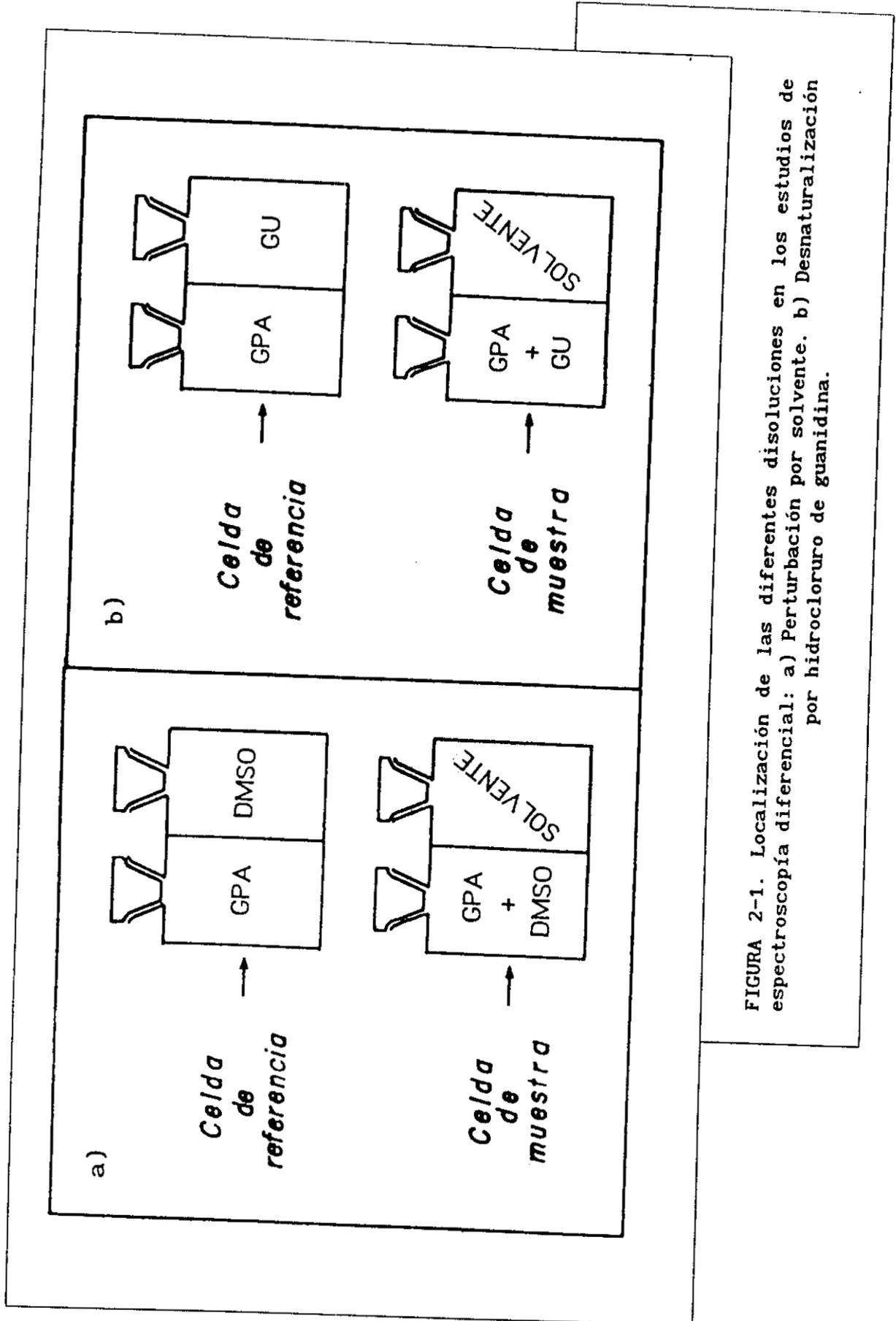


FIGURA 2-1. Localización de las diferentes disoluciones en los estudios de espectroscopía diferencial: a) Perturbación por solvente. b) Desnaturalización por hidrocóloruro de guanidina.

Estudios de Desnaturalización en el Equilibrio. La purificación de la GPA al igual que las determinaciones de homogeneidad y concentración fueron realizados de la manera descrita anteriormente para los estudios de perturbación por solvente. El hidrocloreuro de guanidina (Gu), grado I, fue obtenido de la Compañía Sigma (St. Louis MO, Estados Unidos). Su concentración se determinó a partir del índice de difracción de sus disoluciones, de acuerdo al método de Nozaki (1972). El resto de los reactivos utilizados fueron grado analítico.

Espectroscopía diferencial ultravioleta. Para estos estudios se empleó un espectrofotómetro VARIAN DMS80 de doble haz equipado con graficador de escala expandida. Los espectros diferenciales de la desnaturalización térmica de la GPA, se obtuvieron utilizando celdas espectroscópicamente equivalentes de 1 cm de recorrido óptico. La temperatura en cada celda fue controlada independientemente, con una precisión de $\pm 0.2^{\circ}\text{C}$, por medio de una camisa de circulación de agua, conectada a un baño termostático con recirculador. La celda de referencia fue mantenida a 15°C , mientras que la de muestra se calentó a una velocidad menor a $1^{\circ}\text{C}/\text{min}$. Al alcanzar la temperatura del experimento, la muestra fue incubada durante 10 minutos en esas condiciones. Se determinó que este intervalo de tiempo es suficiente para que la proteína alcance los equilibrios térmico y conformacional. Las muestras estudiadas fueron disueltas en regulador de glicilglicina 0.05 M a un valor de pH de 5.25 y a una concentración de 0.12 mg/ml.

En los estudios diferenciales de la desnaturalización inducida por Gu se emplearon celdas de doble cámara, similares a las empleadas en los estudios de perturbación por solvente, de 1 cm de recorrido óptico por cámara a una temperatura de $20 \pm 0.2^{\circ}\text{C}$. Como se muestra en la figura 2-1b, en la celda de muestra se colocaron una mezcla de proteína y desnaturalizante en una cámara y solvente en la otra. Por otra parte, en la celda de referencia se colocaron soluciones de proteína y desnaturalizante a la misma concentración que en el compartimiento de muestra pero en cámaras separadas. Las diferentes disoluciones de proteína empleadas se prepararon mezclando volúmenes idénticos de una solución patrón de GPA, diferentes volúmenes de una disolución de Gu 8 M y la cantidad necesaria de regulador de acetatos 0.01 M (pH 5.0) para obtener el mismo volumen total. La concentración final de GPA fue de 0.18

mg/ml. Todas estas muestras fueron incubadas durante 8 horas a 4°C antes de efectuar los experimentos.

Dicroísmo circular. Los espectros de dicroísmo circular fueron obtenidos de la manera descrita en la estimación de estructuras secundarias. La temperatura en la celda fue controlada, al igual que en la espectroscopía diferencial, con una camisa de circulación de agua conectada a un baño termostático y con una precisión de $\pm 0.2^\circ\text{C}$. Tanto en los experimentos de espectroscopía diferencial como en los de actividad óptica, la temperatura en la celda fue determinada con un termistor. Las muestras estudiadas se prepararon de la manera descrita en la sección de espectroscopía diferencial.

Análisis Termodinámico de la Desnaturalización Térmica. Al estudiar la estabilidad térmica de una proteína se requiere del empleo de una técnica experimental que posea varias características: ser sensible a los cambios en la conformación de la macromolécula; permitir el control adecuado de la temperatura de la muestra; y tener suficiente sensibilidad para detectar los cambios en una solución muy diluida (del orden de 0.1 mg/ml) de proteína, esto último debido a los problemas de agregación que pueden presentarse a concentraciones mayores y que modifican los parámetros termodinámicos obtenidos del experimento (Ghélis y Yon, 1982).

Algunas técnicas usuales en estos estudios son: la espectroscopía diferencial ultravioleta, el dicroísmo circular, la reactividad de algún grupo químico de la molécula y la pérdida de la actividad biológica. En cualquiera de estos casos, experimentalmente se siguen los cambios de un parámetro en función de la temperatura. Llamemos Y a este parámetro, que puede ser la magnitud del espectro diferencial, elipticidad, reactividad química, etc. En todos los casos se obtiene una curva de Y contra T en la que cada punto experimental tiene una incertidumbre alta debido, entre otras razones, a las condiciones de baja concentración de la muestra en estos experimentos (Cantor y Schimmel, 1980).

Si a la curva obtenida (Y contra T) se le aplica el modelo de una etapa, también conocido como de dos estados, en el cual se considera la existencia en concentración apreciable de sólo dos especies, la nativa (N) y la desnaturalizada (D), se obtiene lo siguiente:

Consideremos al parámetro Y aditivo, de tal forma que a una temperatura T su valor es la suma de las contribuciones de los estados N y D :

$$Y = f_N Y_N + f_D Y_D \quad (2-1)$$

en donde f_N y f_D representan las fracciones de cada uno de esos estados, cumpliéndose que $0 \leq f_N, f_D \leq 1$ y $f_N + f_D = 1$. Los símbolos Y_N y Y_D representan el valor de Y para la muestra en estado nativo y en estado desnaturalizado respectivamente; estos valores en general cambian con la temperatura, aunque esta dependencia es pequeña en la mayoría de los casos. Para estimarla, comúnmente se extrapola la tendencia observada a temperaturas bajas a toda la zona de la transición para Y_N y análogamente en las temperaturas altas para Y_D . De las ecuaciones anteriores se encuentra que para cada temperatura del experimento:

$$f_D = \frac{(Y - Y_N)}{(Y_D - Y_N)} \quad (2-2)$$

Al construir la curva de transición, f_D contra T , se observa una dependencia sigmoideal de la fracción desnaturalizada con tres zonas: la inicial y la final con pendientes pequeñas, y una región intermedia donde los cambios son abruptos y corresponden al grueso del cambio conformacional. Conociendo f_D , la constante del equilibrio $N \rightleftharpoons D$ puede calcularse fácilmente, pues:

$$K = \frac{[D]}{[N]} = \frac{f_D}{f_N} = \frac{f_D}{1 - f_D} \quad (2-3)$$

Dado que podemos conocer el valor de f_D para cada temperatura del experimento, tenemos una serie de valores para la constante de equilibrio a diferentes temperaturas. Puesto que:

$$K = \exp \left(\frac{-\Delta G}{RT} \right) \Rightarrow \ln K = \frac{-\Delta G}{RT} \quad (2-4)$$

$$\text{y } \Delta G = \Delta H - T\Delta S \quad (2-5)$$

se tiene que
$$\ln K = \frac{-\Delta H}{R} \frac{1}{T} + \frac{\Delta S}{R} \quad (2-6)$$

Así, de una curva de van't Hoff ($\ln K$ contra $1/T$) puede calcularse la pendiente, que es $-\Delta H/R$. Si el valor del cambio en la capacidad calorífica de la proteína, ΔC_p , fuese un valor muy pequeño o cero, ΔH sería constante en todo el intervalo del gráfico, la pendiente del mismo sería constante y tendríamos una línea recta. En realidad, para las proteínas el valor de ΔC_p es

mayor que cero, y se estima ΔH de la pendiente de la curva de van't Hoff a la temperatura media de la transición, T_m , es decir a la temperatura a la cual $[N] = [D]$ ($K = 1$). De ese valor de ΔH a la T_m , que se denota ΔH_m , y del valor de ΔH a otras temperaturas, que se obtienen de la pendiente de la curva en diferentes puntos, se puede determinar la variación de ΔH con T . Dado que

$$\Delta H = \Delta H_m + \Delta C_p(T - T_m), \quad (2-7)$$

de la pendiente de un gráfico de ΔH contra T puede estimarse el valor de ΔC_p . Como se observa, la adquisición de datos termodinámicos a partir de valores experimentales, aunque posible teóricamente, conlleva una serie de transformaciones a cocientes, logaritmos de cocientes, pendientes y derivadas de éstas, en donde la incertidumbre inicial de los puntos crece en cada transformación, principalmente en los datos del inicio y del final de la transición.

Se propone en este trabajo el siguiente procedimiento: no transformar los puntos experimentales, sino encontrar una ecuación donde la relación entre f_D y T sea a través de los parámetros termodinámicos. Para tal efecto, de la ecuación (2-3) obtenemos:

$$f_D = \frac{K}{K + 1} \quad (2-8)$$

y sustituyendo (2-4) en (2-8), se tiene

$$f_D = \frac{\exp(-\Delta G/RT)}{1 + \exp(-\Delta G/RT)}$$

Aplicando (2-5), esta última ecuación puede escribirse como:

$$f_D = \frac{\exp\left(\frac{-\Delta H}{R} \frac{1}{T} + \frac{\Delta S}{R}\right)}{1 + \exp\left(\frac{-\Delta H}{R} \frac{1}{T} + \frac{\Delta S}{R}\right)} \quad (2-9)$$

Por otro lado, $\Delta H = \Delta H_m + \Delta C_p \cdot (T - T_m), \quad (2-7)$

$$\Delta S = \Delta S_m + \Delta C_p \cdot [\ln(T/T_m)] \quad (2-10)$$

Ya que $\Delta G = \Delta H - T\Delta S \quad (2-11)$

y sabiendo que cuando $T = T_m$, $\Delta G = 0$, entonces ΔS_m puede expresarse:

$$\Delta S_m = \frac{\Delta H_m}{T_m} \quad (2-12)$$

Al sustituir (2-7), (2-10) y (2-12) en (2-9), obtenemos:

$$f_D = \frac{\exp\left(\frac{\Delta H_m}{RT} + \frac{\Delta C_p(T-T_m)}{RT} + \frac{\Delta H_m}{RT_m} + \frac{\Delta C_p[\ln(T/T_m)]}{R}\right)}{1 + \exp\left(\frac{\Delta H_m}{RT} + \frac{\Delta C_p(T-T_m)}{RT} + \frac{\Delta H_m}{RT_m} + \frac{\Delta C_p[\ln(T/T_m)]}{R}\right)}$$

factorizando el argumento de las exponenciales tenemos que,

$$f_D = \frac{\exp\left\{\frac{\Delta H_m}{RT_m}\left(1 - \frac{T_m}{T}\right) - \frac{\Delta C_p}{R}\left(1 - \frac{T_m}{T}\right) + \ln\left(\frac{T}{T_m}\right)^{\frac{\Delta C_p}{R}}\right\}}{1 + \exp\left\{\frac{\Delta H_m}{RT_m}\left(1 - \frac{T_m}{T}\right) - \frac{\Delta C_p}{R}\left(1 - \frac{T_m}{T}\right) + \ln\left(\frac{T}{T_m}\right)^{\frac{\Delta C_p}{R}}\right\}}$$

y finalmente, agrupando términos comunes, f_D resulta estar descrita por:

$$f_D = \frac{\exp\left\{\left(\frac{\Delta H_m}{RT_m} - \frac{\Delta C_p}{R}\right)\left(1 - \frac{T_m}{T}\right)\right\} \left(\frac{T}{T_m}\right)^{\frac{\Delta C_p}{R}}}{1 + \exp\left\{\left(\frac{\Delta H_m}{RT_m} - \frac{\Delta C_p}{R}\right)\left(1 - \frac{T_m}{T}\right)\right\} \left(\frac{T}{T_m}\right)^{\frac{\Delta C_p}{R}}} \quad (2-13)$$

Esta expresión indica la dependencia de f_D con la temperatura, en función de los parámetros termodinámicos por determinar: ΔH_m , T_m y ΔC_p . Con esta ecuación puede realizarse una regresión no lineal con los valores de f_D y las correspondientes temperaturas, determinando los valores de las magnitudes termodinámicas que minimizan la dispersión de los puntos experimentales a la función que acabamos de deducir (2-13).

Este método tiene la ventaja de que los puntos experimentales no requieren pasar por la serie de transformaciones que se emplean comúnmente para la construcción de la curva de van't Hoff, la obtención de las derivadas de ese gráfico y la regresión lineal de estas últimas para estimar ΔC_p .

Para efectuar la regresión no lineal se desarrolló un programa de cómputo (el listado de éste se encuentra en el apéndice I) que, además de estimar ΔH_m , ΔC_p y T_m , puede calcular la curva teórica de la transición y estimar algunos otros valores. Por ejemplo: al sustituir las ecuaciones (2-7), (2-10) y (2-12) en (2-11), se obtiene:

$$\Delta G = \Delta H_m - T \cdot \frac{\Delta H_m}{T_m} + \Delta C_p \cdot [T - T_m - T \cdot \ln(T/T_m)] \quad (2-14)$$

es decir, la dependencia de la energía libre de Gibbs con la temperatura, de donde puede calcularse el valor a 25°C para compararlo con el de otras proteínas, o bien toda la curva de ΔG contra T , en el intervalo del estudio e incluso su extrapolación fuera de esa zona.

La curva de ΔG contra la temperatura presenta un valor máximo cuando

$$T = T^{\bullet} = T_m \cdot \exp(-\Delta H_m / (T_m \Delta C_p)) \quad (2-15)$$

T^{\bullet} se conoce como temperatura de máxima estabilidad y típicamente se encuentra entre los -10 y 15°C; a esta temperatura la estabilidad de la conformación nativa es máxima.

Análisis Termodinámico de la Desnaturalización por Hidrocloruro de Guanidina.

Los datos obtenidos en estos experimentos fueron analizados empleando las siguientes ecuaciones:

$$f_D = \frac{(Y - Y_N)}{(Y_D - Y_N)} \quad (2-2)$$

$$K_D = \frac{f_D}{1 - f_D} \quad (2-3)$$

$$\Delta G_D = -RT \cdot \ln K_D \quad (2-4)$$

Estas permiten calcular los valores de ΔG_D a diferentes concentraciones de desnaturalizante. La extrapolación de este valor a concentración nula de Gu, $\Delta G_D^{H_2O}$, fue realizada de dos maneras diferentes. En la primera se supuso una dependencia lineal del ΔG_D con la concentración del desnaturalizante en todo el intervalo de concentraciones (Pace, 1975), es decir:

$$\Delta G_D = \Delta G_D^{H_2O} - m[Gu] \quad (2-16)$$

donde [Gu] representa la concentración de hidrocloreuro de guanidina. De acuerdo con Pace (1975), el coeficiente m en esta ecuación describe qué tan abrupta es la transición, es decir, la cooperatividad del proceso de desnaturalización.

El segundo método empleado, se basa en el modelo propuesto por Tanford (1970; Pace, 1975). Este se describe por la siguiente ecuación:

$$\Delta G_D = \Delta G_D^{H_2O} + \alpha \cdot \sum_i \delta g_{tr,i} \quad (2-17)$$

donde α es el cambio promedio en el grado de exposición de los diferentes grupos químicos presentes en la proteína; la suma se efectúa sobre las energías libres de transferencia, $\delta g_{tr,i}$, de cada uno de los i grupos (esto es, cada cadena lateral y cada fragmento de cadena principal) del agua a una disolución de Gu de la misma concentración a la que el ΔG_D fue determinado.

Los valores de $\delta g_{tr,i}$ fueron estimados, por medio de una interpolación parabólica, de los valores reportados por Pace (1975). De una regresión lineal de los valores de ΔG_D en función de $\sum_i \delta g_{tr,i}$, se obtuvo el valor de α así como la estimación del $\Delta G_D^{H_2O}$.

Análisis de Factores. Para determinar si los datos obtenidos de los espectros de desnaturalización se ajustaban a un modelo de dos estados, se analizaron empleando el método de descomposición en vectores propios conocido como análisis de factores (Malinowski y Howery, 1980). Esta técnica se emplea para calcular el número de componentes o factores que pueden reproducir, por combinación lineal, un conjunto de datos espectroscópicos (Hennessey y Johnson, 1981). Los espectros tanto de experimentos de espectroscopía diferencial como de dicroísmo circular, se agruparon en una matriz de datos, [D], de m renglones por n columnas. Cada columna representa uno de los espectros obtenido a una temperatura determinada. A continuación se construyó una matriz cuadrada multiplicando por la izquierda a [D] por su transpuesta. Este arreglo se diagonalizó para encontrar la matriz cuadrada, de dimensión n , que contiene los vectores propios del conjunto de datos, [Q], y la matriz diagonal de valores propios, [λ], esto es

$$[Q]^T [D]^T [D] [Q] = [\lambda]$$

De esta forma, la matriz de datos original se puede reproducir exactamente por

el producto de dos matrices:

$$[D] = [U][Q]^T$$

de donde

$$[U] = [D][Q]$$

Sin embargo, la intención es reproducir la matriz de datos original, dentro de la incertidumbre experimental, con el mínimo número de vectores propios. Para ello se seleccionan los vectores propios más importantes, es decir, aquéllos con los valores propios mayores. A continuación se calcula la desviación estándar residual (DER) de acuerdo con la expresión:

$$DER = \left[\frac{1}{m(n-f)} \sum_{i=f+1}^n \lambda_i \right]^{1/2}$$

en donde f es el número de factores seleccionados para describir a la matriz de datos original, y los valores λ_i son los valores propios de la diagonal de la matriz $[\lambda]$. El valor final de f determinado es el más pequeño que produce una DER menor o igual a la estimación de la incertidumbre experimental.

RESIDUOS AROMATICOS SUPERFICIALES.

Entre la información estructural sobre la GPA se encuentran varios análisis topográficos; en éstos se ha descrito la localización de algunos de sus grupos químicos con respecto a su grado de exposición al solvente. Estudios de titulación espectrofotométrica (Svobodová *et al.*, 1977) determinaron la presencia de tres tirosinas expuestas al solvente y nueve ocultas. En otro trabajo, la reactividad química de las tirosinas indicó que seis de ellas se encuentran accesibles al solvente, cuatro ocultas y dos en un estado intermedio (Schmid *et al.*, 1976). También en este estudio se sugiere que sólo uno de los tres residuos de triptofano presentes en la GPA es parcialmente reactivo. Puede apreciarse que el número de cadenas laterales de tirosina expuestas al solvente difiere en ambos casos, razón por la cual se aplicó un tercer método de análisis para estimar el grado de exposición tanto de los residuos de tirosina como de triptofano expuestos al solvente (Zubillaga-Luna *et al.*, 1991). Esta técnica es la perturbación del espectro de absorción ultravioleta de la GPA, en la cual se propone que el espectro diferencial obtenido por la interacción de la proteína con el agente perturbante (DMSO en nuestro caso) puede ser descrito por la combinación lineal de los espectros diferenciales de compuestos modelo de tirosina y triptofano en las condiciones del experimento. En estos estudios es indispensable que la conformación de la proteína no se vea afectada significativamente por el agente perturbante, pues una desnaturalización de la GPA aumenta el grado de exposición al solvente de algunos de los residuos de aminoácido en estudio. Para demostrar que la GPA no sufre cambios conformacionales drásticos en el intervalo de concentraciones de DMSO empleado, se determinó la intensidad de la perturbación con el aumento de la concentración del perturbante por medio de su absorción diferencial. Puede observarse de la figura 2-2 que existe una relación lineal entre la magnitud de la perturbación y la concentración del DMSO, por lo que se concluye que este perturbante no induce desnaturalización en la glicoproteína en concentraciones de hasta 20% en volumen; ésta fue la concentración de DMSO empleada en los experimentos que aquí se discuten.

Los espectros diferenciales de los compuestos modelo, ATrpEE y ATyrEE, en DMSO al 20% se muestran en la figura 2-3. Estos espectros son esencialmente iguales a los reportados previamente (Herskovitz y Sorensen, 1968a). La determinación de los residuos de tirosina y triptofano expuestos se basa en la suposición de

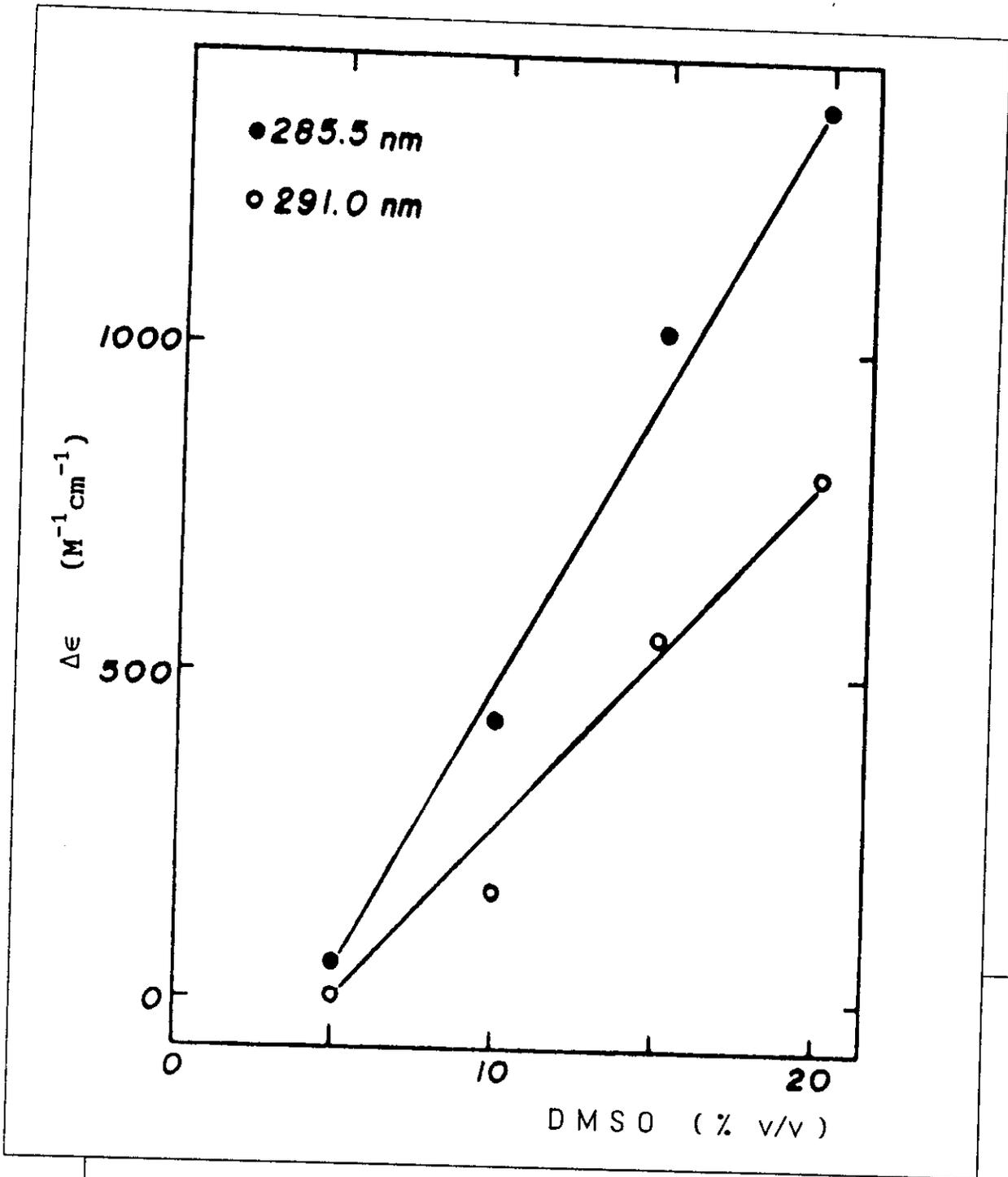


FIGURA 2-2. Absorción diferencial molar de la glicoproteína ácida α_1 en función de la concentración en volumen del agente perturbante.

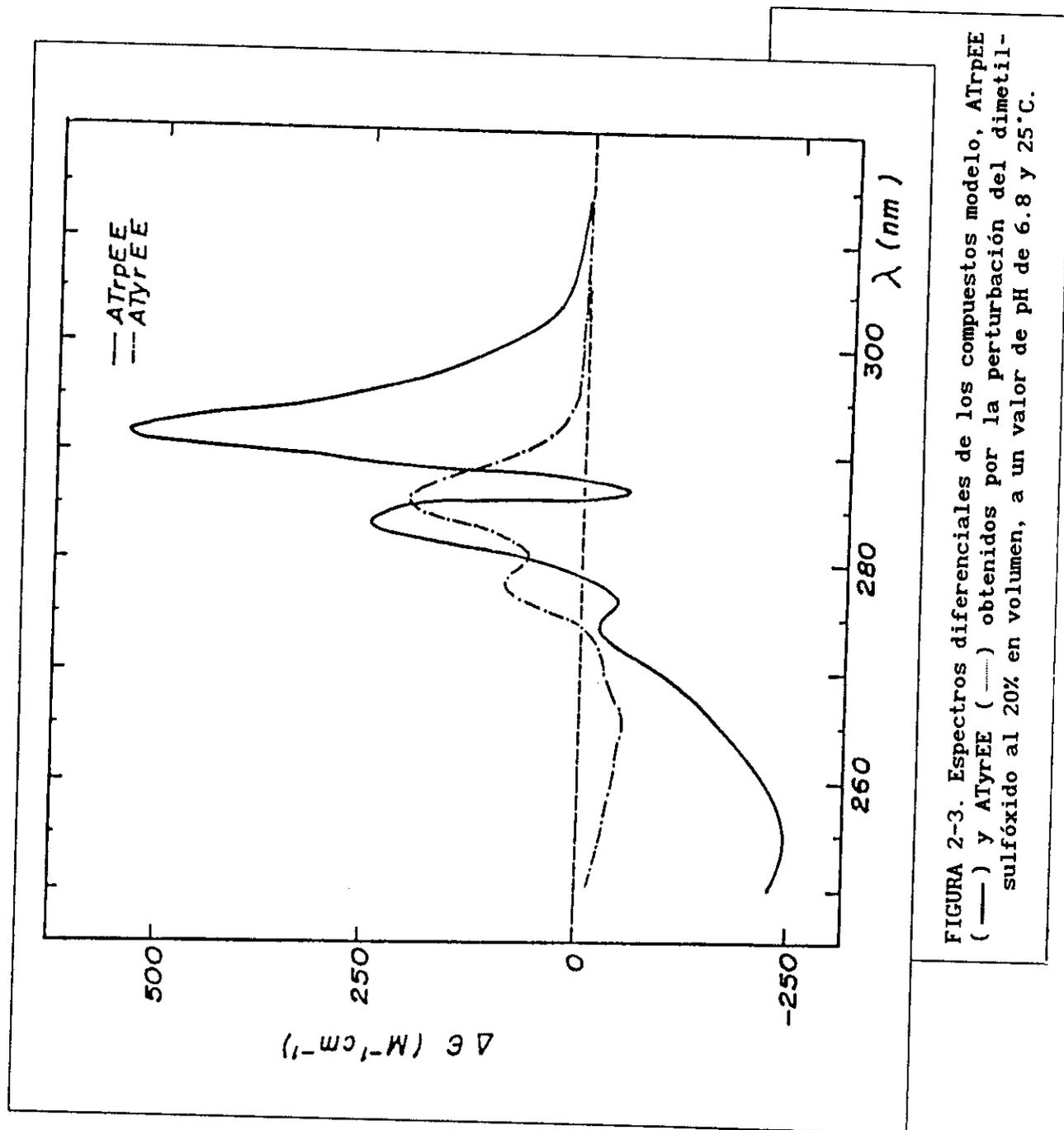


FIGURA 2-3. Espectros diferenciales de los compuestos modelo, ATrpEE (—) y ATyrEE (---) obtenidos por la perturbación del dimetilsulfóxido al 20% en volumen, a un valor de pH de 6.8 y 25°C.

que el espectro diferencial obtenido por la perturbación de la proteína puede expresarse como una combinación lineal de los espectros diferenciales de los compuestos modelo en las mismas condiciones. Esto es, que las perturbaciones son aditivas y que si se eligen dos longitudes de onda, λ_1 y λ_2 , se cumplen las siguientes relaciones:

$$\Delta\epsilon_{\lambda_1}(\text{GPA}) = \Delta\epsilon_{\lambda_1}(\text{ATrpEE}) \cdot X + \Delta\epsilon_{\lambda_1}(\text{ATyrEE}) \cdot Y$$

$$\Delta\epsilon_{\lambda_2}(\text{GPA}) = \Delta\epsilon_{\lambda_2}(\text{ATrpEE}) \cdot X + \Delta\epsilon_{\lambda_2}(\text{ATyrEE}) \cdot Y$$

donde $\Delta\epsilon_{\lambda}$ representa la diferencia de absorptividad molar a una determinada longitud de onda (λ) para la proteína (GPA) y los compuestos modelo de triptofano y tirosina; mientras que X e Y son el número de residuos de triptofano y tirosina expuestos al solvente. Las longitudes de onda seleccionadas fueron 285.5 y 291.0 nm, correspondientes a los máximos de absorción de los espectros diferenciales de los compuestos modelo. Los valores de las diferentes absorptividades molares para cada una de las especies se presentan en la tabla 2-I.

T A B L A 2 - I

Diferencias en la absorción molar para la GPA y los compuestos modelo a pH 6.8 y 25°C por la perturbación de DMSO al 20% en volumen.

λ (nm)	$\Delta\epsilon$ ($M^{-1} \cdot cm^{-1}$)		
	GPA	ATyrEE	ATrpEE
285.5	1414.2	203.0	199.9
291.0	834.9	48.5	535.3

Al resolver el sistema de ecuaciones antes mencionado, se encuentra que el número de triptofanos expuestos es de 0.98 y el de tirosinas de 6.4. Los valores fraccionarios indican que alguno o algunos de estos residuos se encuentran accesibles al solvente de manera parcial. En la tabla 2-II se presentan estos resultados comparados con los de otros trabajos en los que se emplearon diferentes técnicas experimentales.

T A B L A 2 - II
Residuos de triptofano y tirosina expuestos en la GPA.

Método	Triptofanos	Tirosinas	Referencia
Reactividad química	$< 1^a$	6 y 2 semi-expuestas	Schmid, et al., 1976.
Titulación espectrofotométrica	-- ^b	3	Svobodová, et al., 1977.
Espectroscopía diferencial por perturbación de solvente	0.98	6.4	Este trabajo

^a Un valor menor a la unidad indica que la reactividad fue sólo parcial.
^b Este método no determina el número de residuos de triptofano expuestos

En la figura 2-4 se compara la combinación lineal de los compuestos modelo con el espectro diferencial de la proteína y se observa que la curva reconstruida coincide razonablemente con el espectro experimental, por lo que la suposición inicial de la aditividad de los espectros es válida.

El análisis de los resultados mostrados en la tabla 2-II indica que el número de residuos aromáticos superficiales determinados en este trabajo concuerda muy bien con los resultados obtenidos por Schmid et al. (1976), mientras que en el trabajo publicado por Svobodová et al. (1977) se reporta que la curva de titulación espectrofotométrica de los tirosinilos de la GPA es compatible con la presencia de tres de estos residuos con valores de pKa (el logaritmo negativo de la constante de ionización de su protón ácido) de 9.97 y nueve con pKa de 11.83. De ello se concluye en ese reporte que son sólo tres los residuos de tirosina expuestos. Sin embargo, existe la posibilidad de que algunos de estos residuos se encuentren externos pero formando puentes de hidrógeno suficientemente estables para aumentar el valor de sus pKa.

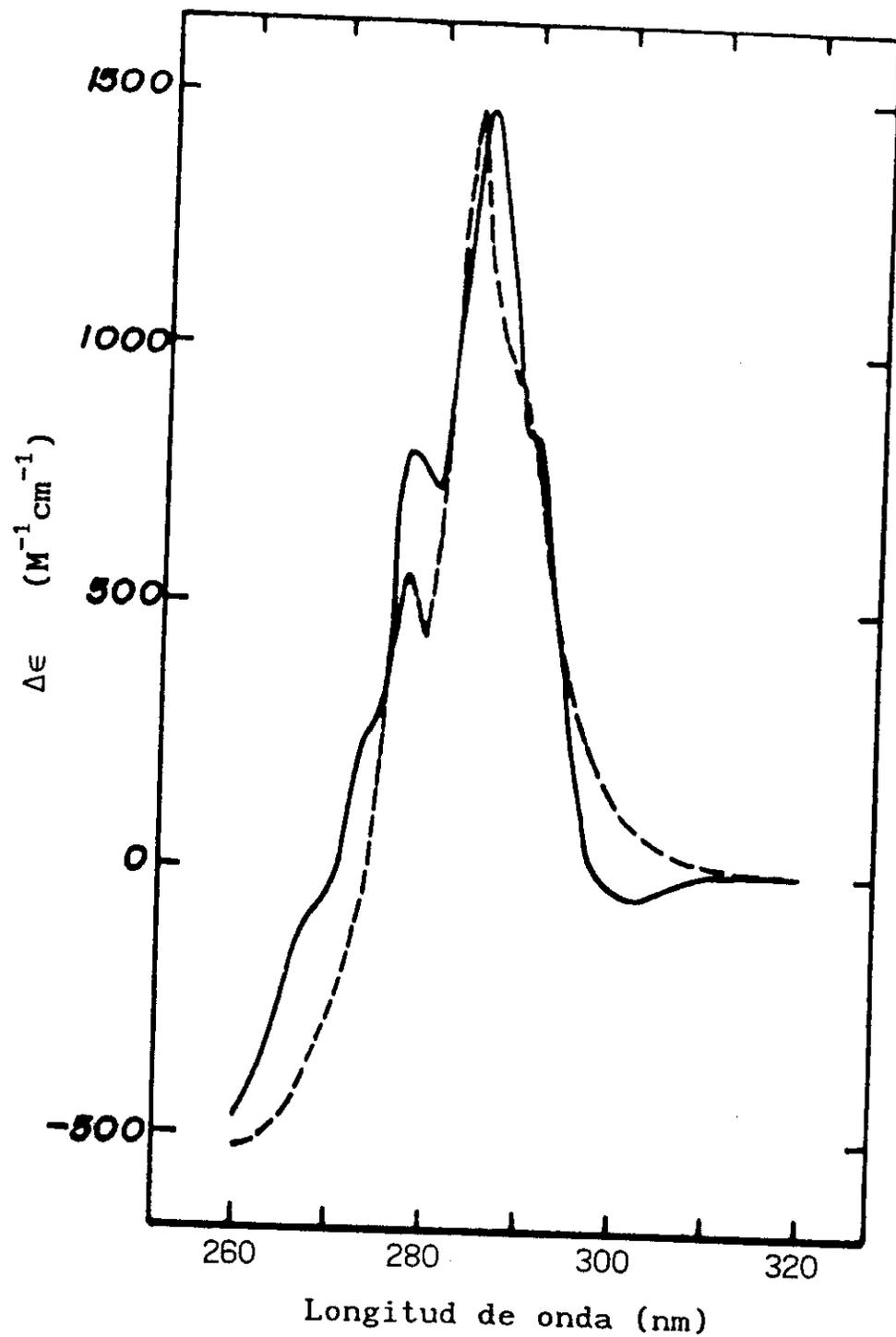


FIGURA 2-4. Espectro diferencial de la GPA en DMSO al 20% (v/v). Con línea punteada se representa la combinación lineal de los espectros diferenciales de los compuestos modelo.

Los estudios de reactividad química indican (por un proceso de fragmentación y secuenciación) que sólo el triptofano 122 es susceptible de modificarse parcialmente, en concordancia con el valor menor a la unidad para el número de triptofanos expuestos obtenido en este trabajo. Se especula en el trabajo de Schmid (1976) que este residuo de triptofano reactivo no se encuentra directamente localizado en el sitio de unión de ligandos de la proteína, pues al estudiar su reactividad en presencia de progesterona ésta no se altera (Kute y Westphal, 1976). Sin embargo, es posible que este residuo se encuentre cercano al sitio de unión, pues en estudios previos se observa una disminución del 17% de la fluorescencia intrínseca de la GPA al unir progesterona, estimándose una distancia de separación entre el triptofano y el esteroide de 9 a 14 Å (Kute y Westphal, 1976). Otros trabajos relacionados con la exposición de residuos aromáticos en la GPA reportan que la presencia de iones Cs^+ inhibe hasta una tercera parte de la fluorescencia intrínseca de la proteína (Friedman et al., 1985) y se observa que el fluoróforo inhibible por este ion se encuentra cercano a una zona de la cadena polipeptídica con carga negativa, la cual atrae electrostáticamente al catión. La presencia de drogas básicas, otras moléculas que la GPA une, disminuye la eficiencia del Cs^+ para inhibir la fluorescencia mas no la fracción inhibible, indicando que el residuo de triptofano es sensible a este ion, que es químicamente reactivo y que probablemente sea el mismo residuo aromático que es perturbado por el DMSO (pues el radio iónico del Cs^+ es de 1.67 Å y el radio promedio del perturbante de 2.0 Å), y se encuentre cercano al sitio de unión de drogas básicas. El ambiente de los residuos de triptofano se discute además en la sección de estudios de desnaturalización.

ESTIMACION DE ESTRUCTURAS SECUNDARIAS.

Los espectros de dicroísmo circular de la glicoproteína ácida α_1 y de la β -lactoglobulina se muestran en la figura 2-5. Puede observarse que ambos espectros tienen forma y magnitud semejantes. Este hecho indica que ambas proteínas comparten un patrón estructural, dado que esas curvas son el reflejo de sus estructuras secundarias (Hennessey y Johnson, 1982; Perczel et al., 1991; 1992). Por ejemplo, la presencia de un sólo mínimo en la región de 214 a 218 nm es característica de un contenido alto de hojas β (Manavalan y Johnson, 1983; Perczel et al., 1992), como es conocido para el caso de la BLG.

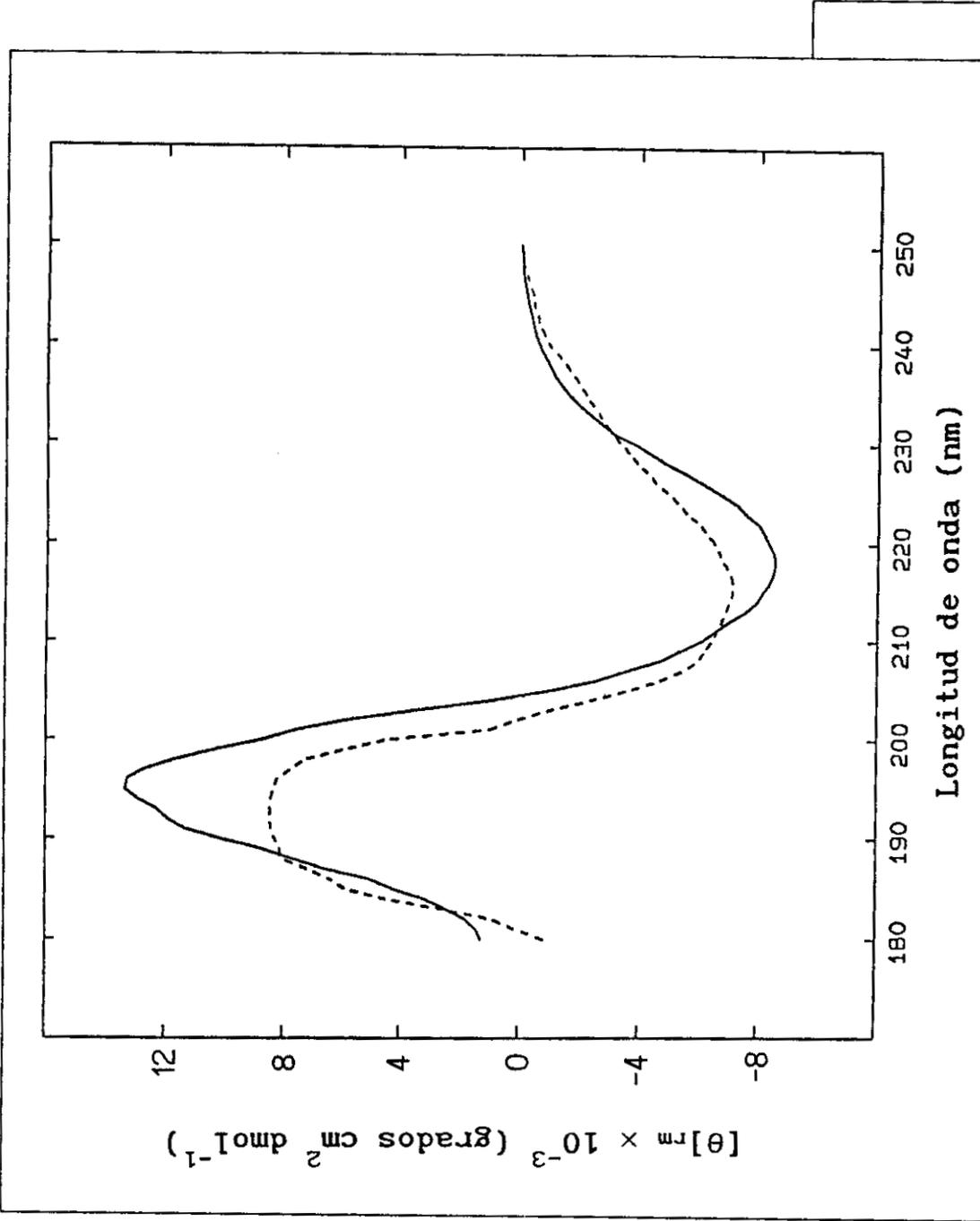


FIGURA 2-5. Espectros de dicroísmo circular de la glicoproteína ácida α_1 (—) y de la β -lactoglobulina (-----) en el ultravioleta lejano.

Sería de gran interés el comparar los espectros de dicroísmo circular de la figura 2-5 con los del resto de las lipocalinas cuya estructura cristalina es conocida; sin embargo, hasta donde sabemos, esta información no ha sido publicada, y la BLG fue la única lipocalina con esta característica a la cual tuvimos acceso para estudios de dicroísmo circular. El análisis de los espectros de esas proteínas permite estimar contenidos de hélices y hojas β de 16 y 33% respectivamente para la GPA y de 18 y 27% en el caso de la BLG. Aunque esta última cifra es menor que el contenido de hojas β observado en la estructura cristalográfica de la BLG, y que alcanza el 40%, los resultados del análisis identifican a las hojas β como el tipo más abundante de estructura secundaria presente en ambas muestras. Es conveniente señalar aquí que análisis previos del espectro de dicroísmo circular de la AGP han estimado el contenido de estas estructuras en valores tan bajos como 23% (Aubert y Loucheux-Lefebvre, 1976). Esta discrepancia puede deberse a que dentro del conjunto de espectros de dicroísmo circular que sirven de base de datos a los métodos de análisis, no se encuentra representada convenientemente la estructura de las lipocalinas. En realidad esto no es poco probable, pues la conformación de los miembros de esta familia no había sido observada antes de 1987, y en las estructuras cristalinas de proteínas con alta resolución (que son las que comúnmente se emplean en esas bases de datos) no se observa una gran cantidad de hojas β paralelas. Justamente, el método de estimación de estructuras secundarias empleado en este trabajo utiliza un mayor número de proteínas ricas en este tipo de estructuras en su conjunto inicial de datos, lo cual se refleja en una estimación más exacta del contenido de hojas β a partir de los espectros de la figura 2-5. De cualquier manera, el análisis tanto cualitativo como cuantitativo de los espectros de estas proteínas aporta evidencias consistentes con la hipótesis de que la parte proteica de la GPA posee un patrón de plegamiento semejante al de la BLG y, por ello, al de las lipocalinas.

Por otra parte, los espectros de dicroísmo circular de la GPA antes y después del proceso de deslipidación resultaron ser esencialmente idénticos, por lo que el ligando hidrofóbico parece no alterar significativamente la estructura secundaria de esta molécula. Este hecho es consistente con la observación previa de que las características fisicoquímicas de la GPA no se ven alteradas por la deslipidación, salvo la afinidad por otros ligandos. Además, permite

suponer que la presencia de progesterona, u otro ligando específico de la GPA, no altera sensiblemente su estructura. A propósito de este último punto, la presencia de las cinco cadenas de carbohidratos unidas a la parte proteica de la GPA tampoco parece ser determinante en la conformación de su cadena polipeptídica, a juzgar de resultados de dicroísmo circular (Aubert y Loucheux-Lefebvre, 1976) y de unión de progesterona (Westphal, 1975).

DESNATURALIZACION TERMICA.

Los cambios conformacionales de la GPA inducidos por el aumento de la temperatura se estudiaron por espectroscopía diferencial ultravioleta y por dicroísmo circular. Los cambios observados en ambos tipos de espectroscopías resultaron ser reversibles en más del 90% al enfriar lentamente las disoluciones de proteína. En la figura 2-6a se muestra un conjunto de espectros diferenciales a diferentes temperaturas. Los espectros correspondientes a temperaturas altas muestran tres bandas negativas, a 277, 284 y 291 nm que pueden asignarse a residuos de tirosina y triptofano que se exponen al solvente acuoso durante la desnaturalización (Donovan, 1969). Sin embargo, los espectros a temperaturas intermedias (curvas 1 y 2 de la figura 2-6a) muestran dos bandas positivas intensas centradas entre 278 y 295 nm. Estas presentan un signo opuesto al esperado por el cambio de ambiente de grupos aromáticos al pasar del medio hidrofóbico del interior de la macromolécula al solvente polar. Se han asignado estas bandas, poco observadas en espectros diferenciales de proteínas, a efectos electrostáticos que sufren los residuos de triptofano (Andrews y Forster, 1972; Anantharayanan y Bigelow, 1969a,b). Estos cambios corresponden a un aumento en la carga neta negativa del microambiente de la cadena lateral de un residuo de triptofano o bien a una disminución de carga positiva. Esta posibilidad no descarta una posible disminución de la polaridad del medio en el que se encuentra el triptofano como una explicación de la aparición de esta banda *anómala* (Andrews y Forster, 1972).

En la figura 2-7 se muestran las curvas de transición de la GPA determinadas por absorción diferencial a 295 y 284 nm. La característica más importante de esta figura es que el cambio de la banda positiva (295 nm) ocurre a temperaturas menores que la banda negativa. Esto sugiere que el proceso de desnaturalización no se ajusta a un modelo de dos estados, existiendo al menos

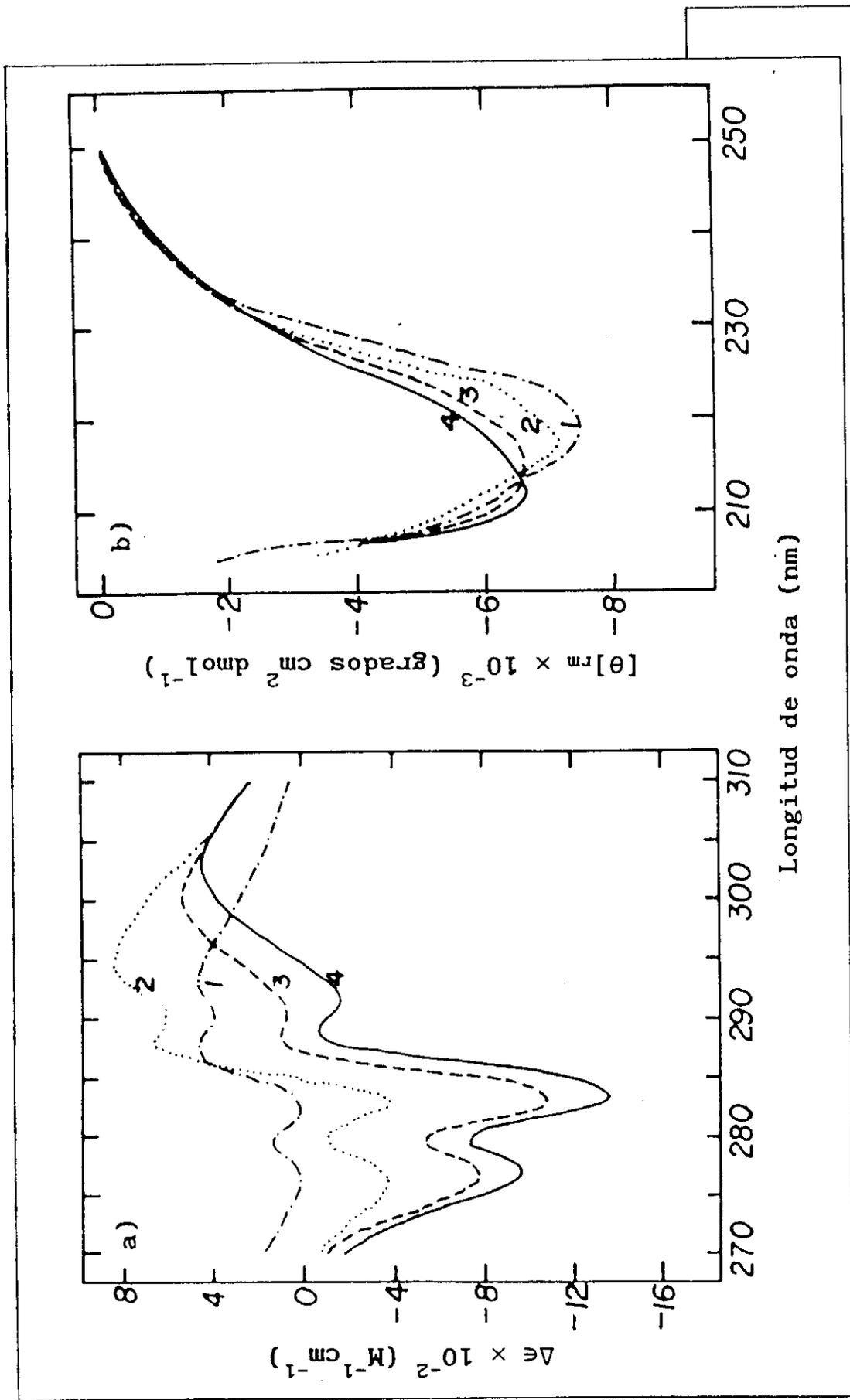


FIGURA 2-6. Desnaturalización térmica de la GPA. a) Espectros diferenciales de absorción ultravioleta a diferentes temperaturas: (1) 29.9, (2) 49.6, (3) 58.2 y (4) 67.0°C. b) Espectros de dicroísmo circular obtenidos a: (1) 16.8, (2) 44.6, (3) 49.5 y (4) 71.2°C.

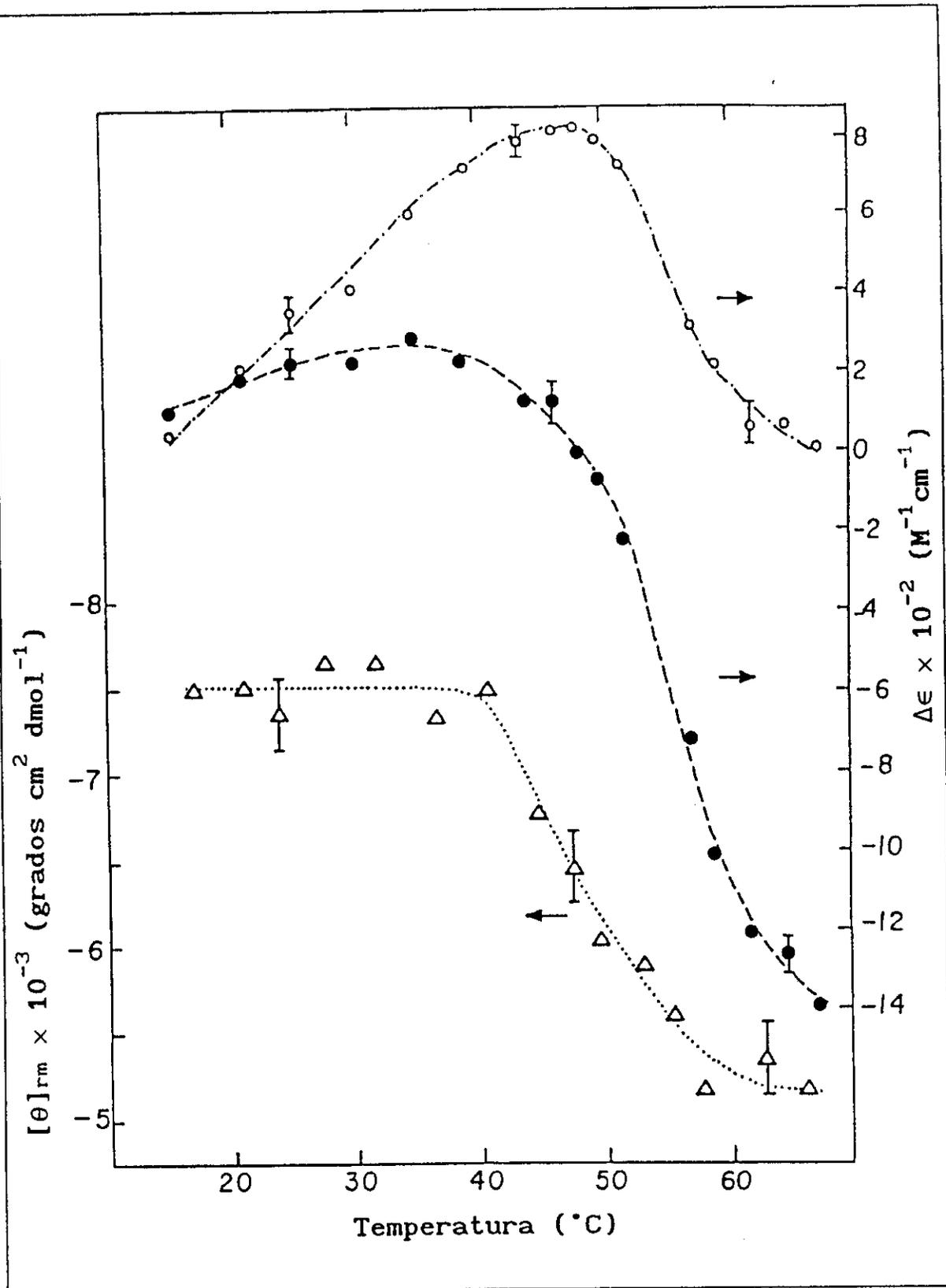
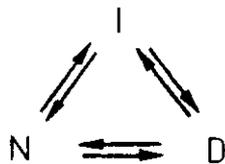


FIGURA 2-7. Efecto de la temperatura sobre las bandas de absorción diferencial a 295 nm (O) y 284 nm (●) y sobre la elipticidad a 220 nm (Δ). Las barras indican la variación experimental observada en diferentes puntos.

un intermediario termodinámicamente estable a concentraciones apreciables. Un mecanismo probable para este proceso podría ser el siguiente:



donde I, N y D representan a las especies intermediaria, nativa y desnaturalizada, respectivamente. Este mecanismo propuesto concuerda con los resultados calorimétricos realizados por Halsall y Kirley (1982), quienes calcularon que el cociente de las entalpías de van't Hoff y calorimétrica es de aproximadamente 1.22, indicando también que el modelo de dos estados no es adecuado para describir la transición conformacional de la GPA.

Al aplicar el método de análisis de factores en un conjunto de 16 espectros diferenciales correspondientes a un intervalo de temperaturas de 15.3 a 67.0°C, se encuentra que para reproducir los datos experimentales son necesarios dos vectores propios. Este resultado muestra la presencia de una especie intermediaria en la desnaturalización térmica de la GPA, ya que si este proceso fuera de dos estados, sería suficiente un factor para describir los espectros diferenciales, dado que el espectro de la especie nativa es un vector nulo. Los dos factores que describen adecuadamente al conjunto de espectros diferenciales no tienen una interpretación física. Simplemente son los dos vectores ortogonales con valores propios más grandes, resultado de la diagonalización del producto matricial $[D]^T[D]$. Para encontrar un par de vectores con significado físico se consideró el espectro diferencial de la GPA a 67°C (la curva 4 de la figura 2-6a) como el correspondiente al estado desnaturalizado. Con este único vector se intentó describir al conjunto completo de espectros experimentales, resultando necesariamente insuficiente. Sin embargo, cuando las diferencias entre los espectros reconstruidos con el vector *desnaturalizado* y los espectros experimentales fueron sometidas a un nuevo análisis de factores, se obtuvo un segundo vector. Al combinar linealmente ambos factores se logró nuevamente describir al conjunto de espectros diferenciales, dentro del error experimental. La forma de ese segundo vector es muy parecida a la de la curva 2 de la figura 2-6a; este espectro diferencial, que posiblemente corresponda (salvo por un factor

multiplicativo) al de la especie intermediaria, muestra dos máximos a 287 y 295 nm. Estas bandas positivas, como se mencionó en la sección anterior, pueden ser debidas a la perturbación de los grupos indol de residuos de triptofano causadas por el solvente o por cargas eléctricas. Esto sugiere que la transición de la conformación nativa a la intermediaria involucra cambios en el microambiente de alguno o algunos de los tres triptofanilos de la GPA, pero sin cambiar significativamente el grado de exposición de sus grupos aromáticos.

En la figura 2-6b se muestra el cambio de los espectros de dicroísmo circular con la temperatura; en la región mostrada, 200 a 250 nm, el cambio en las señales es principalmente debido a la modificación de la estructura secundaria de la glicoproteína, pues ha sido demostrado que la parte de carbohidratos de esta molécula no contribuye sensiblemente a la actividad óptica total en longitudes de onda por arriba de 190 nm (Aubert y Loucheux-Lefebvre, 1976). En este conjunto de espectros (figura 2-6b) puede apreciarse un punto isodicroico, es decir, la existencia de una longitud de onda en la que la señal no se ve afectada por el cambio conformacional. Esta característica es indicativa de que sólo existen dos especies espectroscópicamente diferentes, y que la combinación lineal de sus espectros reproducen el conjunto completo de curvas. Esta interpretación es apoyada por el hecho de que, al aplicar el método de análisis de factores a un conjunto de 19 espectros (en el intervalo de temperaturas de 16.8 a 71.2°C), se encuentra que sólo se requieren dos vectores propios para reconstruir todas las curvas experimentales. La curva de transición (figura 2-8) seguida por la señal de elipticidad a 220 nm coincide aproximadamente con el cambio en la banda diferencial a 284 nm. Estos resultados son consistentes con que el intermediario propuesto en el mecanismo de desnaturalización posea una conformación muy semejante a la nativa (sin gran exposición al solvente de sus grupos hidrofóbicos, y con poco cambio en el contenido original de estructuras secundarias), con cambios estructurales restringidos a una porción pequeña de su geometría.

La presencia de un intermediario, como es el caso para la GPA, impide la aplicación directa de un análisis de van't Hoff para calcular los parámetros termodinámicos de la desnaturalización de una macromolécula, es decir los cambios de entalpía y capacidad calorífica al pasar del estado nativo al desnaturalizado. Sin embargo, puede encontrarse un cambio de entalpía

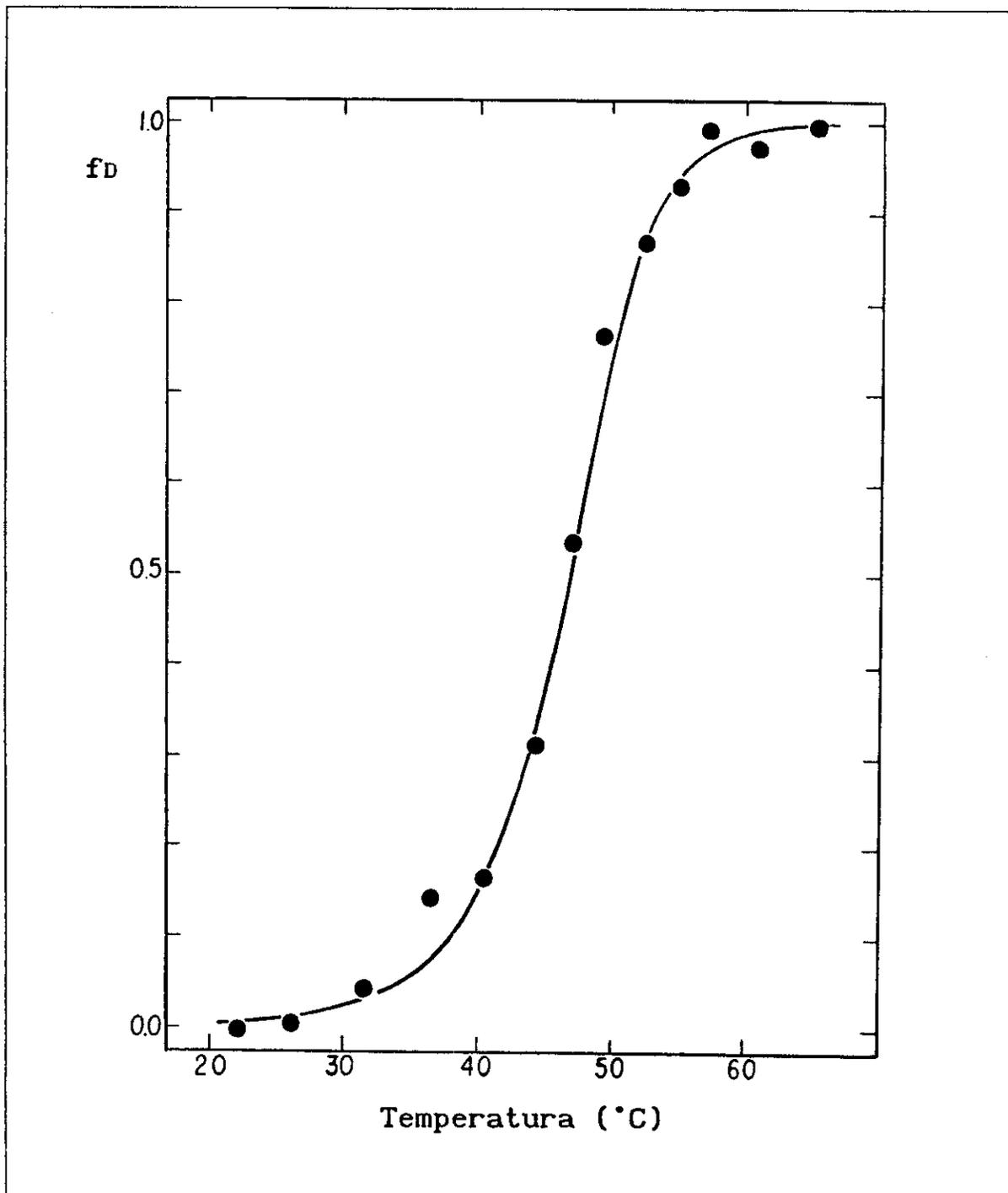


FIGURA 2-8. Curva de transición de la desnaturalización térmica de la GPA. La línea representa a la ecuación 2-13 ajustada mediante una regresión no lineal. Los puntos corresponden a valores calculados de los datos de elipticidad a 220 nm de la figura 2-7.

aparente, ΔH_{vH} , que se encuentra relacionado de una manera compleja con el cambio de entalpía verdadero, ΔH_D (Tanford, 1969; Cantor y Schimmel, 1980). En prácticamente todos los casos en que está involucrado un intermediario, se encuentra que $\Delta H_{vH} < \Delta H_D$. Para nuestro caso particular, en donde participan tres especies (la nativa, N, el intermediario, I, y la desnaturalizada, D) y dos de ellas son indistinguibles por la técnica empleada (N e I), puede calcularse una constante de equilibrio aparente bajo la suposición de dos estados:

$$K_{vH} = \frac{f_D}{(1 - f_D)} = \frac{f_D}{(f_N + f_I)}$$

donde f_D , f_I y f_N representan las fracciones de proteína en cada uno de los tres estados. Al dividir por f_N el numerador y denominador de la ecuación anterior, sustituyendo además las constantes de equilibrio $K_D = f_D/f_N$ y $K_I = f_I/f_N$, se obtiene la siguiente expresión:

$$K_{vH} = \frac{K_D}{(1 + K_I)}$$

A partir de esta expresión y utilizando la relación:

$$\frac{\partial(\ln K)}{\partial T} = \frac{\Delta H}{RT^2}$$

puede demostrarse que el cambio de entalpía aparente puede calcularse de la siguiente manera,

$$\Delta H_{vH} = \Delta H_D - \Delta H_I \cdot \left[\frac{K_I}{(1 + K_I)} \right]$$

en donde los símbolos ΔH_D y ΔH_I representan los cambios de entalpía para los cambios conformacionales de N a D y de N a I, respectivamente.

Por otra parte, dado que $\Delta C_P = \partial\Delta H/\partial T$, al derivar la última expresión respecto a la temperatura y reorganizar, se obtiene

$$\Delta C_{P_{vH}} = \Delta C_{P_D} - \Delta C_{P_I} \cdot \left[\frac{K_I}{(1 + K_I)} \right] - \frac{\Delta H_I^2}{RT^2} \cdot \left[\frac{K_I}{(1 + K_I)^2} \right]$$

en la que los subíndices tienen el mismo significado que en las ecuaciones anteriores. Es muy importante hacer notar que el valor de los tres cocientes, que se encuentran entre paréntesis cuadrados en las dos últimas ecuaciones, es necesariamente menor que la unidad y mayor que cero. Por esta razón, y si los valores numéricos de ΔH_I y ΔC_{P_I} son positivos, se concluye que $\Delta H_{vH} < \Delta H_D$ y que $\Delta C_{P_{vH}} < \Delta C_{P_D}$. Sin embargo, como se estableció en los párrafos precedentes, la naturaleza de la especie intermediaria no difiere sustancialmente de la de la nativa, por lo que la suposición de que los cambios en entalpía y capacidad calorífica asociados con la transición de N a D sean pequeños, es razonable. Esta suposición conduce a que los valores aparentes y reales de la desnaturalización sean aproximadamente iguales; esto es, que $\Delta H_{vH} \approx \Delta H_D$ y que $\Delta C_{P_{vH}} \approx \Delta C_{P_D}$.

Una vez establecido que las magnitudes de los parámetros termodinámicos asociados con la desnaturalización de la GPA pueden aproximarse a los calculados empleando el mecanismo de dos estados, se utilizó el cambio de elipticidad de la solución de glicoproteína, mostrado en la figura 2-7, para calcular las fracciones de proteína desnaturalizada a diferentes temperaturas empleando las ecuaciones (2-1) y (2-2). Los valores calculados se muestran en la figura 2-8, y fueron ajustados por medio de una regresión no lineal (ver apéndice I) a la ecuación 2-13 para estimar la temperatura media de la transición, T_m , y los cambios de entalpía y capacidad calorífica aparentes. Los resultados obtenidos se muestran en la tabla 2-III.

T A B L A 2 - I I I		
Parámetros termodinámicos de la desnaturalización térmica de la GPA.		
T_m (°C)	ΔH_{vH}^a (kJ·mol ⁻¹)	$\Delta C_{P_{vH}}$ (kJ·mol ⁻¹ ·K ⁻¹)
46.6 ± 0.05	252 ± 25	8.8 ± 4.2
^a Valor evaluado a la temperatura media de la transición, T_m .		

Para comparar los valores de esta tabla con los de otros trabajos, se realizó una extrapolación del valor del ΔH_{vH} a 62°C , pues Halsall y Kirley (1982) midieron directamente el cambio de entalpía a esta temperatura en un microcalorímetro, en condiciones en las que la polimerización de la GPA no alteraba notablemente sus resultados ($\text{pH} = 7.4$, $T_m = 62^\circ\text{C}$) obteniendo para el calor absorbido en el proceso, $\Delta H_{cal} = 493 \pm 17 \text{ kJ}\cdot\text{mol}^{-1}$, en buen acuerdo con nuestro resultado de la extrapolación, $\Delta H_{vH}(62^\circ\text{C}) = 401 \pm 13 \text{ kJ}\cdot\text{mol}^{-1}$. Para la comparación del cambio en la capacidad calorífica, no tenemos noticia de que se haya determinado este valor directamente por calorimetría. El único valor reportado se encuentra en un trabajo de la unión de propranolol (una droga básica que se une a la GPA) a la glicoproteína en donde se estima el valor del ΔC_{pD} en $6.7 \pm 6.3 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{K}^{-1}$ (Kirley *et al.*, 1982). A pesar de la gran incertidumbre en ese valor, un $\Delta C_{p_{vH}}$ de $8.8 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{K}^{-1}$ (obtenido de nuestros experimentos) parece ser un valor razonable.

DESNATURALIZACION INDUCIDA POR HIDROCLORURO DE GUANIDINA.

En la figura 2-9a se muestran los espectros obtenidos por espectroscopía diferencial de la GPA a varias concentraciones de Gu a un valor de pH de 5.0. Se demostró la reversibilidad del proceso en alrededor de un 90% al diluir la concentración del desnaturizante a la tercera parte. Puede observarse de los espectros que estos presentan, al igual que los obtenidos en la desnaturalización térmica, dos bandas negativas centradas en 285 y 292 nm, pero en contraste, no se observa la presencia de una banda positiva (banda *anómala*) a ninguna concentración del desnaturizante. Otra diferencia con los estudios de estabilidad térmica, es que los cambios de absorptividad molar ($\Delta\epsilon$) inducidos por la guanidina fueron mayores (figuras 2-6a y 2-9a). Esto puede interpretarse como que el efecto de la desnaturalización por guanidina induce una mayor exposición al solvente de los residuos aromáticos, respecto a la desnaturalización térmica. Al estudiar la estabilidad de la GPA frente a la guanidina por dicroísmo circular, se encontró que el proceso es completamente reversible. Al comparar los espectros de dicroísmo circular de la proteína totalmente desnaturizada de las figuras 2-6b y 2-9b se observa que las señales son aún menores en el segundo caso. Esto también ha sido observado en los estudios de desnaturalización de otras proteínas globulares (Goto y Hamaguchi, 1987; Labhardt, 1982), y se ha interpretado con base a la posible

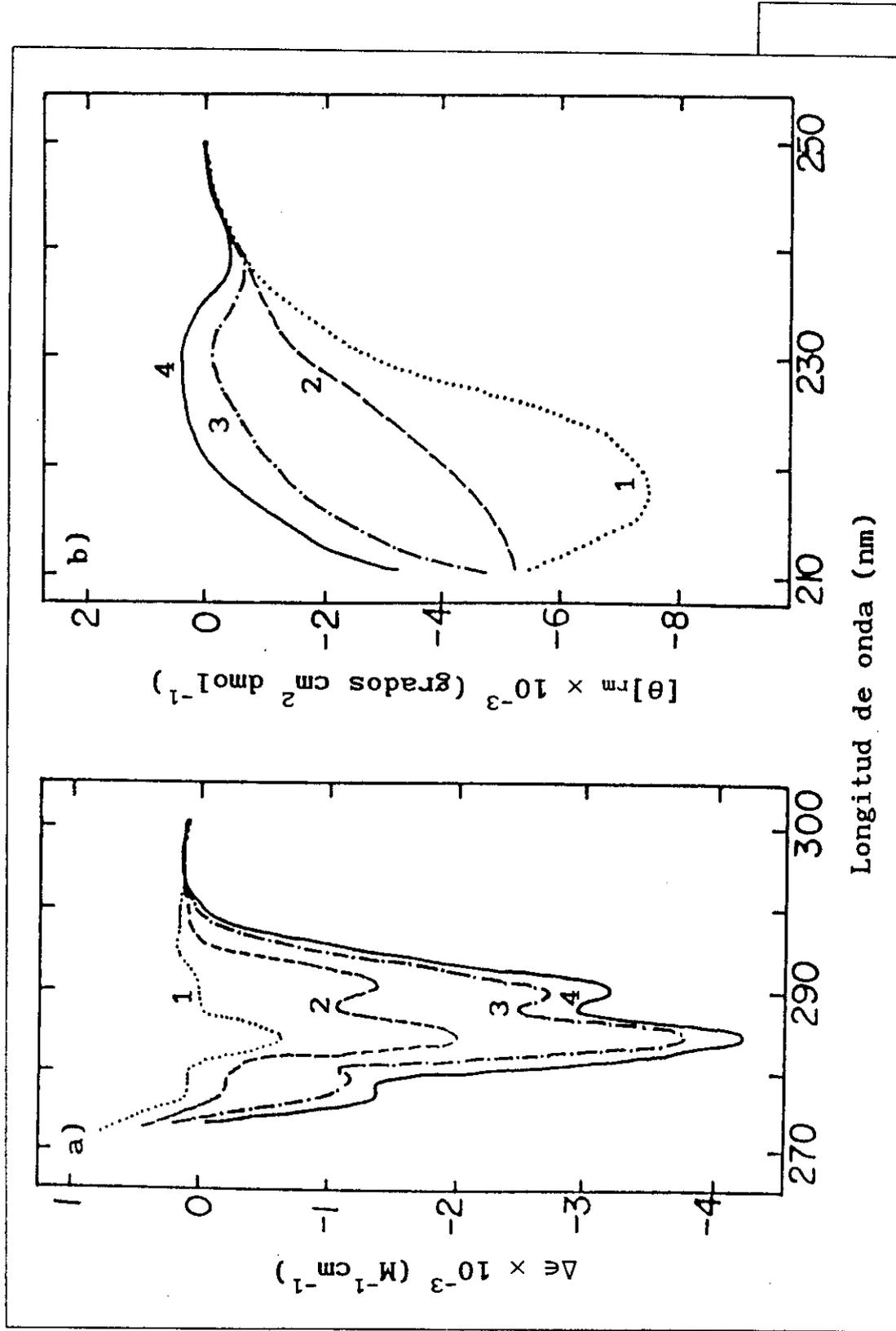


FIGURA 2-9. Desnaturalización de la GPA inducida por hidrocloreuro de guanidina. a) Espectros diferenciales de absorción ultravioleta a concentraciones de Gu de: (1) 1.9, (2) 2.3, (3) 2.8 y (4) 5.0 M. b) Espectros de dicroísmo circular obtenidos a concentraciones de Gu de (1) 0.0, (2) 2.0, (3) 4.0 y (4) 7.0 M.

existencia de estructuras residuales presentes en las especies desnaturalizadas térmicamente (Ptitsyn, 1987; Baldwin, 1986). En la figura 2-10 se muestran las curvas de transición obtenidas por espectroscopía diferencial a 285 nm y por el cambio de elipticidad a 218 nm, en términos de la fracción de proteína desnaturalizada (ecuación 2-2). La suposición de que este proceso ocurre a través de un mecanismo de dos estados se justifica por la coincidencia de las curvas de transición estudiadas con las dos técnicas diferentes (Ghèlis y Yon, 1982). La validez de esta aproximación es respaldada por el hecho de que si se representan las mismas curvas con el cambio de la señal, por cualquiera de los dos métodos experimentales, a otras longitudes de onda, se encuentra nuevamente coincidencia con las ya mencionadas (estos datos no se muestran). Para calcular los cambios de energía libre de la transición se utilizaron los valores de f_D de la figura 2-10 en las ecuaciones 2-3 y 2-4. Las estimaciones de ΔG_D a diferentes concentraciones de guanidina (figura 2-11) fueron ajustadas por el método de mínimos cuadrados a los modelos de *dependencia lineal* (ecuación 2-16) y de *Tanford* (ecuación 2-17) para extrapolar el valor del cambio de energía libre en ausencia de desnaturalizante, $\Delta G_D^{H_2O}$, y estimar los parámetros correspondientes a cada modelo. En la tabla 2-IV se muestran los resultados obtenidos de estos cálculos.

T A B L A 2 - I V
Parámetros de la desnaturalización de GPA inducida por Gu (pH 5.0, 20°C)

Método	Parámetro	[Gu] ^a _{1/2} (M)	$\Delta G_D^{H_2O}$ (kJ·mol ⁻¹)
Extrapolación lineal ^b	$m = 10.2 \text{ kJ}\cdot\text{mol}^{-1}\text{M}^{-1}$	2.1	21.3
Modelo de Tanford ^c	$\alpha = 0.12$	2.1	28.9

^a Concentración de desnaturalizante en el punto medio de la transición.
^b Calculado de la ecuación 2-16. ^c Calculado de la ecuación 2-17.

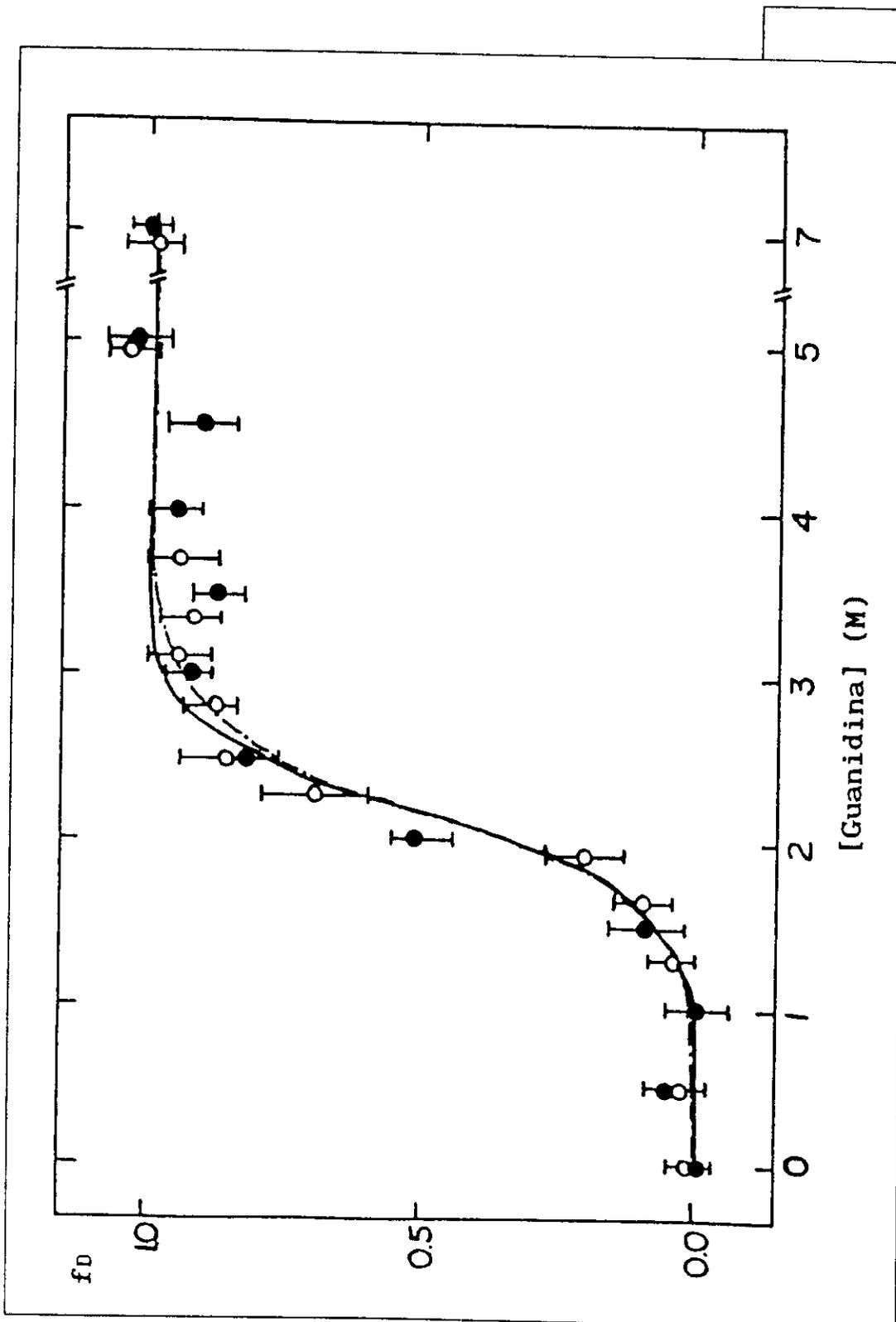


FIGURA 2-10. Curvas de transición de la desnaturalización de la GPA inducida por Gu. Las líneas continua y discontinua representan respectivamente al modelo de extrapolación lineal y de Tanford. Los datos experimentales fueron obtenidos por espectroscopia diferencial (●) y dicroísmo circular (○). Las barras muestran la variación experimental.

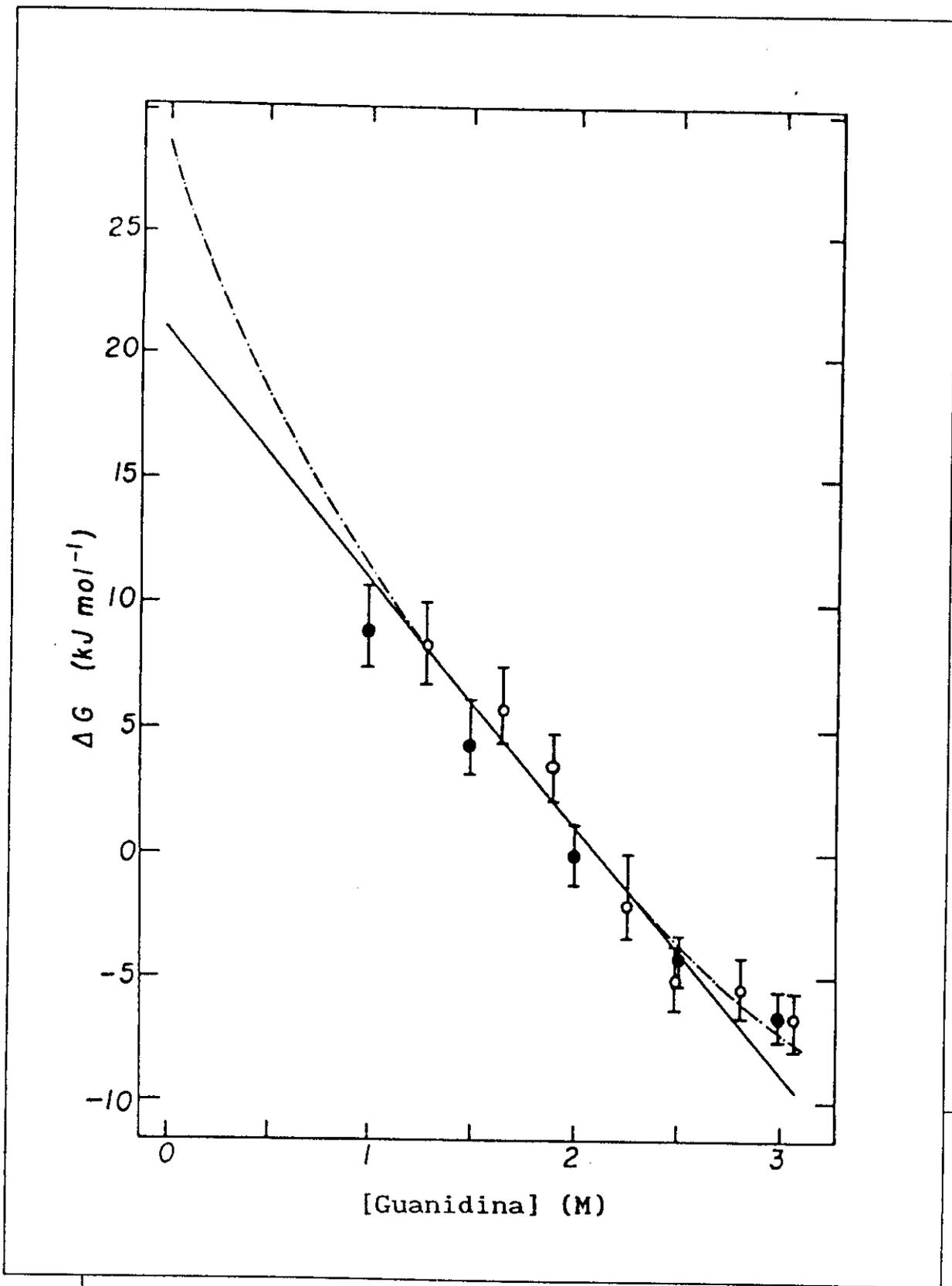


FIGURA 2-11. Métodos de extrapolación para estimar el cambio de energía libre de desnaturalización en ausencia del desnaturalizante. Los símbolos son idénticos a los de la figura 2-10.

CONCLUSIONES DE LOS RESULTADOS EXPERIMENTALES.

Los resultados obtenidos de los experimentos de perturbación por solvente son satisfactorios, a juzgar por la coincidencia con los valores estimados de estudios de reactividad química. De esta manera, la molécula de GPA parece tener un residuo de triptofano parcialmente expuesto y dos ocultos; seis residuos de tirosina expuestos, aproximadamente dos parcialmente accesibles al solvente y el resto en el interior de la molécula. Los estudios de Schmid *et al.* (1976) parecen señalar al triptofano 122 como el que se encuentra en la superficie molecular. En otros trabajos, se estima la distancia entre la progesterona unida a la GPA y una de las cadenas laterales de triptofano de esta proteína, entre 9 y 14 Å (Kute y Westphal, 1976); mientras que al menos uno de los triptofanos parece encontrarse en un medio con carga neta negativa (Friedman *et al.*, 1985). La posibilidad de que esta información corresponda a un mismo triptofano, o bien que sea producida por varios de ellos, se discute en el análisis del modelo de la estructura de la GPA.

La banda positiva que se observa en los espectros diferenciales de la desnaturalización térmica de la GPA, también indica la posibilidad de cambios en el entorno de una cadena lateral de triptofano, en el cual parece haber una disminución de polaridad o un efecto electrostático. La GPA tiene tres residuos de triptofano en las posiciones 25, 122 y 160 de su secuencia (Schmid *et al.*, 1973; Dente *et al.*, 1987). El triptofano 160 se encuentra en una región de la secuencia que contiene un número abundante de cadenas laterales con carga eléctrica, incluyendo un residuo de ácido aspártico y dos de lisina en las posiciones 159, 161 y 162, respectivamente. Esto sugiere que un cambio conformacional en esta región podría provocar cambios en el entorno del triptofano 160.

Por otra parte, si las bandas *anómalas* se debieran a una disminución de la polaridad del medio, el triptofano responsable debería encontrarse inicialmente expuesto al solvente acuoso. Según Schmid *et al.* (1976), el único residuo de triptofano reactivo frente al bromuro de 2-hidroxi-5-nitrobencilo, es el 122. La falta de reactividad de los otros triptofanos es interpretada por Schmid *et al.* como indicativa de que esas cadenas laterales se encuentran en el interior de la molécula.

El perfil de hidrofobicidad de la GPA (figura 2-12), construido de acuerdo al método de Kyte y Doolittle (1982), indica que los residuos 122 y 160 se encuentran en regiones de la secuencia de aminoácidos con una hidrofiliidad mayor que la promedio. En contraste, el triptofano 25 está ubicado en una zona hidrofóbica; esto sugiere que, independientemente del mecanismo por el cual se genera la banda positiva de los espectros diferenciales, la transición del estado nativo al intermediario involucra a los residuos 122 o 160.

La comparación de los espectros de dicroísmo circular de la GPA y de la BLG muestra que ambas macromoléculas poseen aproximadamente el mismo contenido de estructuras secundarias, donde las hojas β antiparalelas son la estructura repetitiva más abundante. Esto apoya la idea de que la glicoproteína posee una conformación semejante a la de las lipocalinas. Por otra parte, existe evidencia para suponer que la presencia de un ligando específico, o de las cadenas de carbohidratos que la GPA posee, no afectan de manera notable la conformación de su porción proteica.

En los estudios de estabilidad térmica y frente al hidrocioruro de guanidina de la glicoproteína ácida α_1 , se demostró que en las condiciones estudiadas la reversibilidad del proceso es casi total. Aunque el análisis de los datos experimentales indica que durante la desnaturalización térmica de la GPA existe un intermediario termodinámicamente estable en concentraciones apreciables, este intermediario parece tener una conformación semejante a la nativa. La especie intermediaria tiene un contenido de estructuras secundarias, y un núcleo de cadenas laterales hidrofóbicas, similar a las de la especie nativa, aunque con cambios locales en el entorno de una cadena lateral de triptofano. La presencia de este intermediario es consistente con las observaciones calorimétricas reportadas por Halsall y Kirley (1982), quienes proponen dos explicaciones posibles para sus resultados: la variabilidad de la secuencia de aminoácidos de la GPA, o bien que esta glicoproteína posee una región (que es fácilmente perturbada por la presencia de etanol en el medio) cuyo cambio conformacional casi no afecta la endoterma calorimétrica. Esta zona podría ser aquella en la que se encuentra el residuo de triptofano responsable de las bandas diferenciales positivas. Es así que, aunque la desnaturalización térmica de la GPA no se ajuste a un mecanismo de dos estados, los parámetros termodinámicos calculados de un análisis de van't Hoff se aproximan a los valores reales de la transición. Por esto, y pese a

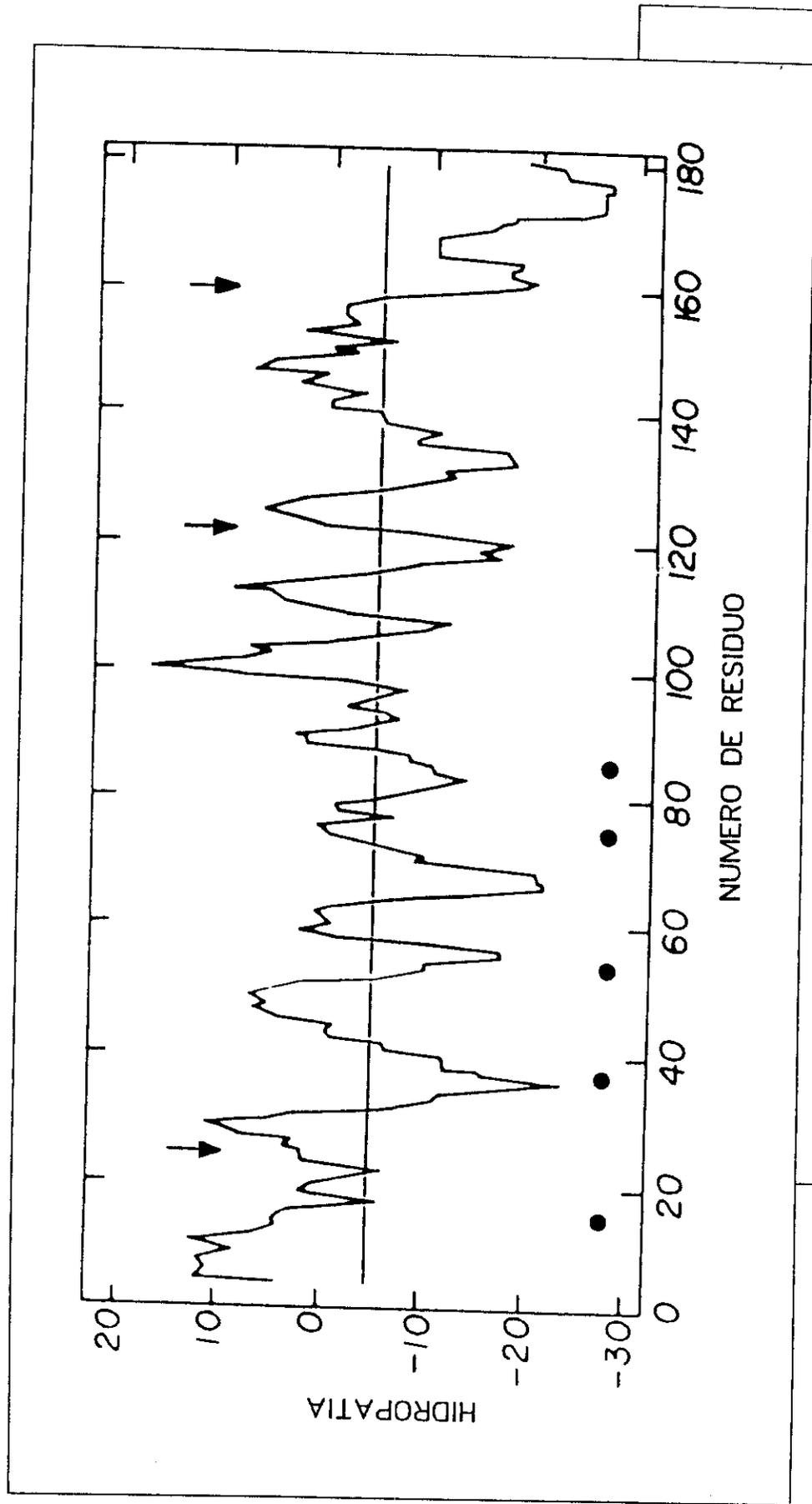


FIGURA 2-12. Perfil de hidrofobicidad de la secuencia de la GPA. Los máximos corresponden a regiones de hidrofobicidad relativa y los mínimos a zonas hidrofílicas. Las flechas señalan las posiciones de los residuos de triptofano. Los puntos muestran a los residuos de asparragina donde se unen las cadenas de carbohidratos.

las incertidumbres de gran magnitud inherentes a este tipo de análisis, el valor del $\Delta C_{p,D} = 8.8 \pm 4.2 \text{ kJ}\cdot\text{mol}^{-1}\text{K}^{-1}$ puede emplearse para estimar los valores de los parámetros termodinámicos a otras temperaturas. De acuerdo a las ecuaciones 2-14 y 2-15 y utilizando el valor del ΔH_D calorimétrico ($493 \text{ kJ}\cdot\text{mol}^{-1}$ a 62°C), la temperatura de máxima estabilidad y el valor del ΔG_D a esta temperatura son de 8.5°C y $21.8 \text{ kJ}\cdot\text{mol}^{-1}$, respectivamente.

Al extrapolar a 110°C el cambio de entalpía del proceso, se obtiene un valor de $915 \pm 200 \text{ kJ}\cdot\text{mol}^{-1}$. El cambio de entalpía específico, considerando sólo la parte proteica de la GPA (21.5 kDa), resulta ser de $38 \pm 8 \text{ J}\cdot\text{g}^{-1}$. Este valor es menor que $54 \text{ J}\cdot\text{g}^{-1}$, el cual es característico de proteínas globulares cuya estructura es compacta, y es similar al de proteínas con conformaciones poco empacadas (Privalov, 1979).

Por otra parte, en la desnaturalización de la GPA inducida por Gu no se encuentra evidencia de intermediarios. Con la información obtenida de estos experimentos no es posible concluir si el intermediario observado en la desnaturalización térmica es inestable en presencia del desnaturalizante, o simplemente no es posible detectarlo bajo las condiciones experimentales. Sin embargo, la transición conformacional parece ser tan abrupta como la de otras proteínas, al menos en lo que respecta a la pendiente de la ecuación 2-16, m , que es una medida de la cooperatividad del proceso (Pace, 1975). Los valores reportados para este parámetro se encuentran en el intervalo de 7.9 a $16.2 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{M}^{-1}$ (Green y Pace, 1974), en concordancia con $10.2 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{M}^{-1}$ para el caso de la GPA. No obstante, la extrapolación del cambio de energía libre, $\Delta G_D^{\text{H}_2\text{O}}$, para este proceso fue de alrededor de $25 \text{ kJ}\cdot\text{mol}^{-1}$, valor apenas superior al límite inferior de la estabilidad para proteínas globulares en condiciones fisiológicas (Privalov, 1979; Pace 1975).

Los resultados del análisis de Tanford (ecuación 2-17) de la desnaturalización de la GPA producida por Gu, arrojan un valor de α de 0.12 . Este valor representa el cambio en el grado de exposición al solvente de los grupos químicos de la macromoléculas. Las magnitudes observadas en estudios de proteínas típicas se encuentran entre 0.17 y 0.35 . En el caso de la GPA, el valor tan pequeño observado para α podría deberse a que el modelo de dos estados no describe adecuadamente la transición, como se ha mostrado en el caso de la desnaturalización de la cadena ligera de una inmunoglobulina tipo κ

(Rowe y Tanford, 1973). Otra probable explicación estaría relacionada con que el cambio en la accesibilidad de los grupos químicos de la GPA debido a su desnaturalización fuera pequeño. Esto significaría que, en términos relativos con otras proteínas, la GPA tuviera una mayor exposición al solvente de su estructura molecular nativa, o bien que la conformación desnaturalizada tuviera ocultos al solvente un número mayor de grupos químicos.

La magnitud de las bandas en los espectros diferenciales a altas concentraciones de guanidina, pueden indicar aproximadamente cuantos residuos de tirosina y triptofano pasan del medio hidrofóbico, del interior de la proteína, al solvente acuoso durante la desnaturalización; suponiendo que la exposición de estos residuos causarían diferencias de absorción molar de -700 y $-130 \text{ M}^{-1}\text{cm}^{-1}$ a 285 y 292 nm , respectivamente, para el caso de la tirosina, mientras que para el triptofano fueran de -1190 y $-1690 \text{ M}^{-1}\text{cm}^{-1}$ (nuevamente a 285 y 292 nm) (Donovan, 1969; Herskovits y Sorensen, 1968), el espectro diferencial de la GPA sería producido por la exposición de aproximadamente dos triptofanos y tres tirosinilos. Esto implica que, en lo que respecta a estas cadenas laterales aromáticas, la exposición es prácticamente total, pues se mencionó anteriormente que la estructura nativa de la GPA tiene dos triptofanos y aproximadamente tres tirosinas internas.

En contraste, la interpretación de que el valor pequeño del parámetro α (ecuación 2-17) es debido a que la GPA nativa tiene una fracción de grupos químicos expuestos al solvente mayor que otras proteínas, es apoyada por la gran hidrofiliidad del fragmento carboxilo terminal generado por bromuro de cianógeno (Ikenaka *et al.*, 1972). En la figura 2-12 se puede observar que los últimos aminoácidos en la secuencia de la glicoproteína poseen alta hidrofiliidad.

Puede describirse la estructura nativa de la GPA como una molécula con una región poco empacada, es decir, con un plegamiento menos compacto que el del resto de la cadena. Esto podría explicar simultáneamente la presencia de la banda *anómala* en la desnaturalización térmica y la aparente gran accesibilidad del solvente a la estructura nativa. Esta zona de plegamiento *laxo* podría estar localizada en el segmento del aminoácido 122 al extremo carboxilo terminal, en donde no se encuentran cadenas de carbohidratos unidos a la parte proteica de la GPA y la polaridad de las cadenas laterales es muy alta.

CAPITULO 3.

ANALISIS TEORICO DE LA SECUENCIA DE AMINOACIDOS.

METODOS.

En el presente capítulo se ha preferido emplear en algunas ocasiones las abreviaturas de una sola letra para los aminoácidos (Apéndice II); esto debido a la frecuencia y cantidad de referencias a diferentes residuos, necesaria para la descripción de los resultados que se exponen. Para la traducción de codones, se empleó el código genético universal (Jukes, 1978).

Análisis teórico de la composición de aminoácidos. La predicción de la clase estructural a partir de la composición de aminoácidos (Nakashima *et al.*, 1986), emplea los resultados del análisis de 135 proteínas con estructura tridimensional conocida. La composición de estas proteínas se expresó en un espacio de dimensión 20, en donde cada eje ortogonal corresponde a un aminoácido. Los autores del método examinaron el patrón de distribución de las proteínas en este espacio respecto a las cinco clases estructurales: α , β , α/β , $\alpha+\beta$ e irregular, según la clasificación propuesta por Levitt y Chothia (1976). Encontraron que los porcentajes de aminoácidos constituyentes de las proteínas de las clases α , β y α/β se localizan en diferentes regiones del espacio multidimensional, lo cual permite la clasificación estructural de una proteína a partir de la composición de aminoácidos. Sin embargo, los puntos que representan a las proteínas con estructuras de los tipos $\alpha+\beta$ e irregulares, se encuentran en regiones ampliamente dispersas en el espacio multidimensional, y se traslapan con las zonas correspondientes a otras clases estructurales (figura 3-1a, página 50). Nakashima *et al.* desarrollaron entonces un método sencillo para predecir el tipo de doblado de una proteína con una exactitud del 70%. Para cada aminoácido (i) de una proteína (j) se calcula una composición normalizada $a(i,j)$:

$$a(i,j) = \frac{c(i,j) - \text{prom}(i)}{\sigma(i)} \quad (3-1)$$

donde $c(i,j)$ es el porcentaje del aminoácido (i) en la proteína (j); $\text{prom}(i)$ es la composición promedio del aminoácido (i) en la base de 135 proteínas y $\sigma(i)$ la desviación estándar de este último valor. Empleando las composiciones normalizadas, el método calcula las distancias entre pares de puntos del espacio de composiciones utilizando la siguiente expresión

$$d(j,k) = \sqrt{\sum_{i=1}^{20} (a(i,j) - a(i,k))^2} \quad (3-2)$$

en la que los índices j y k identifican cada uno de los puntos; éstos pertenecen a la composición de la proteína en estudio (j) y al origen de los ejes o a los centros de las regiones de las diferentes clases estructurales (k). Si la distancia al origen del punto correspondiente a la proteína en estudio es mayor a 11 unidades, se le asigna a esta macromolécula la clase irregular. En caso contrario, se le asigna la clase estructural cuyo centro esté más cercano.

Perfiles de hidrofobicidad. La hidropatía se ha definido como la tendencia hidrofílica o hidrofóbica de la cadena lateral de un aminoácido dado. En el método utilizado, se asigna un determinado valor de hidropatía a cada residuo de la secuencia de la proteína en estudio; a continuación, los valores de los aminoácidos de un segmento de la cadena se suman, y el resultado se asigna al residuo central de la zona. Este segmento es desplazado un residuo sobre la secuencia para calcular el siguiente valor de hidropatía. El proceso continúa hasta haber realizado el cálculo en toda la cadena polipeptídica. El número de aminoácidos involucrados en cada segmento, generalmente un número impar, es elegido por el usuario. En el caso de la GPA, fue empleada una ventana de 7 residuos. Ha sido observado que para proteínas globulares existe una gran correspondencia entre las porciones internas de la cadena principal y las regiones que aparecen en el lado hidrofóbico de la línea de hidrofobicidad promedio, así como entre las zonas externas y las regiones del lado hidrofílico (Kyte y Doolittle, 1982).

Antigenicidad. Se estudia la antigenicidad de los segmentos de la cadena de la GPA debido a que los sitios de la proteína reconocidos por los anticuerpos se encuentran necesariamente expuestos al medio, aportando así información sobre la topografía de la glicoproteína.

La predicción de antigenicidad se realizó con el método desarrollado por Hoop y Woods (1981). Este se basa en el análisis de la secuencia de aminoácidos para determinar las regiones más hidrofílicas. Para ello se emplea una escala de polaridad, similar a la empleada por Kyte y Doolittle (1982), asignando a cada aminoácido un valor numérico y promediándolos en cada segmento de seis residuos de la cadena polipeptídica. En el conjunto de proteínas tomadas como control, los autores encontraron un 100% de correlación entre las regiones con el máximo valor estimado de antigenicidad y los sitios realmente reconocidos por un anticuerpo.

Flexibilidad. La flexibilidad de la cadena polipeptídica tiene relación con la localización de segmentos de ésta que sean reconocidos por anticuerpos, es decir, regiones expuestas al solvente. Para el desarrollo del método, Karplus y Schulz (1985) emplearon una base integrada por 31 proteínas de estructura conocida con una resolución de al menos 3 Å, que tuviesen más de 30 residuos y que no presentaran homologías entre ellas mayores al 50%. Como medida de la flexibilidad de las cadenas polipeptídicas, consideraron los factores de temperatura (B) de los carbonos α en las estructuras cristalográficas. Estos factores representan la dispersión de cada átomo a partir de su posición promedio. A continuación establecieron una clasificación de los 20 diferentes aminoácidos, de acuerdo al promedio de sus factores de temperatura en la base de datos, separándolos en *rígidos* (A, C, F, H, I, L, M, V, W y Y) y *flexibles*. Finalmente, se calcularon los promedios del factor B independientemente para cada tipo de aminoácido que tuviese 0, 1, ó 2 vecinos rígidios en la cadena polipeptídica. Esta última escala permite estimar la flexibilidad de la proteína en la n -ésima posición de la secuencia por el promedio pesado de los valores correspondientes a cada residuo en la región $n-3$ a $n+3$, utilizando como peso los siguientes valores: 0.25, 0.5, 0.75, 1.0, 0.75, 0.5 y 0.25.

Probabilidad superficial. El método de Nishikawa y Ooi (1986) permite estimar la posición radial de una zona de la cadena respecto al centro geométrico de la misma. Se define esta posición con base en el número de carbonos α que rodean a cada aminoácido en una esfera de 14 Å de radio; este número es una medida del grado de exposición al solvente y se relaciona con su distancia al centro de masa de la molécula. En este procedimiento, los autores emplean 92 proteínas para generar un conjunto de parámetros. Con éstos, se estima la posición radial de cada residuo sumando las contribuciones de los nueve aminoácidos adyacentes a cada lado de él mismo:

$$N_{14}_i = \sum_{j=-9}^{j=+9} p(j, i+j) \quad (3-3)$$

en donde N_{14}_i es el número estimado de carbonos α en la esfera de 14 Å alrededor del aminoácido i y p es el parámetro correspondiente al residuo en la posición $i+j$ de la secuencia, tomando en cuenta su posición relativa j respecto al aminoácido i .

Predicción de estructuras secundarias. Ha sido establecido que la exactitud en la predicción de la estructura secundaria de una proteína aumenta al utilizar simultáneamente varios métodos predictivos (Fasman, 1989). Por esta razón, en el presente trabajo fueron empleados tres criterios para estimar la posición de estas estructuras con base en la secuencia de aminoácidos.

El procedimiento de Novotny y Auffray (1984) consiste en el análisis de un conjunto de curvas que pueden emplearse para predecir la estructura secundaria de una proteína: a) la propensión de cada región de la cadena a formar hélices α , hojas β y giros, de acuerdo a los parámetros de Chou y Fasman (1978); b) el perfil de hidrofobicidad empleando la escala de Rose y Roys (1980) y un criterio de suavizado de la curva con una ventana móvil de siete puntos:

$$N_i = \frac{7N_i + 3[2(N_{i-1} + N_{i+1}) + N_{i-2} + N_{i+2}] + 2(N_{i-3} + N_{i+3})}{21} \quad (3-4)$$

en donde N representa el valor de hidrofobicidad correspondiente a la posición de la secuencia que marca el subíndice; c) las posiciones de los residuos cargados (D, E, H, K y R). El número de ciclos de suavizado recomendado por los autores es de 0 a 1 si se desea obtener detalle del perfil dentro de cada estructura secundaria, o de 3 a 6 para la localización de los elementos principales de la estructura secundaria.

El método de Garnier *et al.* (1978) ha mostrado tener una exactitud de aproximadamente 57% al aplicarse a proteínas con estructura conocida (Busetta y Hospital, 1982; Kabsch y Sander, 1983), valor superior al alcanzado por el método clásico de Chou y Fasman (1978). En este procedimiento de predicción, la conformación de cada aminoácido (hélice, cadena extendida, giros e irregular) se establece, utilizando los formalismos de la teoría de la información, tomando en cuenta la contribución de los ocho residuos previos y los ocho posteriores al sitio en cuestión, independientemente para cada tipo de estructura secundaria. El usuario asigna los valores a cuatro parámetros (uno para cada conformación) que son restados del resultado del cálculo anterior para ajustar la predicción a los valores óptimos. La conformación de cada residuo se asigna según el valor más alto obtenido para cada tipo de estructura secundaria.

La probabilidad de que una porción de la secuencia se encuentre formando un giro β , según el método propuesto por Chou y Fasman (1979), se calcula como:

$$p = f(1) \times f(1+1) \times f(1+2) \times f(1+3) \quad (3-5)$$

en donde los valores de f son las frecuencias de ocurrencia de cada aminoácido en las posiciones 1, 2, 3 y 4 de un giro β . Se predicen como regiones de giros aquellas con un valor de p mayor de 7.5×10^{-3} , eliminando las regiones (según los parámetros propuestos por Chou y Fasman (1978)) que posean un promedio de potenciales de formación de giro en el tetrapéptido en cuestión menor a 1.0 o menor al promedio de potenciales de formación de hélices u hojas β . Además, el tetrapéptido con el menor valor de p dentro de los propuestos es eliminado.

Significación de alineamientos. Para determinar la significación o relevancia de las homologías entre la GPA y las lipocalinas con estructura terciaria conocida, se aplicó el método propuesto por Needleman y Wunsch (1970) con las modificaciones sugeridas por Dayhoff (1978) y Feng *et al.* (1985). En este sistema, se crea una matriz rectangular cuyas dimensiones están determinadas por las longitudes de las secuencias a compararse. Este arreglo representa todos los posibles alineamientos de cada aminoácido de una secuencia con los de la otra, asignando a cada elemento de la matriz el valor correspondiente según los datos de la matriz MDM-78 propuesta por Schwartz y Dayhoff (1978) para la comparación de pares de aminoácidos. Una vez concluida la etapa anterior, la matriz se transforma en un conjunto de valores acumulativos al sumar el máximo valor de los elementos de la matriz inicial a los que se encuentran a la derecha en los renglones inferiores. Una trayectoria sobre los valores de la matriz acumulada, que inicia en el primer elemento del primer renglón, y se dirige diagonalmente hacia el elemento opuesto, representa un alineamiento posible, y el valor obtenido al final de la trayectoria es el puntaje o marcador de ese alineamiento. Si la trayectoria se aleja de la diagonal, se produce una inserción de aminoácidos en una de las secuencias; el número de inserciones se controla por un parámetro de penalización (P) que el usuario asigna y se resta del puntaje del alineamiento por cada inserción. Para evitar la presencia de inserciones exageradamente grandes, el usuario otorga un valor al parámetro de tendencia (T) que disminuye el puntaje por cada sitio en las inserciones. El puntaje del alineamiento (M) es comparado con la distribución estadística de un conjunto de puntajes, que se calculan de los alineamientos de una serie de proteínas con la misma composición y longitud que las estudiadas, pero con secuencias diferentes. Estas secuencias se generan cambiando aleatoriamente la posición de los aminoácidos de las proteínas originales. A partir del conjunto de puntajes descritos

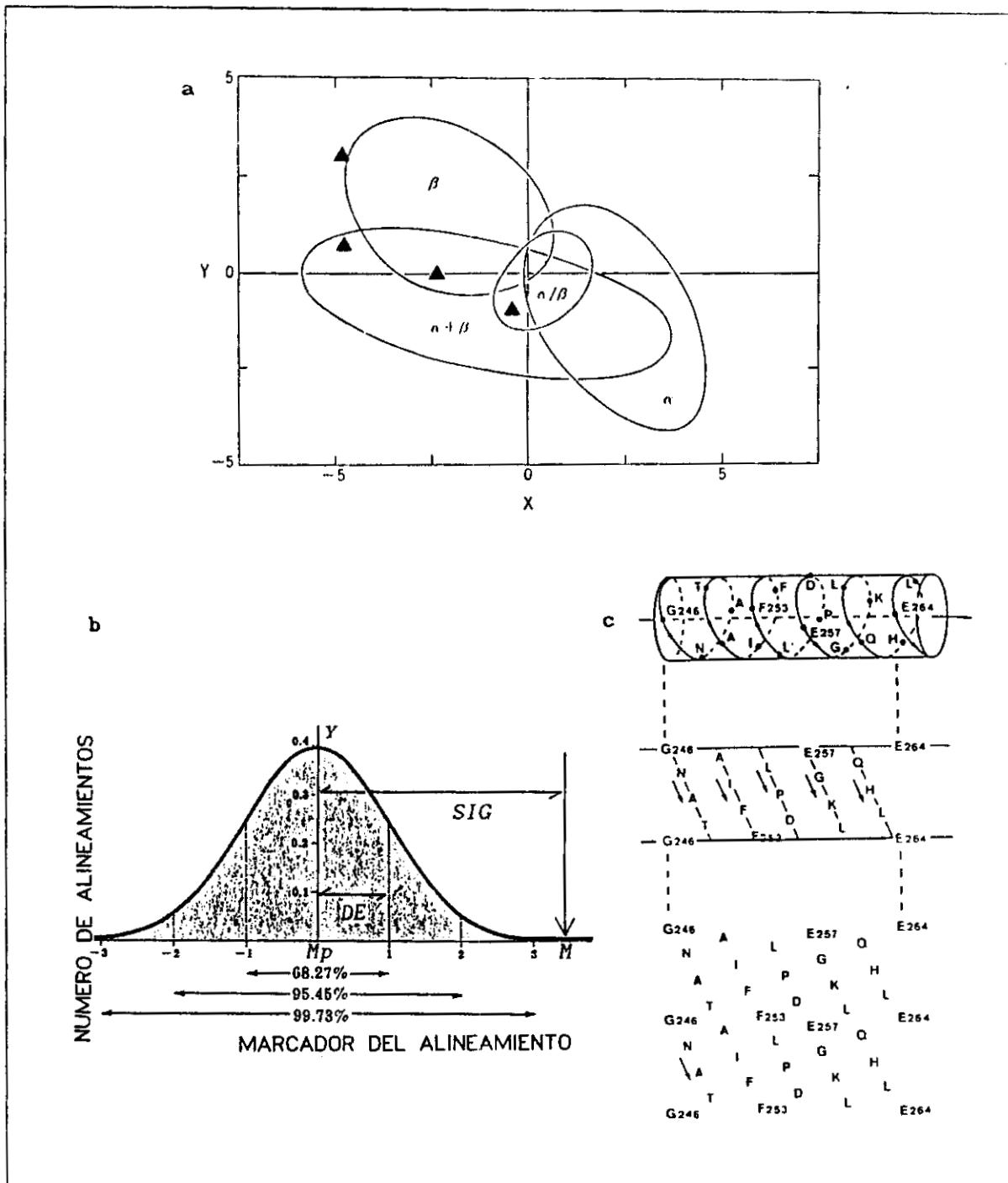


FIGURA 3-1 a) Esquema del espacio de composiciones proyectado en un plano. El eje de las abscisas representa el carácter helicoidal de las composiciones de aminoácidos y el de las ordenadas está relacionado con la tendencia a formar hojas β . Los triángulos señalan las posiciones de las proteínas irregulares, y el origen corresponde a la composición promedio. b) Significación en el marcador o puntaje del alineamiento entre dos proteínas (M) en base al promedio de ellos (Mp) para las secuencias aleatorias, y la desviación estándar (DE). c) Representación de la secuencia de aminoácidos para el análisis de núcleos hidrofóbicos.

anteriormente, se calculan el valor promedio, M_p , y la desviación estándar, DE . Finalmente, la significación del alineamiento, SIG , se calcula en unidades de desviaciones estándar según la expresión:

$$SIG = \frac{M - M_p}{DE} \quad (3-6)$$

El valor obtenido representa la diferencia en el marcador del alineamiento de dos proteínas, respecto al esperado de un par de ellas sin homología entre sí, pero con la misma composición y longitud (figura 3-1b). Un valor para SIG de 3.0 se utiliza comúnmente para considerar que un par de proteínas tienen homología estructural. El criterio empleado en caso de que no exista evidencia independiente de similitud entre las proteínas es que SIG sea mayor de 5.0 desviaciones estándar. Los autores del método recomiendan utilizar un mínimo de 100 comparaciones de secuencias aleatorias para obtener resultados confiables estadísticamente.

Alineamiento. Se emplearon diferentes procedimientos para alinear secuencias de aminoácidos; en todos ellos se utilizan parámetros que el usuario asigna para controlar el número y la longitud de las inserciones. El propuesto por Myers y Miller (1988) calcula inicialmente un valor óptimo de alineamiento para las secuencias completas. A continuación toma el aminoácido central de una de las proteínas y lo ajusta, comparándolo con los aminoácidos de la segunda, hasta obtener el máximo valor posible. En cada mitad de la secuencia el proceso se repite; y posteriormente en cada cuarta parte. Este movimiento de fragmentos de proteína sobre los de una segunda secuencia, continúa hasta que todos los aminoácidos han sido alineados. Los valores del alineamiento entre cada par de residuos son tomados de la matriz de comparación MDM-78 propuesta por Schwartz y Dayhoff (1978) con base en la mutabilidad de cada aminoácido.

Otro método de alineamiento empleado en este trabajo, es el de Higgins y Sharp (1988), el cual emplea la misma matriz de comparación de aminoácidos (MDM-78) que se mencionó en el párrafo anterior. Este método permite estudiar simultáneamente varias secuencias de aminoácidos, alineándolas por pares de una manera muy semejante a las descritas previamente, y construyendo después un dendrograma o árbol taxonómico con base en los datos de similitud. El grupo de secuencias se alinea empezando con las más semejantes y continúa hasta incorporar a todas las proteínas en el alineamiento.

Núcleos hidrofóbicos. El procedimiento de Gaboriaud *et al.* (1987) para la comparación y alineamiento de secuencias de proteínas se basa en el análisis de núcleos hidrofóbicos; éstos se determinan en un arreglo bidimensional de la secuencia de aminoácidos. El método ha probado ser sensible aún en los casos en que las proteínas comparadas posean homología muy pequeña. La representación de los aminoácidos se realiza en una serie de renglones, en las posiciones que ocuparían si la secuencia se enrollara en una hélice α clásica con 3.6 residuos por vuelta y se cortara a lo largo del eje. La secuencia sobre el cilindro formado coincide en el mismo renglón al cabo de 5 vueltas, o sea, 18 residuos. La misma serie de renglones se coloca de manera contigua para facilitar el análisis, pues al desdoblar la superficie del cilindro se alejan algunos aminoácidos consecutivos en la secuencia (figura 3-1c). En esa representación, los residuos hidrofóbicos contiguos se denominan *núcleos hidrofóbicos*. Se definen como contiguos en el plano al aminoácido i , a aquellos residuos con posiciones $i\pm 4$, $i\pm 3$ e $i\pm 1$. Se consideran hidrofóbicos los residuos de I, L, F, W, M, Y y V; como rompedor de núcleos a la P y como miméticos a la A y la C, es decir, sólo se consideran hidrofóbicos si se encuentran en un ambiente no polar. Una vez representadas las dos secuencias de aminoácidos a alinear, se buscan las semejanzas entre los patrones de núcleos hidrofóbicos de las dos proteínas, considerando no sólo su posición en la secuencia sino su tamaño, forma y orientación. Un alineamiento adecuado, según este procedimiento, es aquél que incorpora inserciones o pérdidas de aminoácidos en las regiones hidrofílicas o cercanas a residuos de P o G.

Proyecciones helicoidales. Esta representación de segmentos de la secuencia de aminoácidos de una proteína permite visualizar la posición de las cadenas laterales, tal como se orientarían en una hélice α vista a través de su eje. Esta proyección de hélices en dos dimensiones fue propuesta por primera vez por Shiffer y Edmundson (1967), y consiste en dibujar un círculo sobre el que se representan las cadenas laterales de los aminoácidos que integran la posible hélice, separados por un arco de 100° , considerando que existen 3.6 residuos por vuelta. También es posible calcular la hidrofobicidad promedio del segmento analizado y el momento hidrofóbico de la proyección de la hélice según el sistema de Eisenberg *et al.* (1982).

COMPOSICION Y SECUENCIA.

La secuencia de aminoácidos de la glicoproteína ácida humana fue obtenida por Schmid (1973; Ikenaka *et al.*, 1972) hidrolizando la cadena polipeptídica con bromuro de cianógeno, tripsina y quimotripsina, seguida de la secuenciación de los fragmentos y de su sobreposición para construir la secuencia final. En ese trabajo se encontró que la parte proteica de esta glicoproteína estaba formada de 181 aminoácidos, de los cuales 21 sufrían reemplazo. Las dos variantes de la GPA fueron encontradas en una proporción de 3 a 1 (Ikenaka *et al.*, 1972; Dente *et al.*, 1985). En otros estudios, se han reportado otras variantes que en menores proporciones integran esta glicoproteína. No debe olvidarse que una fuente de la microheterogeneidad de esta macromolécula reside en la parte constituida por carbohidratos. Por esta razón no puede asegurarse que tales variantes menores contengan nuevos reemplazos de aminoácidos. Estudios más recientes sobre la estructura primaria de la GPA, en los que se han secuenciado los genes que codifican a esta proteína (Dente *et al.*, 1985; 1987), muestran la existencia de tres genes organizados en seis exones. Dos de éstos son idénticos entre sí; el tercero difiere de los anteriores cambiando la naturaleza de 22 de los 183 aminoácidos que integran a la GPA.

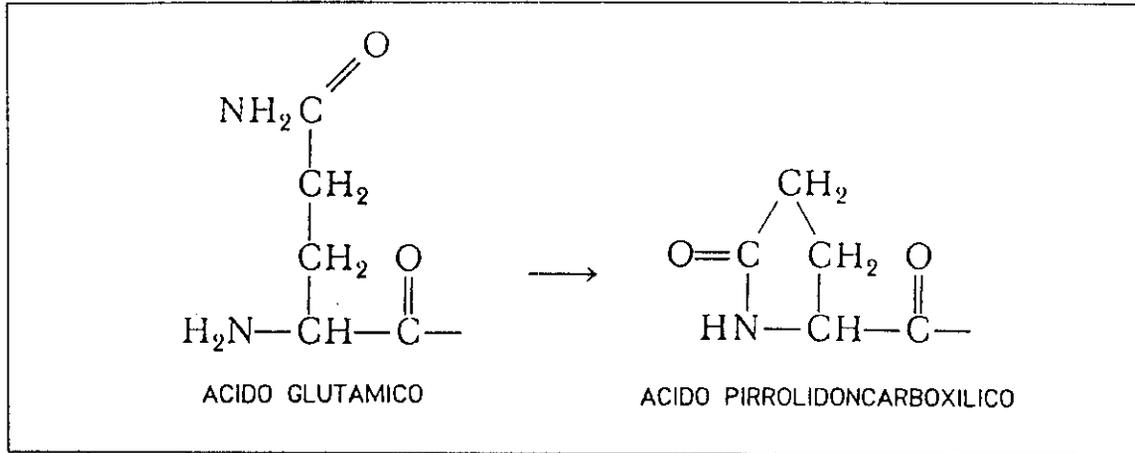


FIGURA 3-2. Condensación de la cadena lateral de un residuo de ácido glutámico con su grupo amino libre para formar un residuo de ácido pirrolidoncarboxílico o piroglutámico.

En este trabajo se eligió la secuencia deducida del gen A de la GPA por ser la variante más abundante, además de que la información obtenida de los ácidos nucleicos es más confiable que la secuenciación de péptidos. Por esta razón, las referencias en este trabajo respecto a la secuencia de la GPA, involucran a los 183 aminoácidos reportados por Dente *et al.* (1985), salvo cuando

explícitamente se mencione la variante B obtenida del gen menos expresado. En ambos casos se usa como primer aminoácido el ácido glutámico condensado, con el que se forma un residuo de ácido piroglutámico o pirrolidonicarboxílico (X en las abreviaturas empleadas en este trabajo; figura 3-2). La numeración de los aminoácidos de la GPA utilizada aquí, coincide con la de Dente *et al.* (1985).

En la figura 3-3 se muestran diferentes esquemas de la estructura de la GPA; la cadena polipeptídica se representa con línea continua, y en ella pueden apreciarse las cinco cadenas de carbohidratos unidas a los residuos N15, N38, N54, N75 y N85, así como los puentes disulfuro que unen la C5 con la C147 y la C72 con la C164 (figura 3-3a). Por otra parte, se muestran (figura 3-3b) las diferentes posiciones de los carbohidratos (*), los reemplazos de aminoácidos (|) y las uniones entre exones del gen de la GPA (↑). En esta figura (3-3b) puede apreciarse también que las cinco uniones de carbohidratos a la cadena polipeptídica de la GPA se localizan en la mitad de la cadena que corresponde al extremo amino; estos carbohidratos se encuentran separados por una distancia de 23, 16, 21 y 10 residuos de aminoácidos en la estructura primaria. Además, la segunda mitad tiene una gran cantidad de residuos cargados; la GPA presenta una elevada solubilidad, aún en estado desnaturalizado, característica que le confieren tanto los carbohidratos unidos a la primera mitad, y en especial el ácido siálico, como la gran cantidad de cargas eléctricas de la segunda mitad. También destaca la distribución de los reemplazos de aminoácido en la cadena; puede apreciarse una región de ocho residuos, entre los aminoácidos 110 y 117, donde se concentran 7 de las sustituciones de cadenas laterales. Otras zonas con menor grado de variabilidad se encuentran entre los aminoácidos 70-77, 88-103, y 149-156, donde se observa el reemplazo de 4, 6 y 3 residuos en una zona de 8, 16 y 8 aminoácidos, respectivamente. Las otras sustituciones se encuentran aisladas en los aminoácidos 20 y 126 para hacer un total de 22. De estas sustituciones de aminoácidos, 18 son causadas por el cambio de una sola base en el codón correspondiente, mientras que las dos restantes necesitan del cambio en dos de las tres bases del codón para cambiar el aminoácido de esa posición. En la figura 3-1c se muestra la secuencia de la GPA, y en la tabla 3-I se resume la información sobre las sustituciones en la parte proteica de esta glicoproteína.

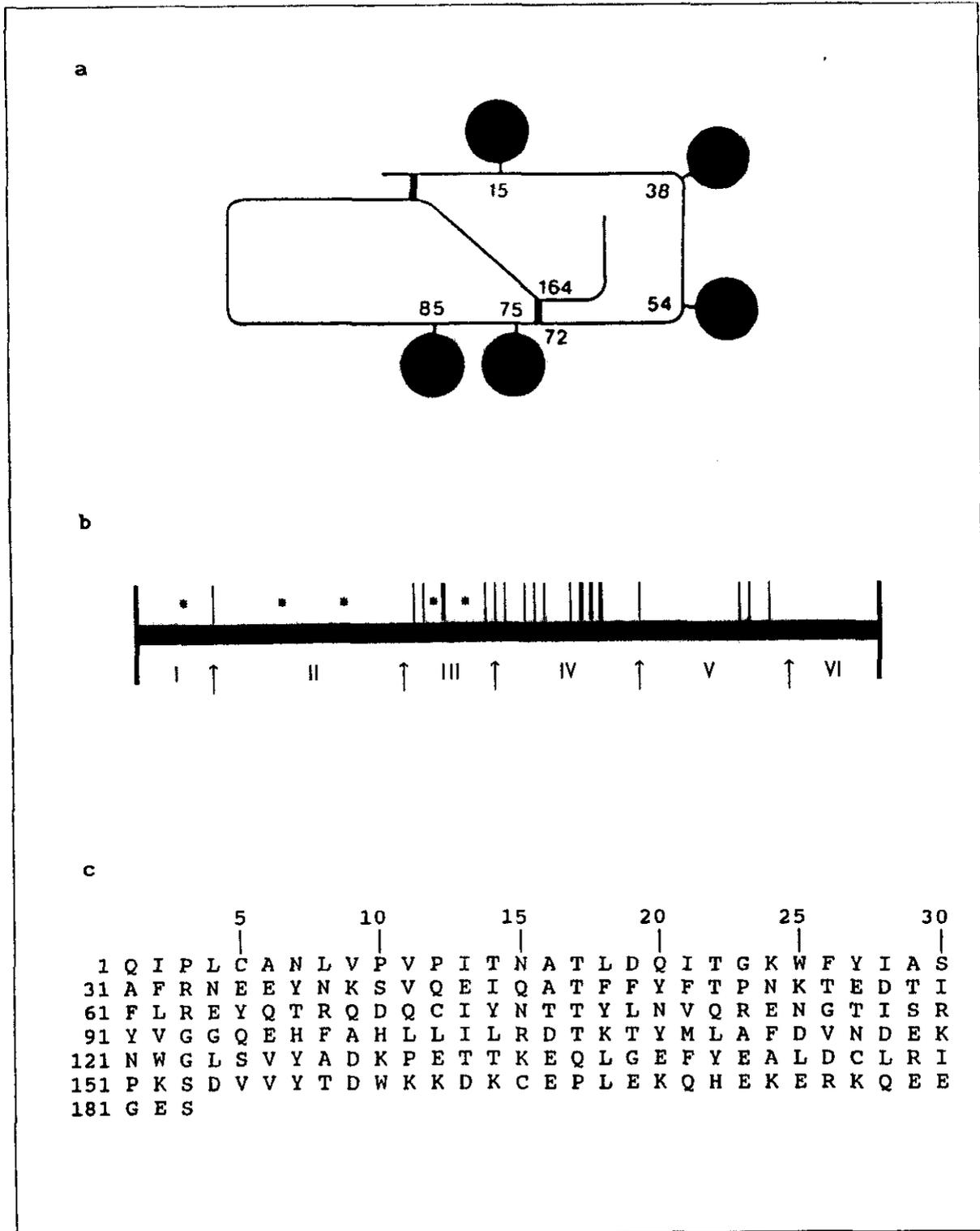


FIGURA 3-3. Esquemas de la estructura de la GPA. a) Los círculos sombreados corresponden a las cadenas de carbohidratos unidas a residuos de asparragina. Las líneas gruesas representan enlaces disulfuro. b) Posición de los residuos de carbohidrato (*), reemplazos de aminoácidos entre las dos variantes (|) y uniones entre exones (↑). c) Secuencia de aminoácidos.

De la tabla 3-I puede apreciarse que no existe reemplazo de His, Pro y Trp lo cual está de acuerdo con la baja mutabilidad reportada para estos residuos (Dayhoff, 1978). Sin embargo Arg, Leu, Phe y Tyr poseen aún menor mutabilidad y presentan reemplazo. Los aminoácidos involucrados con mayor frecuencia en los reemplazos son: Phe con 7 ocasiones, Val con 6, Leu y Arg con 4, y Asp, Asn, Ile y Ser con 3. Los intercambios que ocurren dos veces son Q ↔ R, D ↔ N, F ↔ I, S ↔ T y F ↔ V.

T A B L A 3 - I
Sustituciones de aminoácidos en las variantes de la GPA.

Posición	Sustitución	Codones	Núm. cambios	Bases del codón	Exón
<u>GEN A → GEN B</u>					
20	Gln→Arg	CAG→CGG	1	2a.	I
70	Asp→Asn	GAC→AAC	1	1a.	III
73	Ile→Phe	ATC→TTC	1	1a.	III
76	Thr→Ser	ACC→TCC	1	1a.	III
77	Thr→Ser	ACC→AGT	2	2a. y 3a.	III
88	Ile→Val	ATC→GTC	1	1a.	III
92	Val→Glu	GTG→GAG	1	2a.	IV
95	Gln→Arg	CAA→CGA	1	2a.	IV
98	Phe→Val	TTC→GTT	2	1a. y 3a.	IV
101	Leu→Asn	TTG→CTG	1	1a.	IV
103	Ile→Phe	ATC→TTC	1	1a.	IV
110	Tyr→Leu	TAC→TTG	2	2a. y 3a.	IV
112	Leu→Phe	CTT→TTT	1	1a.	IV
113	Ala→Gly	GCT→GGT	1	2a.	IV
114	Phe→Ser	TTT→TCC	2	2a. y 3a.	IV
115	Asp→Tyr	GAC→TAC	1	1a.	IV
116	Val→Leu	GTG→CTG	1	1a.	IV
117	Asn→Asp	AAC→GAC	1	1a.	IV
126	Val→Phe	GTC→TTC	1	1a.	IV
149	Arg→Cys	CGC→TGC	1	1a.	V
152	Lys→Arg	AAG→AGG	1	2a.	V
156	Val→Met	GTG→ATG	1	1a.	V

La composición de aminoácidos de la GPA y de otras 15 lipocalinas se muestra en la tabla 3-II. En ésta puede apreciarse que el número de residuos de cisteína presentes (y, por lo tanto, de posibles puentes disulfuro) es, a excepción de la PBP que carece de ellos, mayor o igual a dos; y en el 70% de las secuencias, mayor o igual a cuatro. Esto sugiere que, aunque el número de puentes disulfuro tiende a conservarse en la familia, no es una característica indispensable para la estabilidad estructural. En el caso de los residuos de prolina, los cuales han mostrado intervenir en el mecanismo de doblado de diferentes proteínas (Schmid, 1992), se observa una variación desde uno hasta 12 residuos. Esto parece indicar que los residuos de prolina no son esenciales en el mecanismo por el cual se adquiere la conformación nativa.

En la figura 3-4 se representa la frecuencia de aparición de los veinte aminoácidos en diferentes conjuntos de proteínas. Se muestra la composición en proteínas de vertebrados (excluyendo las humanas), de humanos (Doolittle, 1989) y en las secuencias de las lipocalinas de la tabla 3-II, excluyendo la GPA, la cual se exhibe independientemente. Se observa que la glicoproteína contiene un número menor de residuos de L, A, G, S, V, H, y M que los promedios de otras proteínas, incluso que el conjunto de 15 lipocalinas ya mencionado. Por otra parte, la glicoproteína presenta en su secuencia, mayor cantidad de residuos de E, T, Q, F, y Y que los promedios de los conjuntos mostrados. Al agrupar los diferentes aminoácidos de acuerdo a sus características hidropáticas (tabla 3-II y figura 3-4) y realizar nuevamente las comparaciones, resulta que el conjunto de lipocalinas no se aparta notablemente de los porcentajes correspondientes a las proteínas de vertebrados o humanos, salvo por un aumento moderado en aminoácidos aromáticos y ácidos. En el caso de la GPA, se observan acrecentadas estas mismas diferencias, aunadas a la disminución de residuos alifáticos, por una parte, y al aumento en el número de aminoácidos polares, por la otra. La GPA contiene un mayor número de aminoácidos polares (y sobre todo ácidos) con un aumento en el porcentaje de residuos aromáticos que, tal vez, compense la disminución en la cantidad de cadenas laterales alifáticas.

Aplicando el método de Nakashima *et al.*, puede predecirse el tipo de plegamiento general de una proteína a partir sólo de su composición. La tabla 3-III muestra los resultados del análisis de la secuencia completa de la GPA y de la aplicación del método a dos segmentos de esta proteína: 1-130 (GP1) y 129-183 (GP2). Como se describirá más adelante, se predice que existe un

TABLA 3 - II. Composición de aminoácidos

PROTEINA	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	Mis	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val	*	ARO	ALI	POL	+/-	MFO	HFI	
GPA	9	8	10	11	4	12	19	7	3	10	14	14	1	9	7	6	16	3	11	9	25	30	23	43	44	55	66	99
PTB	8	1	11	10	4	2	12	15	6	7	6	20	0	5	6	12	7	5	16	20	27	22	26	41	32	49	67	81
TNS	15	2	12	14	4	4	10	11	6	9	11	19	2	9	8	13	9	3	14	14	27	24	26	51	38	51	77	89
BLG	15	3	5	10	5	9	16	4	2	10	22	15	4	4	8	7	8	2	4	9	20	26	10	60	29	46	70	75
PTR	13	14	9	18	6	9	9	11	2	4	13	10	4	10	5	11	9	4	8	13	26	27	22	47	38	53	69	91
AFR	8	3	14	6	4	8	14	10	2	15	9	10	3	9	5	3	13	1	7	7	15	20	17	42	38	35	59	73
ATH	9	10	6	8	3	6	15	14	4	13	11	10	5	6	12	10	17	3	8	11	24	23	17	49	39	47	66	86
AZM	7	10	10	12	3	3	13	12	3	10	14	13	4	10	1	8	10	1	7	9	26	25	18	44	31	51	62	82
ALD	10	4	16	7	5	7	12	6	2	11	15	1	3	7	12	7	11	4	7	12	17	19	18	51	41	36	69	77
CBG	21	15	1	8	3	17	9	13	5	4	17	5	1	11	8	10	9	2	6	17	25	17	19	60	37	42	79	79
P14	9	7	9	10	4	7	11	4	4	8	23	10	8	6	10	5	13	3	4	7	21	21	13	55	34	42	68	76
PDA	15	5	3	11	2	2	13	10	5	7	17	14	4	7	3	11	14	1	8	14	24	24	16	57	30	48	73	78
POR	9	2	6	9	5	11	12	10	5	8	19	9	7	6	7	12	2	5	12	26	26	21	14	46	36	47	60	83
POL	7	8	12	13	5	9	15	9	8	7	11	10	1	5	3	6	10	1	6	12	26	28	12	38	37	54	50	91
PRO	8	9	12	11	3	4	19	9	5	11	16	9	3	9	2	10	9	1	4	6	23	30	14	44	35	53	58	85
PTP	5	5	11	13	0	6	19	9	6	5	11	15	0	9	5	6	11	3	5	15	26	32	17	36	34	58	53	92
PUR	15	7	5	15	6	10	7	13	0	8	10	10	5	7	9	14	10	3	12	9	17	22	22	47	39	39	69	78

PROTEINA	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	Mis	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val	*	ARO	ALI	POL	+/-	MFO	HFI	
GPA	4.9	4.4	5.5	6.0	2.2	6.6	10.4	3.8	1.6	5.5	7.7	7.7	.5	4.9	3.8	3.3	8.7	1.6	6.0	4.9	13.7	16.4	12.6	23.5	26.0	30.1	36.1	54.1
PTB	4.6	.6	6.4	5.8	2.3	1.2	6.9	8.7	3.5	4.0	3.5	11.6	.0	2.9	3.5	6.9	4.0	2.9	9.2	11.6	15.6	12.7	15.0	23.7	18.5	29.3	38.7	46.8
TNS	7.9	1.1	6.3	7.4	2.1	2.1	5.3	5.8	3.2	4.8	5.8	10.1	1.1	4.8	4.2	6.9	4.8	1.6	7.4	7.4	14.3	12.7	13.8	27.0	20.1	27.0	40.7	47.1
BLG	9.3	1.9	3.1	6.2	3.1	5.6	9.9	2.5	1.2	6.2	13.6	9.3	2.5	2.5	4.9	4.3	4.9	1.2	2.5	5.6	12.3	16.0	6.2	37.0	17.9	28.4	43.2	46.3
PTR	7.1	7.7	4.9	9.9	3.3	4.9	4.9	6.0	1.1	2.2	7.1	5.5	2.2	5.5	2.7	6.0	4.9	2.2	4.4	7.1	14.3	16.8	12.1	25.8	20.9	29.1	37.9	50.0
AFR	5.3	2.0	9.3	4.0	2.6	5.3	9.3	6.6	1.3	9.9	6.0	6.6	2.0	6.0	3.3	2.0	8.6	.7	4.6	4.6	9.9	13.2	11.3	27.8	25.2	23.2	39.1	48.3
ATH	5.0	5.5	3.3	4.4	1.7	3.3	8.3	7.7	2.2	7.2	6.1	5.5	2.8	3.3	6.6	5.5	9.4	1.7	4.4	6.1	13.3	12.7	9.4	27.1	21.5	26.0	36.5	47.5
AZM	4.4	6.3	6.3	7.5	1.9	1.9	8.1	7.5	1.9	6.3	8.8	8.1	2.5	6.3	.6	5.0	6.3	.6	4.4	5.6	16.3	15.6	11.3	27.5	19.4	31.9	38.8	51.3
ALD	5.9	2.4	9.5	4.1	3.0	4.1	7.1	3.6	1.2	6.5	8.9	6.5	1.8	4.1	7.1	4.1	6.5	2.4	4.1	7.1	10.1	11.2	10.7	30.2	24.3	21.3	43.4	43.4
CBG	11.5	8.2	.5	4.4	1.6	9.3	4.9	7.1	2.7	2.2	9.3	2.7	.5	6.0	4.4	5.5	4.9	1.1	3.3	9.3	13.7	9.3	10.4	33.0	20.3	23.1	63.4	43.4
P14	5.6	4.3	5.6	6.2	2.5	4.3	6.8	2.5	2.5	4.9	14.2	6.2	4.9	3.7	6.2	3.1	8.0	1.9	2.5	4.3	13.0	13.0	8.0	34.0	21.0	25.9	42.0	46.9
PDA	9.0	3.0	1.8	6.6	1.2	1.2	7.8	6.0	3.0	4.2	10.2	8.4	2.4	4.2	1.8	6.6	8.4	.6	4.8	8.4	14.5	14.5	9.6	34.3	18.1	28.9	44.0	47.0
POR	5.5	1.2	3.7	5.5	3.0	6.7	7.3	6.1	3.0	4.9	4.9	11.6	5.5	4.3	3.7	4.3	7.3	1.2	3.0	7.3	15.9	12.8	6.5	28.0	22.0	28.7	36.6	50.6
POL	4.4	5.1	7.6	8.2	3.2	5.7	9.5	5.7	5.1	4.4	7.0	6.3	.6	3.2	1.9	3.8	6.3	.6	3.8	7.6	16.5	17.7	7.6	24.1	23.4	34.2	31.6	57.6
PRO	5.0	5.6	7.5	6.9	1.9	2.5	11.9	5.6	3.1	6.9	10.0	5.6	1.9	5.6	1.3	6.3	5.6	.6	2.5	3.8	14.4	18.8	8.8	27.5	21.9	33.1	36.3	55.0
PTP	3.1	3.1	6.9	8.2	.0	3.8	11.9	5.7	3.8	3.1	6.9	9.4	.0	5.7	3.1	3.8	6.9	1.9	3.1	9.4	16.4	20.1	10.7	22.6	21.4	36.5	33.3	57.9
PUR	8.6	4.0	2.9	8.6	3.4	5.7	4.0	7.4	.0	4.6	5.7	5.7	2.9	4.0	5.1	8.0	5.7	1.7	6.9	5.1	9.7	12.6	12.6	26.9	22.3	22.3	39.4	44.6

FAMILIA	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	Mis	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val	*	ARO	ALI	POL	+/-	MFO	HFI	
VERTEBRADOS	7.2	4.9	4.1	5.4	2.5	3.9	6.8	7.6	2.2	5.0	8.7	6.5	2.3	4.1	4.9	7.0	5.6	1.3	3.2	6.8	13.6	12.2	8.6	21.4	20.6	25.8	30.0	46.4
HUMANOS	6.3	4.9	4.1	5.1	3.0	4.3	6.8	7.1	2.7	4.3	9.2	5.8	2.0	4.0	5.6	7.5	6.1	1.5	3.3	6.5	13.4	11.9	8.8	19.5	22.0	25.3	28.3	47.3
LIPOCALINAS	6.1	3.6	5.6	6.7	2.3	3.9	7.9	5.8	2.4	5.3	7.8	7.8	2.2	4.5	3.8	5.2	6.5	1.5	4.6	6.8	13.8	14.6	10.6	17.6	21.2	28.4	28.2	49.6

NUMERO DE AMINOACIDOS

PORCENTAJE DE AMINOACIDOS

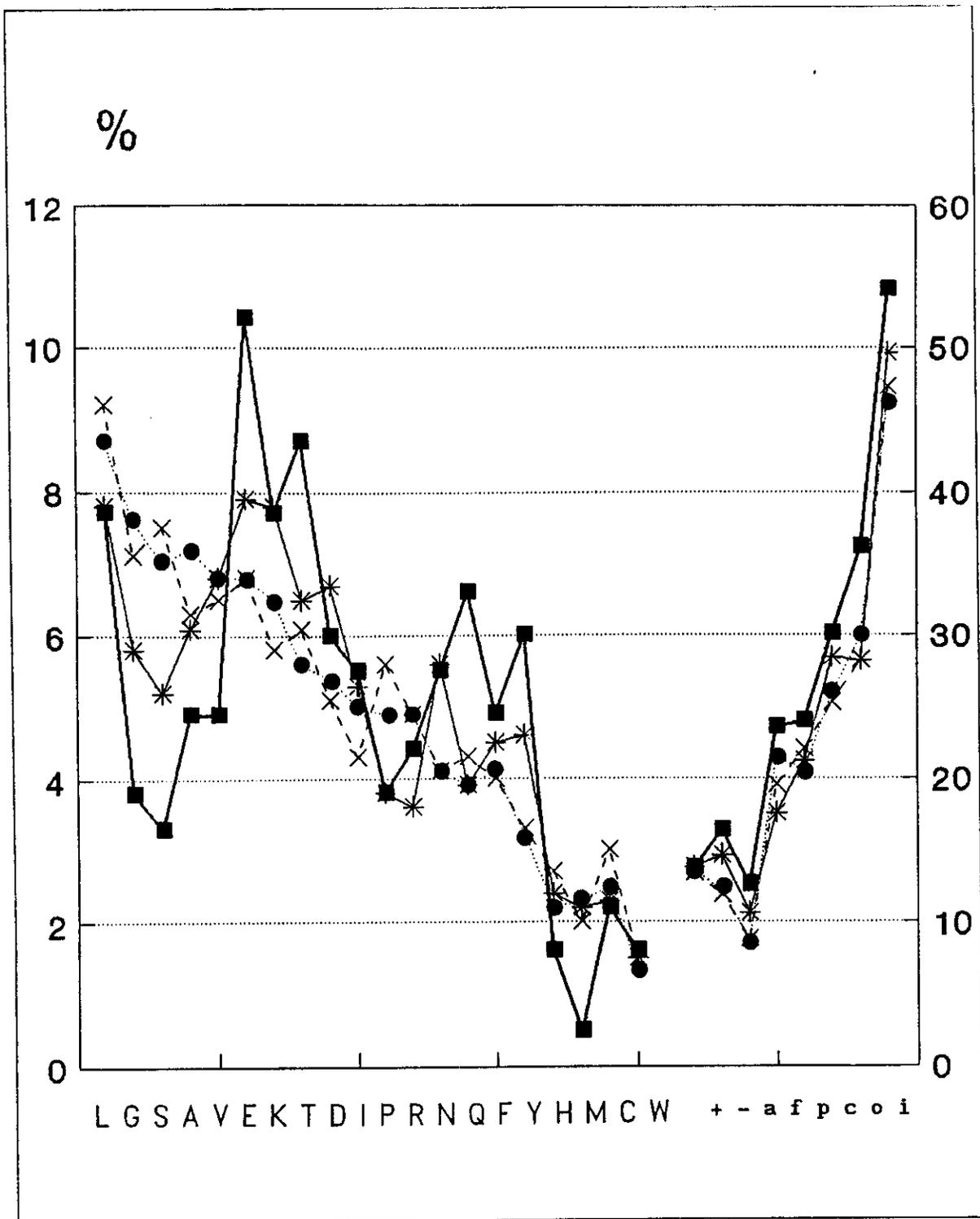


FIGURA 3-4. Composición de aminoácidos de diferentes grupos de proteínas [vertebrados (●), humanos (×) y lipocalinas (*)] y de la GPA (■). También se muestran los porcentajes de aminoácidos ácidos (D, E; -), básicos (H, K, R; +), aromáticos (F, W, Y; a), alifáticos (A, I, L, M, V; f), polares (Q, N, S, T; p), cargados (ácidos + básicos; c), hidrofóbicos (aromáticos + alifáticos; o) e hidrofílicos (cargados + polares; i).

número alto de hojas β en el primer segmento y hélices en el segundo. Las clasificaciones que se obtienen en estos análisis asignan a la secuencia de la GPA la clase $\alpha+\beta$, al igual que a la proteína transportadora de bilina (PTB) y a la insecticiana (INS), dos de las lipocalinas con estructura conocida. En el caso de los segmentos GP1 y GP2, su composición los clasifica como $\alpha+\beta$ y α , respectivamente; esto indica que los aminoácidos en el segmento GP2 son típicamente helicoidales. Sin embargo, en el caso de la transportadora de retinol (PTR) y de la β -lactoglobulina (BLG), las otras dos lipocalinas cuya estructura tridimensional ha sido determinada, se asignan las clases α/β y α , respectivamente. No debe olvidarse que las regiones de cada clase estructural en el espacio de composiciones, se traslapan sensiblemente (figura 3-1a).

T A B L A 3 - III

Clasificación estructural en base a la composición de aminoácidos.

CLASE	AGP	GP1	GP2	PTB	INS	PTR	BLG
Origen	3.98	4.64	7.48	5.81	3.72	3.61	3.82
α	4.69	5.63	7.15	6.08	3.73	4.04	3.57
β	4.89	5.23	8.43	6.03	4.29	4.28	5.17
α/β	4.55	5.28	7.53	5.70	3.85	3.82	3.88
$\alpha+\beta$	4.40	4.47	8.58	5.16	3.72	4.92	5.69
Irregular	8.36	9.12	9.55	9.05	8.05	7.78	7.89
Asignación:	$\alpha+\beta$	$\alpha+\beta$	α	$\alpha+\beta$	$\alpha+\beta$	α/β	α

Como puede observarse de los perfiles de la figura 3-5a, la estructura primaria de la GPA muestra zonas de alta hidrofobicidad, alternadas con regiones polares, y un segmento de gran hidrofiliidad, debida a la concentración de aminoácidos con carga eléctrica (11 de 15 residuos) en el extremo carboxilo. Los perfiles de las otras lipocalinas, aunque también alternan porciones hidrofílicas e hidrofóbicas, presentan gran variabilidad en sus patrones, sobre todo en la intensidad de los picos. Como se mencionó, las cadenas de carbohidratos de la GPA se unen a la parte proteica en cinco

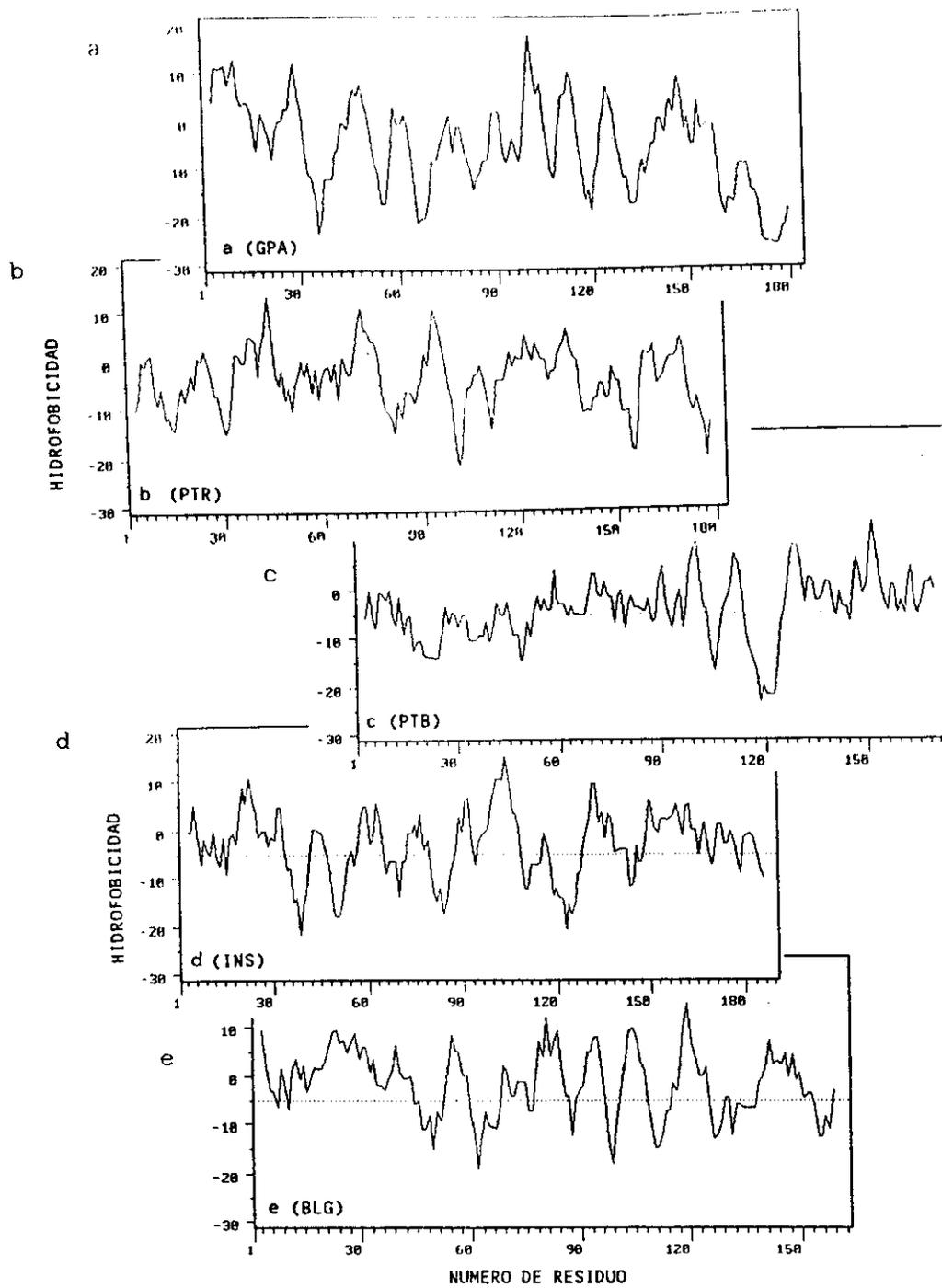


FIGURA 3-5. Perfiles de hidrofobicidad calculados en intervalos de siete aminoácidos de la GPA (a) y de las lipocalinas con estructura tridimensional conocida: PTR (b), PTB (c), INS (d) y BLG (e). La línea punteada indica el valor promedio observado en proteínas.

cadenas laterales de asparagina: N15, N38, N54, N75 y N85. Las posiciones de estos aminoácidos en la secuencia, como se muestra en la figura 3-3, se encuentran en la primera mitad de la cadena polipeptídica. Tres de ellos en regiones marcadamente hidrofílicas y los dos restantes a pocos aminoácidos de distancia de zonas con esta característica.

En la figura 3-6 se muestran los perfiles de antigenicidad, flexibilidad y probabilidad superficial para la cadena polipeptídica de la GPA. En estas gráficas destacan las regiones vecinas a los aminoácidos 28, 50, 62, 77, 89, 101, 113, 125, 146 y 158 por presentar valores pequeños de antigenicidad y flexibilidad, así como valores grandes para N14 (el número de carbonos α cercanos). Los tres criterios señalan a estos segmentos como internos, al igual que el perfil de hidrofobicidad de la figura 3-5a. De manera contrastante, se observan varios segmentos con alta probabilidad de encontrarse en la superficie molecular, es decir, son zonas muy flexibles y altamente antigénicas, para las que la estimación de N14 es pequeña. Estas porciones de la secuencia se encuentran alrededor de los aminoácidos 22, 34, 55, 68, 85, 95, 106, 119, 133, 152, 163 y 178. Esta información será de gran utilidad en la asignación de estructuras secundarias y en el alineamiento con una de las lipocalinas caracterizadas estructuralmente.

Estudiando el alineamiento de las secuencias de las glicoproteínas ácidas α_1 humana, de rata, de conejo y la porción conocida de las variantes de la de ratón (figura 3-7a), puede observarse que las de fuentes no humanas no conservan la G23 característica de las lipocalinas, aunque mantienen el resto de la huella digital. La glicoproteína de rata no posee el puente disulfuro C5-C147, pero todas las secuencias conservan el C72-C165. Otra característica es que algunos sitios de N-glicosilación o no están presentes o se encuentran en diferentes posiciones de la cadena. Comparando por pares la secuencia de la variante A (*Humana 1*) con las de otros organismos, se encuentra la mayor homología con la glicoproteína de conejo (104 aminoácidos idénticos = 56.8% y 23 similares = 12.6%, sin necesidad de inserciones), mientras que la semejanza menor es con la glicoproteína ácida α_1 de ratón 1 (62 idénticos = 44.9% de los 138 residuos conocidos, 25 similares = 18.1% y la necesidad de una inserción). Las homologías son menores si se analiza la variante B de la GPA. En el alineamiento entre las variantes de GPA y la glicoproteína ácida α_1 de rata, se requiere de cuatro inserciones, tres en la GPA y una en la otra secuencia. Los aminoácidos completamente conservados, marcados con un asterisco en la

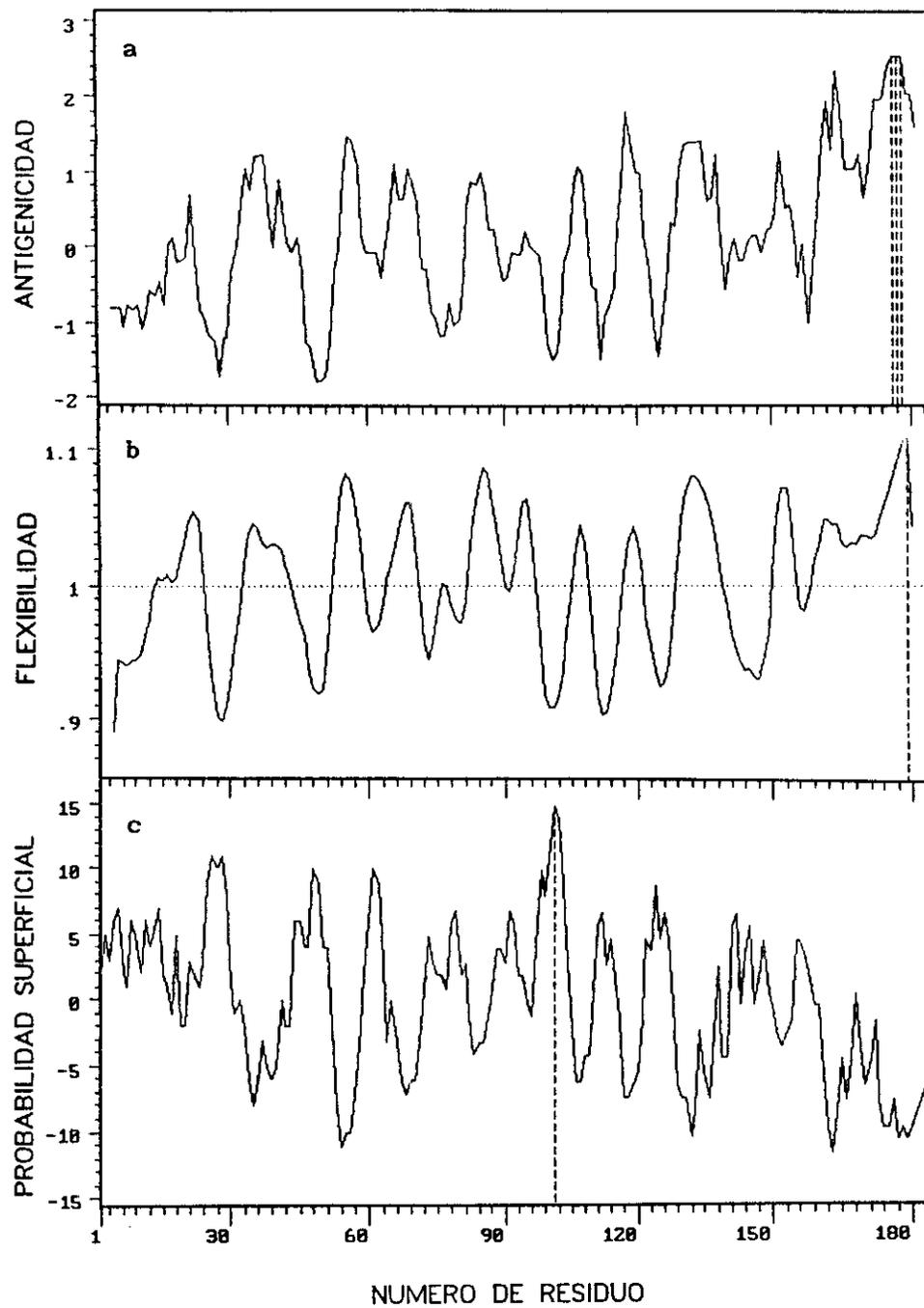
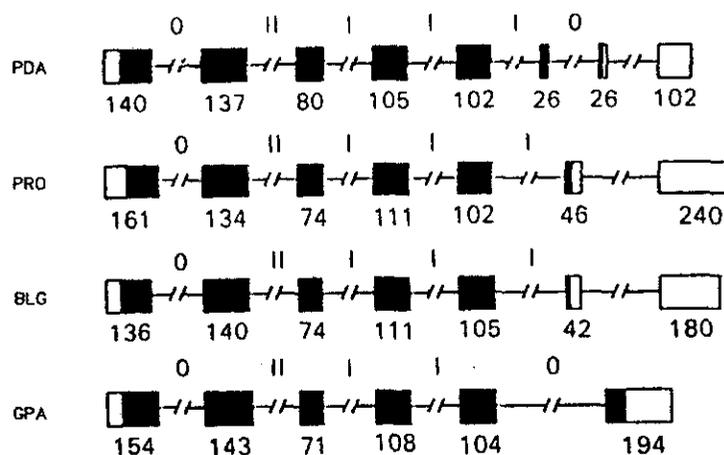


FIGURA 3-6. Perfiles de antigenicidad (a), flexibilidad (b) y probabilidad superficial (c) de la secuencia de aminoácidos de la GPA.

Humana 1	QIPLCANLV-PVPITNATLDQITGKWFYIASAFRNEEYKNSVQEIQATFFFYFTPN	54
Humana 2	QIPLCANLV-PVPITNATLDRIITGKWFYIASAFRNEEYKNSVQEIQATFFFYFTPN	54
Conejo	QDPACANFS-TSPI TNATLDQLSHKWFYFASAFRNPYKQLVQHTQAAFFFYFTA	54
Rata	QNPEPANITLGIPIITNGTLRWLSDKWFYMGAAFRDPVFKQAVQTIQTEYFYLTPN	55
Ratón 1	-----LTPN	4
Ratón 2	-----LTPN	4
	..*	
Humana 1	KTEDTIFLREYQTRQDQCIYNTTYLNVQRENGTISRIVGGQEHFAHLLILRDTKT	109
Humana 2	KTEDTIFLREYQTRQDQCFYNSSYLNVQRENGTISRVEGGREHVAHNLI FRDTKT	109
Conejo	KEEDTLLREYITNTNCFYNSSIVRVQRENGTLSKHDGIRNSVADLLLLLRDPGS	109
Rata	LINDTIELREFQTTDDQCVYNFTHLGVQRENGTLSKAGAVKIFAHLLIVLKKHGT	110
Ratón 1	LINDTMELREYHTIDHCVYNSTHLGIQRENGTLSKYVGGVKIFADLIVLKMIGA	59
Ratón 2	LINDTMELRESQTI GDQCVYNSTHLGFQRENGTFSKYEGGVETFAHLIVLPKHGA	59
	**.* ** * . * * * . * * * * * * * . * . * . . .	
Humana 1	YMLAFDVNDEKNWGLSVYADKPETTKEQLGFEYALDCLRIKSDVVYTDWKKDK	164
Humana 2	LMFGSYLDDEKNWGLSFYADKPETTKEQLGFEYALDCLCIPRSDVMYTDWKKDK	164
Conejo	FLLVFFAGKEQDKGMSLYTDKPKASTEQLLEEFYEALTCGLMKNTEVVYTDWTKDL	164
Rata	FMLAFNLTDEN-RGLSFYAKKPDLSPELRKIFQQAVKDVGMDESEIVFVDWTKDK	164
Ratón 1	FMLAFDLKDEKKRGLSLNAKRPDITPELREVFQKAVTHVGMDESEIIFVDWKKDR	114
Ratón 2	FMLAFDLKDEKKRGLSLYAKRPDITPELQEVFQKAVTHVGMDESEIIFVDWKKDR	114
	..* . * . * * * . . . * . * . . . * * * * * *	
Humana 1	CEPLEKQH---EKERKQE--EGES	183
Humana 2	CEPLEKQH---EKERKQE--EGES	183
Conejo	CEPLEKQH---EEERKKE--KAES	183
Rata	CSEQQKQLELEKETKKETKK-DP	187
Ratón 1	CSQQEKQLELEKETKKDPEEGQA	138
Ratón 2	CGQQEKQLELEKETKKDPEEGQA	138
	* . * * . * * * * . . .	

a



b

FIGURA 3-7. a) Alineamiento de las secuencias de aminoácidos de la glicoproteína ácida α_1 de diversos mamíferos. Los números distinguen variantes del mismo organismo. Se marcan las posiciones idénticas en todas las secuencias (*) y las conservadas (·). b) Estructura de los genes que codifican lipocalinas. Los rectángulos representan los exones, se muestra en la parte inferior su tamaño en pares de bases. Los símbolos 0, I y II indican si el intrón se inserta entre codones o después de la primera o segunda bases, respectivamente.

figura 3-7a, se encuentran distribuidos en toda la secuencia, por lo que se puede afirmar que todas estas glicoproteínas mantienen las características de la GPA, en particular un segmento rico en cadenas laterales con carga eléctrica en el extremo carboxilo.

La figura 3-7b muestra la organización de los genes que codifican a algunas de las lipocalinas. Destaca el hecho de que los cuatro genes poseen el mismo número de exones, de longitudes muy similares y que conservan notablemente la fase, esto es, la posición entre las tres bases del codón en el cual se insertan los intrones correspondientes. Esto es una evidencia sólida para proponer la relación entre las macromoléculas estudiadas, sugiriendo que provienen de una misma proteína ancestral.

PREDICCIÓN DE ESTRUCTURA SECUNDARIA.

Respecto a la estructura secundaria de la glicoproteína ácida α_1 , la figura 3-8 muestra las probabilidades de formación de hélices α , hojas β y giros para las diferentes regiones de la cadena principal, empleando diferentes algoritmos. Es interesante notar que los segmentos de la cadena polipeptídica 22-29, 45-50, 72-81, 87-92 y 109-113 muestran simultáneamente bajo los métodos de Novotny y Auffray (1984) y Garnier *et al.* (1978), una alta probabilidad de encontrarse en conformación de cadena β . Estas regiones, como se mostrará en el alineamiento entre la GPA y la lipocalina elegida como modelo inicial, corresponden a las porciones de hoja β denominadas A, B, D, E y G en la estructura de las lipocalinas. Por otra parte, se observa la coincidencia en las tendencias a formar hélices de las zonas 133-145 y 168-182. Esta última, localizada en el extremo carboxilo de la cadena, es el segmento extremadamente polar en donde se encuentran diez aminoácidos con carga eléctrica a pH fisiológico: seis residuos de ácido glutámico, tres de lisina y uno de arginina (además de dos residuos de glutamina y uno de histidina).

La predicción de giros β en la cadena de GPA (figura 3-8) asigna la máxima probabilidad a los residuos TPNK (52-55), en donde se localiza uno de los cinco sitios de glicosilación, y probabilidades altas de encontrarse en esta conformación a los segmentos 38-41, 68-71, 83-86, 92-95, 106-109, 117-120, 130-133 y 151-154. Estas regiones parecen estar en el exterior de la molécula, a juzgar por los perfiles de hidrofobicidad y predicciones superficiales de las figuras 3-5 y 3-6. La asignación general de estructuras secundarias en la GPA, se asemeja a lo propuesto hace más de 15 años (Aubert y Locheux-Lefebvre,

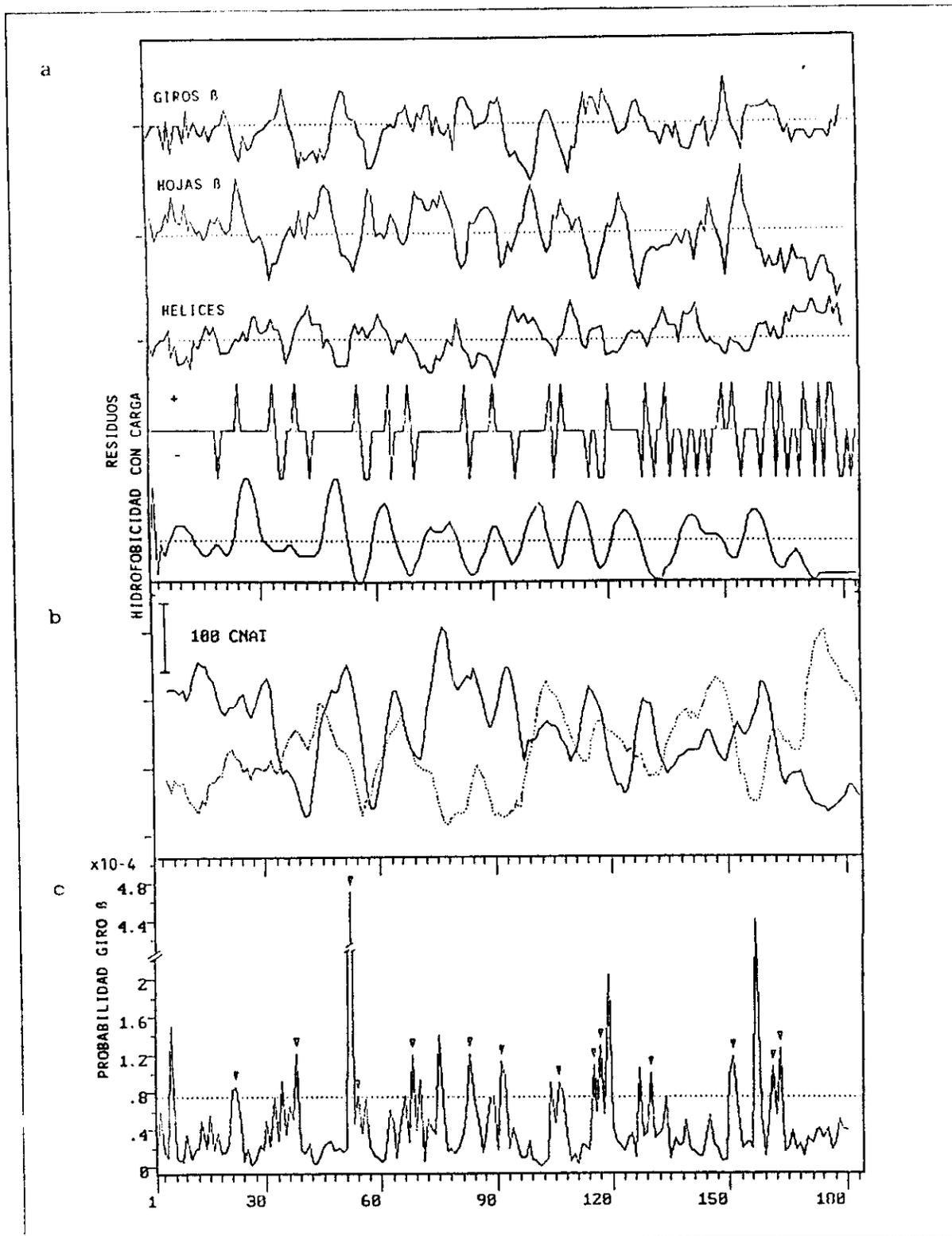


FIGURA 3-8. Predicción de estructuras secundarias en la GPA. a) Perfiles de propensión, posición de residuos cargados e hidrofobicidad según Novotny *et al.* b) Método de Garnier *et al.*, perfiles de propensión de hojas B (—) y hélices (---). c) Probabilidad de giros B.

asemejan a los aminoácidos del extremo amino de la cadena ligera tipo κ de la inmunoglobulina G (Schmid *et al.*, 1973), zona que contiene parte del segmento variable y una región hipervariable (Davies y Padlan, 1975). Los residuos del 77 al 125 de la GPA son homólogos con una región del tercer dominio constante de la inmunoglobulina H (Ikenaka *et al.*, 1972), y los residuos 112 a 132 con la cadena α de la haptoglobina (Ikenaka *et al.*, 1972; Schmid, 1972). Además, se ha reportado semejanza de la GPA con los dos dominios extracelulares del receptor del factor de crecimiento epidérmico (Toh *et al.*, 1985); algunos de estos receptores han mostrado tener homología con los miembros de la familia de las inmunoglobulinas. Por otra parte, hace alrededor de 4 años, se encontró semejanza entre el patrón de plegamiento propuesto para esta glicoproteína (Rojo-Domínguez, 1988) y las proteínas transportadora de retinol y β -lactoglobulina. Tal similitud fue apoyada al comparar las secuencias de estas macromoléculas y encontrar algunas zonas con homología. En esas fechas, otros investigadores empezaron a encontrar de manera independiente, diferentes secuencias de proteínas que también poseían semejanza en estructura primaria con las que se acaban de mencionar. Pervaiz y Brew (1987) publicaron un trabajo donde no sólo incluían a la GPA en este grupo, sino que además asignaron a este conjunto de proteínas el nombre de lipocalinas. Posteriormente, se conoció por difracción de rayos X la estructura de otras dos proteínas de la familia; dos insecticianinas.

Homología con las lipocalinas. Al comparar la secuencia de la GPA con las secuencias conocidas, se observa que esta glicoproteína posee cerca del extremo amino la *huella digital*, o secuencia característica de las lipocalinas: [DENST]-X-X-X-[LIVFY]-X-G-X-W-[FYWRH]- X-[LIVM]. La relación entre la GPA y las lipocalinas había sido propuesta anteriormente (Rojo-Domínguez, 1988; Pervaiz y Brew, 1987). Los miembros de esta familia poseen una sola cadena polipeptídica con un número de aminoácidos entre 160 y 190, y la capacidad de unir ligandos hidrofóbicos en su interior. Actualmente, se conocen más de cuarenta secuencias de lipocalinas (Nagata *et al.*, 1991; Boguski, 1992), sin embargo, se ha determinado la estructura cristalina para sólo cuatro de ellas: la transportadora de retinol (TRR) (Cowan *et al.*, 1990), la β -lactoglobulina (BLG) (Mónaco, *et al.*, 1987), la insecticianina (INS) (Holden *et al.*, 1987) y la proteína transportadora de bilina (PTB) (Huber *et al.*, 1987a,b). Al comparar las estructuras de estas proteínas, se observa que a pesar de la baja homología en secuencia, su arquitectura se conserva: un

barril (o cáliz) formado por ocho cadenas β antiparalelas y una hélice en su extremo carboxilo (figura 3-10). En el presente trabajo, las cadenas β serán denominadas en los sucesivos con letras mayúsculas, de la A a la H, según el orden de aparición en la secuencia de aminoácidos. La tabla 3-IV contiene un resumen de las características de algunos miembros de esta familia. Se ha sugerido que el estudio de la variabilidad en las secuencias de los miembros de una familia homóloga aporta mayor información que el análisis en pares de ellas (States y Boguski, 1991). Sin embargo, en el caso de las lipocalinas, esto es particularmente difícil debido a la poca conservación de patrones característicos.

Aunque la GPA posee la secuencia característica de las lipocalinas, así como un patrón de enlaces disulfuro similar, la capacidad de unir ligandos hidrofóbicos, y un número de aminoácidos semejantes a las de las lipocalinas conocidas, su homología se encuentra en el umbral o incluso, ligeramente por debajo de lo que se consideraría estadísticamente significativo. Sin embargo, dada la conocida degeneración del código de doblado para esta familia, y la serie de argumentos experimentales mencionados previamente, se puede afirmar que la GPA pertenece a este conjunto de proteínas y que conserva el patrón estructural determinado por difracción de rayos X en cuatro de las lipocalinas. Uno de estos argumentos experimentales se mencionó en el capítulo anterior, y consiste en la semejanza en estructura secundaria entre la GPA y una de las lipocalinas con estructura tridimensional conocida: la β -lactoglobulina. Otro fundamento de gran peso en favor de la estrecha relación entre la GPA y las lipocalinas, es que los genes que las codifican poseen la misma organización (Girotti, 1992); esto es, la estructura de exones e intrones en los genes de lipocalinas caracterizadas es idéntico a la de los genes de la glicoproteína ácida α_1 , tanto de humano (Dente *et al.*, 1987) como de rata (Liao *et al.*, 1985; Reinke y Feigelson, 1985) y de conejo (Ray y Ray, 1992). Esta evidencia permite suponer que la estructura tridimensional de la GPA puede ser modelada tomando como punto de partida la arquitectura de una lipocalina, efectuando un alineamiento apropiado entre las secuencias de aminoácidos de ambas proteínas.

Análisis topológico entre las lipocalinas y las inmunoglobulinas. Es conocido que la arquitectura de los miembros de la familia de las lipocalinas, al igual que la de los dominios de las inmunoglobulinas, está basada en dos juegos de hojas- β antiparalelas colocadas una sobre la otra (figura 3-10).

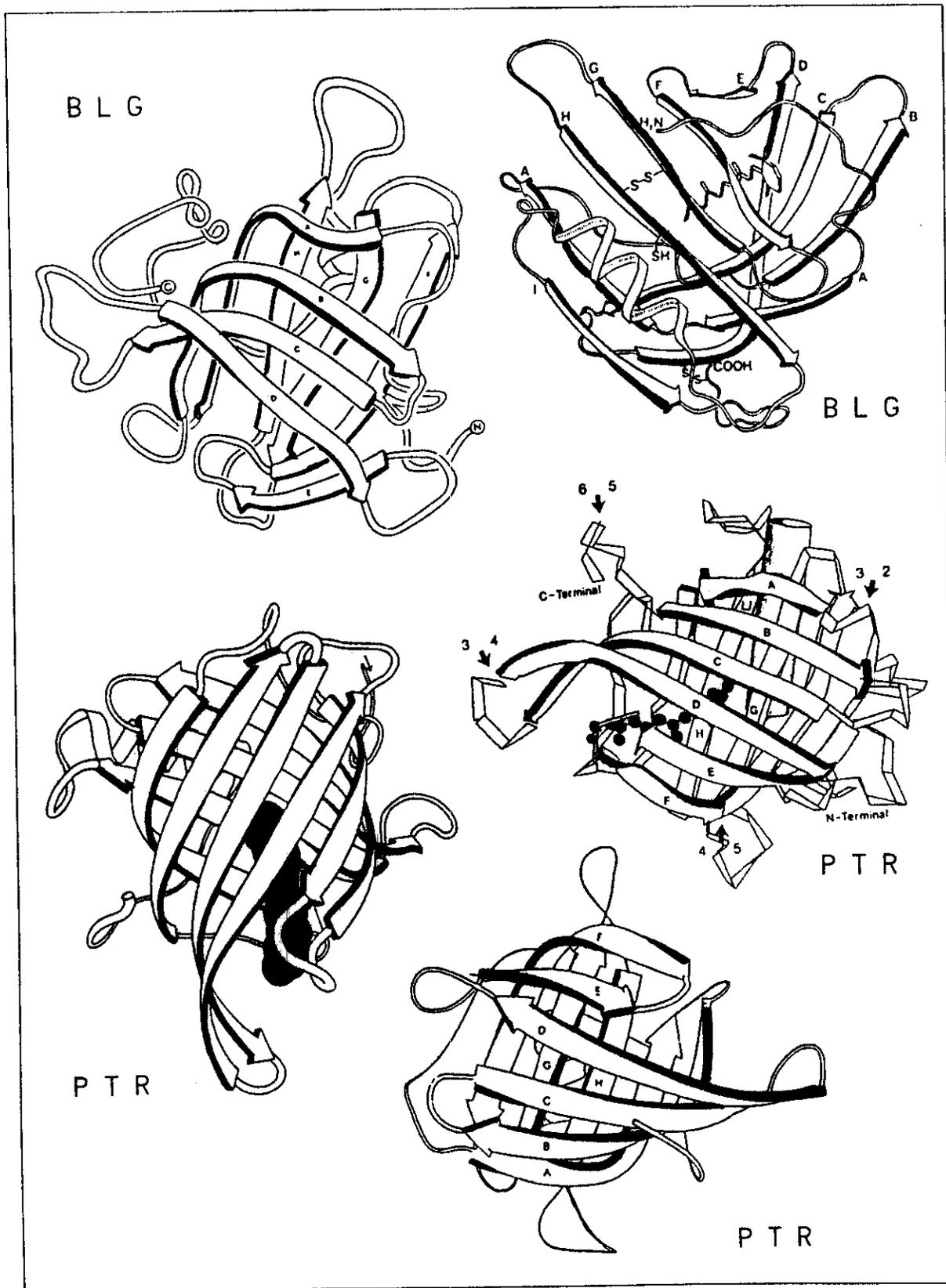


FIGURA 3-10. Esquemas de las estructuras tridimensionales de las lipocalinas

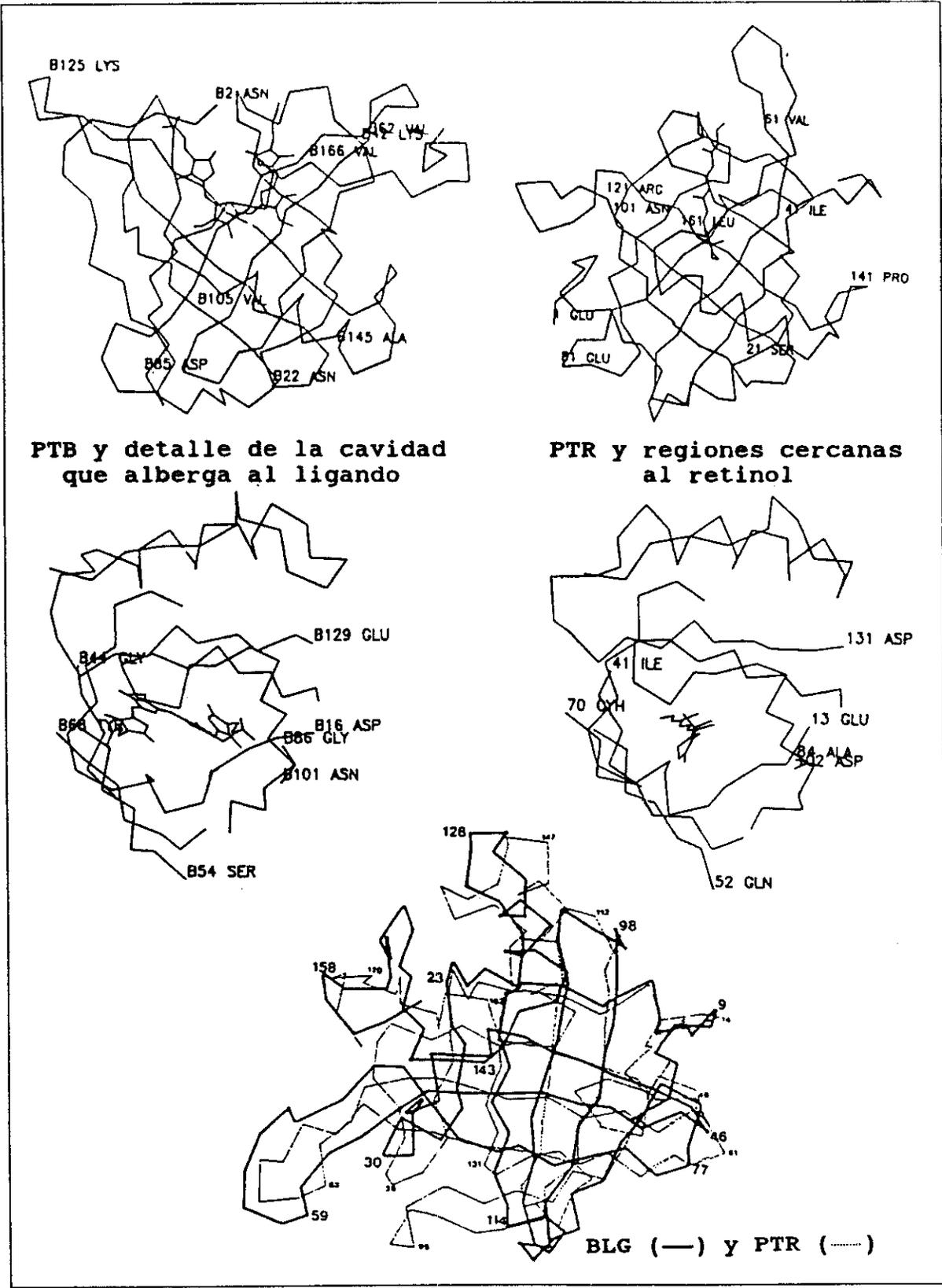


FIGURA 3-10. Esquemas de las estructuras tridimensionales de las lipocalinas (continuación).

T A B L A 3 - IV

Propiedades físicas de algunas lipocalinas conocidas.

Proteína	M M	#	Fuente	Ligando
Glicoproteína ácida α_1	21560	183	Plasma humano	fármacos, esteroides
Receptora de olor	18212	158	Epitelio nasal, rata	moléculas fragantes
Microglobulina α_2	18472	160	Hígado, rata	•
Principal urinaria	18437	160	Hígado, ratón	*
de Bowman	18661	164	Epitelio nasal, rana	moléculas fragantes
Transportadora de retinol	20957	182	Hígado humano	retinol
Apolipoproteína D	19303	169	Plasma humano	colesterol
Insecticianina	21379	189	Hemolinfa, insecto	biliverdina IX
Transportadora de bilina	19676	173	Hemolinfa, insecto	pigmentos
Andrógeno dependiente	18480	166	Epidídimo, rata	*
Microglobulina α_1	20433	181	Plasma y orina humanos	retinol
β -Lactoglobulina	18281	162	Suero de leche bovina	retinol
Purpurina	19538	175	Retina, pollo	retinol, heparina
P14	18784	162	Placenta, humano	•
Afrodísina	17232	151	Hamster	feromonas
Cadena γ del complemento 8	20327	182	Plasma humano	•

M M, masa molecular; #, número de aminoácidos; *, aún no determinado

Las inmunoglobulinas son un conjunto de proteínas con una gran variabilidad en su secuencia de aminoácidos. Se acepta que el número de mutaciones puntuales por cada 100 residuos en un periodo de 10 millones de años es mayor de la centena (Wilson *et al.*, 1977); en el conjunto de proteínas analizado, esta mutabilidad es alcanzada sólo por las toxinas de veneno de serpiente, y son seguidas muy de lejos por otras proteínas. La estructura básica de una molécula de inmunoglobulina es una Y, compuesta de cuatro cadenas, dos ligeras y dos pesadas, unidas entre sí por puentes disulfuro. En realidad, la estructura compleja de las moléculas de inmunoglobulina puede simplificarse al notar que está formada por 12 dominios estructurales, prácticamente idénticos, cada uno de ellos de alrededor de 110 residuos. Cuatro dominios constituyen las cadenas pesadas y dos las ligeras. Cada dominio se integra por dos capas de siete hojas β antiparalelas, una sobre otra, con un ángulo entre las direcciones de las cadenas β de aproximadamente 30° (Davis *et al.*, 1975). Una secuencia conservada en estos dominios es la zona cercana al puente disulfuro, la cual consiste de los siguientes aminoácidos: [FY]-X-C-X-[VA]-X-H, en donde X representa cualquier residuo. La conectividad (Richardson, 1977) de las cadenas en las hojas β es: +1, +3, -1, -1, +3, +1.

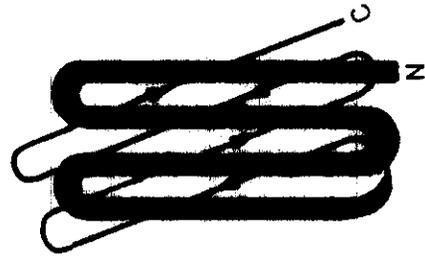
Por otro lado, las lipocalinas fueron identificadas como una familia de proteínas hace apenas un lustro (Pervaiz y Brew, 1987), pues aunque se conocían algunas de las secuencias de aminoácidos de sus miembros, éstas poseen tan baja homología, que fueron relacionadas entre sí hasta que sus estructuras tridimensionales se determinaron por difracción de rayos X y pudo compararse el patrón general de doblado, prácticamente idéntico, que ellas poseen. Como se mencionó previamente, se conocen más de 40 secuencias de lipocalinas con la secuencia [DENST]-X-X-X-[LIVFY]-X-G-X-W-[FYWRH]-X-[LIVM] cercana al amino terminal. En la figura 3-10 el patrón general de doblado que estas proteínas mantienen, a pesar de tener una homología de apenas un 15%. En la misma figura podemos observar que dos hojas β antiparalelas (cada una formada por cuatro cadenas, A-D y E-H) se colocan una sobre la otra con la dirección de sus cadenas casi perpendiculares y dejando un espacio vacío entre sí. En este espacio se aloja una molécula hidrofóbica (representada como un sólido negro); debido a esta característica (la formación de un cáliz que alberga a un ligando lipídico) se dio el nombre de lipocalinas a la familia. Adicionalmente, la estructura tridimensional de estas macromoléculas posee un segmento de hélice α y, en algunas moléculas, una novena cadena en

conformación β , ambas en el extremo carboxilo. La conectividad de las cadenas que forman el cáliz en las lipocalinas es (+1)7, es decir, cada cadena en las hojas β se encuentra formando puentes de hidrógeno con las cadenas inmediatas en la secuencia.

Aunque las dos familias de proteínas mencionadas en el párrafo anterior realizan funciones biológicas diferentes, tienen en común un código de doblado extremadamente degenerado, causante de la gran variabilidad en las secuencias de aminoácidos que mantienen el patrón estructural tridimensional. Además, los miembros de ambos conjuntos comparten una arquitectura basada fundamentalmente en dos juegos de hojas β cuyos planos son aproximadamente paralelos. Existen además otras dos relaciones entre ellas: una lipocalina (precisamente la GPA) con homología parcial con algunas inmunoglobulinas, y una estrecha relación topológica en las conectividades de las cadenas β de ambas familias.

La figura 3-11 muestra esquemáticamente, en una serie de etapas, cómo puede cambiarse la topología de las lipocalinas (3-11a) a la de las inmunoglobulinas (3-11d). En la parte superior de estos diagramas son representadas con trazo grueso las cadenas en la hoja β superior y con trazo delgado las correspondientes a la hoja β inferior. Las flechas indican la dirección de la cadena del extremo amino (N) al carboxilo terminal (C). Dado que las inmunoglobulinas poseen sólo siete cadenas, se muestra con línea punteada la octava. En la parte inferior se muestra una perspectiva donde las cadenas que forman las hojas β son perpendiculares al plano del papel. Los círculos representan la dirección hacia dentro del plano y los triángulos hacia fuera. La conectividad de las cadenas se representa con líneas gruesas y delgadas cuando se encuentran cercanas o alejadas del observador, respectivamente, mientras que los números muestran la posición consecutiva de las cadenas en la secuencia de aminoácidos. La serie de movimientos implicados para pasar de una topología a otra invierte la posición de las cadenas laterales internas y externas en las porciones de hojas β que sufren desplazamiento. Por esta razón, no es probable que una proteína perteneciente a una de estas familias pueda cambiar a la arquitectura de la otra familia. Con la información que actualmente se dispone, es difícil establecer si la relación topológica entre estas familias de proteínas es producto de la evolución convergente o divergente. Las lipocalinas son codificadas por seis exones, pero sólo uno (o en ocasiones dos) codifica para cada dominio estructural de las inmunoglobulinas. La homología entre los miembros de una misma familia es

LIPOCALINAS



INMUNOGLOBULINAS

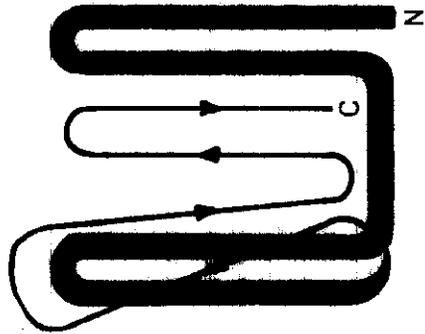
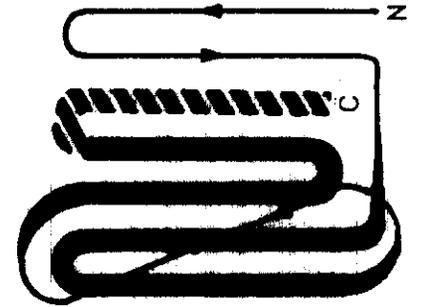
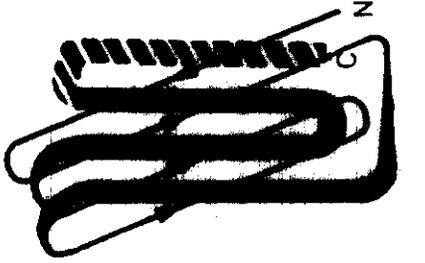


FIGURA 3-11. Relación topológica entre las conectividades de cadenas- β de las lipocalinas y de las inmunoglobulinas.

frecuentemente escasa e incluso no significativa. La evolución y las mutaciones han borrado la huella de un eslabón entre ambos conjuntos de macromoléculas, si es que éste existió en algún momento. De cualquier forma, la glicoproteína ácida α_1 podría ser la proteína conocida más cercana a este eslabón, aunque la homología con las inmunoglobulinas es sólo local.

Homología con otras proteínas. Ha sido establecida la homología de porciones de la secuencia de aminoácidos de la GPA con otras proteínas, además de las ya mencionadas con las lipocalinas e inmunoglobulinas. Por ejemplo, se ha señalado que los dos dominios extracelulares del receptor del factor de crecimiento epidérmico (los cuales guardan homología entre sí), presentan homología con la estructura primaria de la glicoproteína (Toh *et al.*, 1985). Puede observarse en la figura 3-9e que esta homología se mantiene en una región amplia de la secuencia de la GPA; esta región incluye, como será descrito en la siguiente sección, las ocho regiones topológicamente equivalentes a las cadenas β de las lipocalinas y a la hélice I. Desafortunadamente, se desconoce la estructura tridimensional de los dominios extracelulares de ese receptor para compararla con el modelo de la GPA o con la arquitectura de las lipocalinas; pero la conformación de otros receptores de factores de crecimiento muestra la presencia de dominios formados de hojas β antiparalelas, semejantes a las observadas en las lipocalinas o en los dominios de las inmunoglobulinas (DeVos *et al.*, 1992). Esto hace pensar que las lipocalinas padrían no sólo ser una familia de proteínas, sino que su topología pudiera constituir un módulo estructural, del tipo que funciona como bloque en la construcción de la geometría tridimensional de proteínas de alta masa molecular (Baron *et al.*, 1991).

Por otra parte, se ha encontrado homología entre el segmento 106-132 de la GPA y una zona de la cadena α de la haptoglobina (HAP) (figura 3-9f) (Ikenaka *et al.*, 1972). Esta región de la glicoproteína coincide casi exactamente con el exón número cuatro del gen que la codifica, sugiriendo que ambos segmentos pudieron tener un origen común. Sería muy interesante analizar la estructura del gen que codifica a la cadena α de la haptoglobina, buscando que el segmento homólogo con la GPA realmente coincidiera con un exón. Hasta donde sabemos, aún no se tiene esta información.

ALINEAMIENTO.

Dado el número tan escaso de aminoácidos idénticos entre la GPA y los otros miembros de las lipocalinas, el alineamiento presenta gran dificultad. Para seleccionar la secuencia, con estructura tridimensional conocida, que se usaría como molde para modelar a la GPA, se aplicó el método de Needleman y Wunsch (1970) para determinar que tan significativa es la homología de esta secuencia con cada una de las cuatro lipocalinas cristalizadas. Con el objeto de determinar los parámetros más favorables para las comparaciones de las secuencias de aminoácidos, se muestra en la figura 3-12a la dependencia del valor de significación (en número de desviaciones estándar) con las magnitudes de la penalización por inserciones (P) y la tendencia de alineamiento (T) para la comparación de 100 secuencias aleatorias. Los valores elegidos fueron $P = 30$ y $T = 110$. Durante la comparación de las secuencias, se observó que un aumento en el número de comparaciones con secuencias aleatoriamente generadas, disminuía la variabilidad en el valor de significación, pero también su magnitud. En la figura 3-12b se muestran estas observaciones para la comparación de la secuencia de la GPA con las cuatro lipocalinas con estructura cristalina conocida; puede notarse que el valor de significación (SIG) decrece de una manera exponencial con el número de comparaciones. Las barras en la figura denotan la variabilidad en SIG al repetir cada cálculo cinco veces, y los puntos corresponden al promedio de estos valores. Las barras de variabilidad en el valor de significación decrecen en magnitud hacia la derecha del gráfico, indicativo de que estos valores son más precisos al aumentar el número de muestras en el análisis estadístico. De cualquier manera, la relevancia de la homología entre la GPA y la proteína transportadora de bilina se mantiene sistemáticamente por encima de las otras lipocalinas. Valores de 3.1, 4.3, 4.7 y 5.2 desviaciones estándar equivalen a probabilidades de homología accidental de una en mil, una en cien mil, una en un millón y una en diez millones, respectivamente. Fue elegida como molde, o estructura inicial, la proteína transportadora de bilina (PTB) por ser la que mostró la homología más significativa, y una de las que fueron clasificadas como de la clase $\alpha+\beta$ por el método de Nakashima *et al.* (1986).

La aplicación directa de los métodos de alineamiento tradicionales arrojó resultados ambiguos. Por esta razón, se decidió alinear las secuencias anclándolas en diversos sitios y efectuando el alineamiento en cada segmento por separado. Los criterios para el anclaje fueron los siguientes:

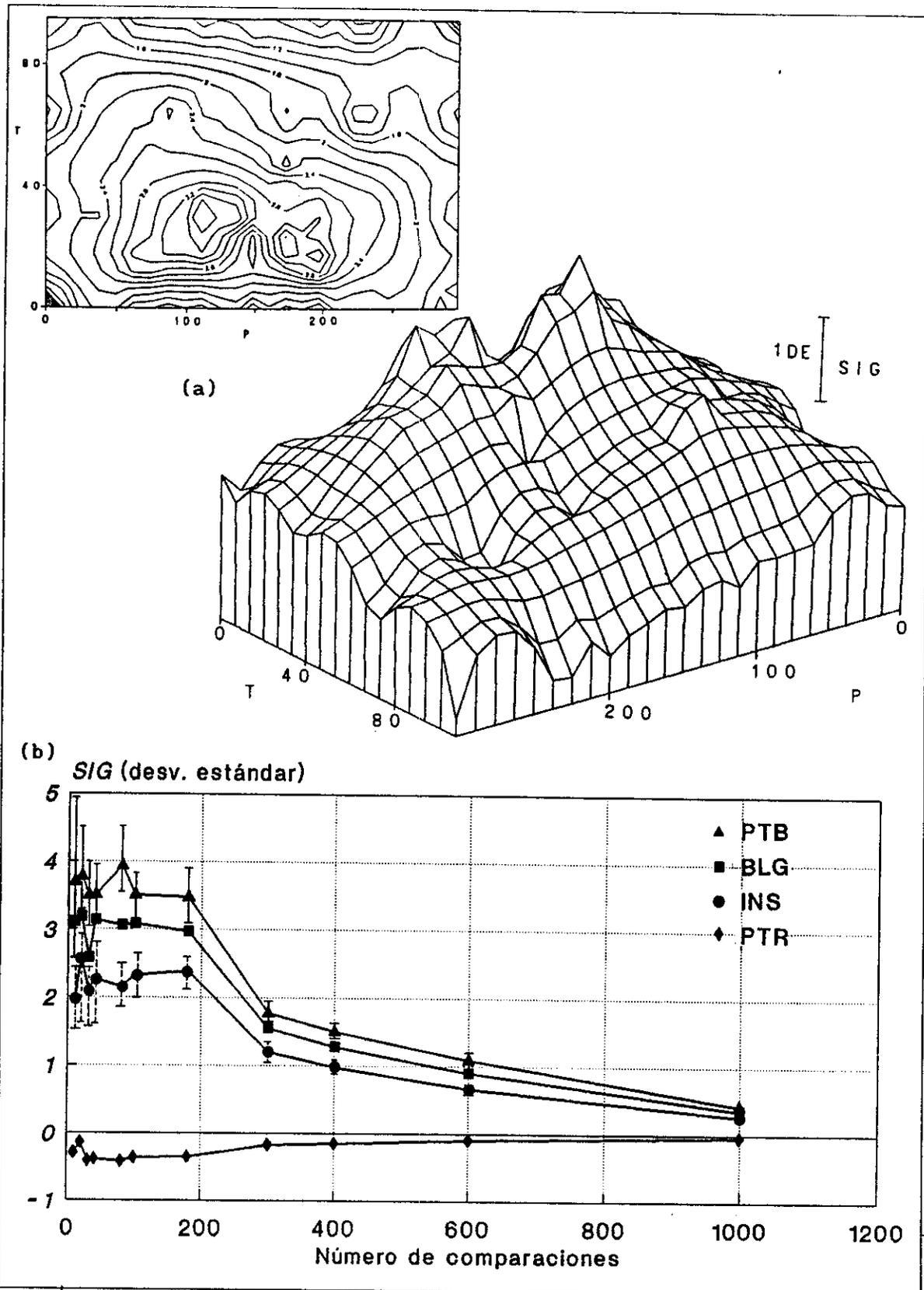


FIGURA 3-12. Dependencia de la significación con los parámetros de inserción (a) y el número de comparaciones (b).

- a) Posición de las dos regiones de mayor homología entre la GPA y la PTB (residuos 23-29 y 105-112)
- b) Posición de los residuos de cisteína conservados: C5 y C165.
- c) Localización de los intrones en el gen que codifica a la GPA. Se sabe que las posiciones de éstos se conservan en todos los genes conocidos de lipocalinas. Dado que no se ha determinado la secuencia del gen que codifica a la PTB, fue necesario identificar las posiciones de su cadena polipeptídica equivalentes con los sitios de inserción de intrones de otras lipocalinas. Estas posiciones se encuentran entre los aminoácidos 20-21 y 162-163 y en los residuos 68, 91 y 128.
- d) La coincidencia de regiones en la GPA con alta probabilidad de adoptar estructuras secundarias o de ocupar regiones expuestas al solvente con las zonas de la PTB que presentan esta característica.
- e) La comparación de las estructuras tridimensionales de las otras tres lipocalinas, buscando las regiones de mayor y menor conservación de las características estructurales.

De esta forma, la secuencia de la GPA fue anclada en siete regiones; éstas fueron:

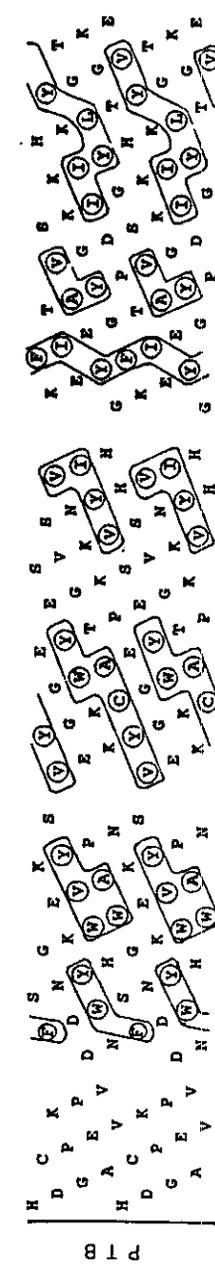
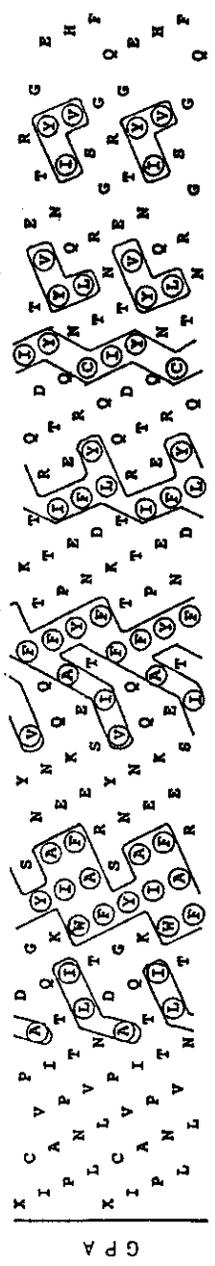
- a) La C5 por ser un aminoácido conservado involucrado en la formación de puentes disulfuro y en una zona de homología C-X-[NE]-[LV]-P-V.
- b) La G23 parte de la secuencia característica de las lipocalinas y cercana a la posición del primer intrón.
- c) La N54, que une carbohidratos, debe encontrarse en una región externa. Esta es la zona con mayor probabilidad de formar un giro β en la secuencia de la GPA, por lo que se ancló con el giro β más cercano de la PTB.
- d) El D106, perteneciente a una de las regiones conservadas en las lipocalinas.
- e) La S125, presente en una zona conservada W-X-L-S y cercana a un sitio de inserción de intrón.
- f) La C165, involucrada en un puente disulfuro y cerca de un sitio de unión de exones.

Los segmentos entre estos sitios de anclaje fueron alineados empleando los algoritmos de Myers y Miller (1988) y Higgins y Sharp (1989) con diferentes parámetros en cada región, y con un ajuste final a *mano*, considerando la exposición necesaria de los cinco residuos de asparagina que unen carbohidratos y que por lo tanto deben encontrarse en regiones expuestas al

solvente acuoso. El alineamiento final (figura 3-13) muestra 26 residuos idénticos en estas proteínas (14.2%) y 25 semejantes (13.7%), con 4 inserciones en la secuencia de cada proteína. Estas inserciones están localizadas en posiciones de la cadena que corresponden a giros- β o a estructuras no repetitivas, a excepción del que se observa entre los aminoácidos 103 y 104, el cual corresponde a la distorsión de una cadena β (*β bulge*). En esta figura (3-13) puede apreciarse que los últimos 15 aminoácidos de la GPA no tienen equivalente en la PTB. Sin embargo, basándose en la tendencia de este segmento a formar estructuras secundarias, es posible que se encuentre en conformación helicoidal, aunque es difícil precisar los límites de esta estructura. En adelante se denominará a esta región *hélice II*, y a los residuos 132-145 *hélice I*. Estas dos zonas se representan en la figura 3-14, en donde se muestra las proyecciones de las cadenas laterales a lo largo del eje de las hélices propuestas. Puede observarse que la hélice I tiene un lado hidrofóbico, con dos residuos de L y uno de Y, F y A, y otro hidrofílico en donde se localizan cuatro residuos de ácido glutámico y uno de K y T. En contraste, la hélice II, con una gran cantidad de aminoácidos con carga eléctrica (seis residuos de E, tres de K y uno de R) y polares (dos residuos de Q, uno de H y uno de S), sólo posee un residuo hidrofóbico, la L168. Estas características de la hélice II y la falta de una conformación equivalente en la PTB hicieron imposible el modelado de esta región. Se espera que la hélice I exponga su lado polar al solvente, orientando el otro hacia el resto de la molécula; en el caso de la otra hélice, es difícil proponer una posible orientación, aunque la gran cantidad de residuos con carga eléctrica negativa la hacen un buen sitio de interacción de los fármacos básicos que une la GPA.

Al analizar los núcleos hidrofóbicos de las dos secuencias alineadas (figura 3-13) con el propósito de observar la correspondencia en la localización de las estructuras secundarias en ambas proteínas, se encuentra que los núcleos corresponden adecuadamente a segmentos de la cadena en conformación de cadenas β o hélices. Además, resulta que los núcleos hidrofóbicos ocurren en regiones equivalentes en las dos cadenas polipeptídicas, lo que resulta un argumento adicional en favor de que el alineamiento correlaciona de manera apropiada las estructuras secundarias de ambas macromoléculas. Un detalle más de importancia relevante es que las cadenas β C, F y H, que no fueron detectadas concluyentemente en la predicción de estructura secundaria de la GPA, aparecen claramente delimitadas en el patrón de núcleos hidrofóbicos.

1 2 3 4 5 6 7 8 9
 GPA XIPLCANLVPPIAATLDDITGCKWFIASAFRNEEYKNSVQEIQAATYFFPFRNKTEDLFLREYQTRDQDCIYNTIYLNQVREN GTISRYVGG QEHF R V
 PTB HDGACPEKPV DNFDWSNHYGKWEVAKYRNS VERYKCGWAEYTPGKSVKSNHYHVR GREYFIETGATVPVGDGSKIGKIYKLLTYGGVTK E



10 11 12 13 14 15 16 17 18
 GPA AHLIL RDTKTHLAFDIN LFGSYLD DENWGLSVADKPTTEKQELGFEYALDCLIRKSDVYVDKDKKPELEKQHERERQEGES
 PTB MFWNLSTDNKNIIGYCKRYDEDKRQHGDQVWVLSRSKVLGTEAKTAVENYLGISPT VDSQKLVYDFSEACNVR

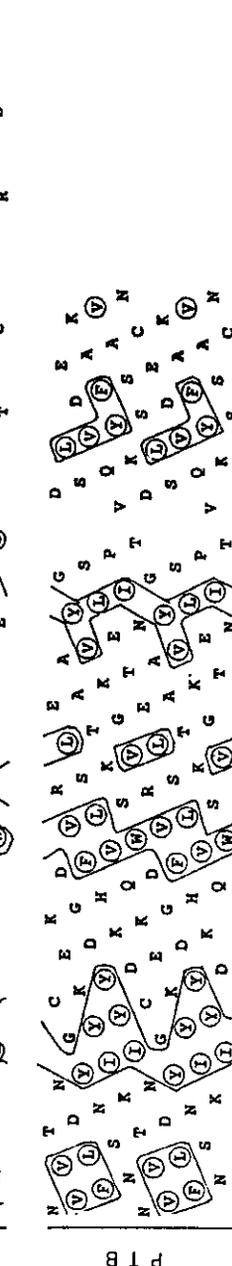
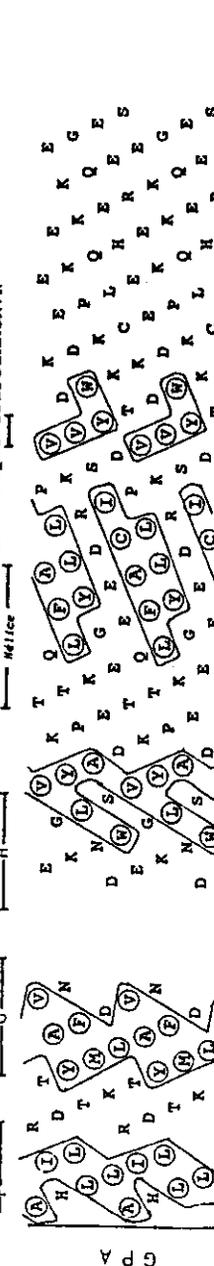


FIGURA 3.13.
 Alineamiento entre la GPA y la PTB. Se muestran además la correspondencia de los núcleos hidrofóbicos, los reemplazos de aminoácidos y la posición de los residuos de N que unen cadenas de carbohidratos (*). Las barras horizontales indican la posición de los elementos de estructura secundaria en ambas proteínas.

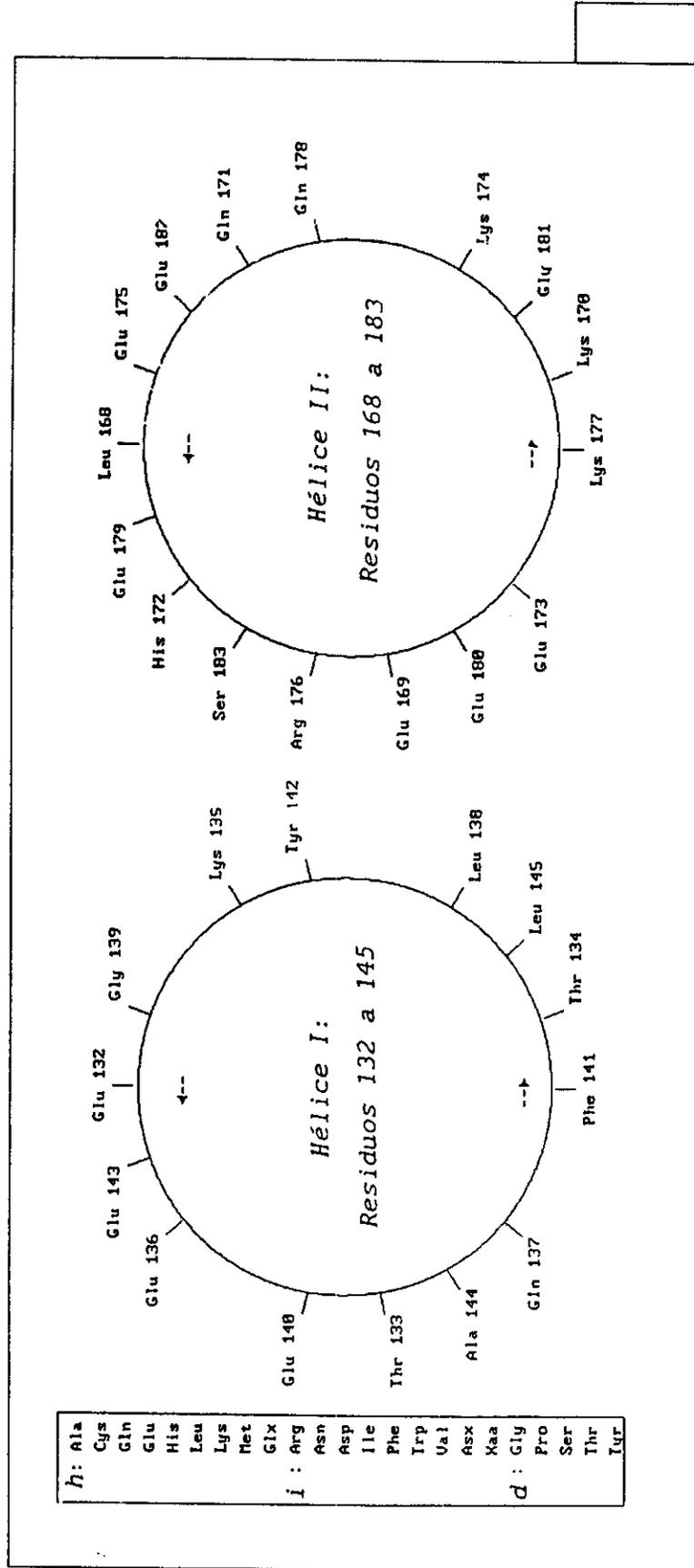


FIGURA 3-14. Proyecciones helicoidales de dos segmentos de la GPA. El cuadro de la izquierda es la clasificación de los aminoácidos en helicoidales (*h*), indiferentes (*i*) y desestabilizantes (*d*) según Levitt (1978).

CONCLUSIONES DEL ANALISIS DE SECUENCIA.

A juzgar de la información obtenida de la secuenciación de los tres genes que codifican a la GPA y de los genes correspondientes a la glicoproteína ácida α_1 de otros mamíferos, puede asegurarse que la parte proteica de de esta macromolécula está integrada por 183 residuos de aminoácidos y no por 181, como se había reportado previamente (Schmid, 1975). La GPA se caracteriza por ser una molécula de alta hidrofiliidad, con cinco cadenas de carbohidratos ricos en ácido siálico en su primera mitad y un número alto de residuos de aminoácido polares y con carga eléctrica en la segunda. Esta gran concentración de aminoácidos hidrofílicos en la porción final de la cadena se manifiesta en la composición global de aminoácidos; en ella también destacan un mayor número de aminoácidos aromáticos y menor de alifáticos respecto a la composición promedio de otras proteínas.

La clasificación estructural de la GPA con base en su secuencia le asigna la clase $\alpha+\beta$, tanto a la cadena completa como al segmento 1-130 (GP1), mientras que identifica a los últimos 55 aminoácidos (porción denominada GP2) como helicoidales. Cabe recordar que las lipocalinas pertenecen justamente a esta clase estructural y que la composición de aminoácidos de la PTB también la clasifica en $\alpha+\beta$. Al estudiar este resultado, simultáneamente con los del análisis de los espectros de dicroísmo circular de la GPA y la BLG, y con los de la predicción de estructuras secundarias con base en la secuencia de aminoácidos, se encuentra que las diferentes metodologías concuerdan con una estructura integrada principalmente por hojas β antiparalelas, y un contenido menor de hélices; estas últimas localizadas fundamentalmente en la porción final de la cadena.

El número y localización de las diferentes cadenas β y hélices en esta glicoproteína pueden estimarse a partir de la homología estructural que muestra con las lipocalinas, y en particular con la PTB, del análisis comparativo de núcleos hidrofóbicos en las secuencias de estas dos proteínas, el alineamiento entre ellas y la localización de las regiones externas, antigénicas y de alta probabilidad de encontrarse en giros β . Respecto a este último punto, es interesante resaltar el buen acuerdo de los resultados obtenidos con las diferentes metodologías aplicadas. Se señala consistentemente que las regiones cercanas a los aminoácidos 22, 34, 55, 68, 85, 95, 106, 119, 133, 152, 163 y 178 se encuentran en la superficie molecular, delimitando así la posición de hojas β y hélices.

Por supuesto que el empleo del alineamiento y comparación de núcleos hidrofóbicos entre la GPA y la PTB debe ser justificado. Sin embargo, la evidencia señala que la glicoproteína realmente pertenece a la familia de las lipocalinas, no sólo por la presencia de la huella digital en su secuencia, la conservación de la región cercana al aminoácido 108, los residuos de C conservados y el número de aminoácidos que la componen. Tal vez las evidencias más importantes para clasificar a la GPA como una lipocalina sea la semejanza de la estructura del gen que la codifica con las que se conocen de otras lipocalinas, aunada a que la homología en las secuencias de aminoácidos de la GPA y la PTB, aunque pequeña, tiene una significancia mayor a tres desviaciones estándar.

Toda la infraestructura mencionada anteriormente permite proponer un alineamiento confiable que relacione la mayor parte de los segmentos topológicamente equivalentes entre las secuencias de estas proteínas. Este resultado es fundamental para el modelado molecular de la GPA que se describe en el siguiente capítulo.

Una característica muy interesante de la familia de las lipocalinas es la gran degeneración de su código de doblado. Las secuencias de aminoácidos que son compatibles con la topología de esta familia son muy variadas. Destaca, por ejemplo, que las glicoproteínas ácidas α_1 de diferentes organismos presentan un número bajo de aminoácidos idénticos al compararlas por pares (entre 39 y 57%) y sólo 55 aminoácidos (29%) son invariantes en el conjunto de secuencias de la figura 3-7a. Más aún, las glicoproteínas conocidas de fuentes no humanas no conservan la G23, invariante en el resto de las lipocalinas. Al comparar la secuencia de la GPA con otras lipocalinas se hace patente su baja homología. Uno de los casos extremos es el alineamiento entre la GPA y la PTR, en donde la significancia es incluso negativa, es decir, que el promedio de las comparaciones aleatorias es mayor que el marcador de su alineamiento normal. Un resultado como éste no indica que ambas proteínas no tengan relación, señala que la homología entre ellas es demasiado baja para que sea posible concluir algo sobre su relación estructural.

A pesar de la gran variabilidad de estas secuencias, es posible encontrar homología interna en la GPA y homología de ésta con diversas proteínas. Aparentemente, un denominador común en todas las estructuras involucradas es la presencia de hojas β antiparalelas. Podríamos estar estudiando un conjunto

de proteínas, formado por varias familias, en cuyos miembros exista una región de su secuencia de aminoácidos que induzca la formación de ese tipo de estructura secundaria. La homología de esta zona entre las diferentes proteínas que la poseen ha sido oscurecida por las mutaciones y la evolución, evitando poder reconocer su localización, proponer una secuencia de consenso o identificar las propiedades fisicoquímicas que retiene.

Las relaciones de homología de la GPA con otras proteínas, abren un campo de estudio futuro muy interesante en donde se buscaría respuesta a preguntas tales como: ¿qué características se conservan en todas las sumamente variables secuencias de aminoácidos, que mantienen la arquitectura general basada en juegos de hojas- β ? ¿Cuáles son las propiedades de estas secuencias que dirigen el plegamiento de la cadena polipeptídica hacia una topología particular? ¿Observamos en todos los casos descritos anteriormente la convergencia evolutiva hacia un patrón estructural estable o conveniente? o bien, ¿las proteínas, dominios y familias mencionadas son producto de una macromolécula ancestral común? No es posible responder a estas preguntas con la información disponible hasta el momento. De cualquier manera, los estudios estructurales en familias de proteínas han probado ser una fuente muy abundante de información sobre la compleja arquitectura de estas macromoléculas (Petsko, 1991) y, en el futuro, todo hace suponer que el análisis comparativo de diferentes conjuntos de proteínas nos proporcionará nueva información para responder algunas de las preguntas (y generar nuevas) que hasta ahora se han acumulado.

CAPITULO 4.

MODELADO MOLECULAR.

MÉTODOS.

Debido a la dificultad para cristalizar a la glicoproteína ácida α_1 y para obtener la información detallada sobre su estructura por algún otro medio (por ejemplo, utilizando la resonancia magnética nuclear), la predicción de la estructura tridimensional de esta macromolécula es de gran interés. El modelo de la GPA contribuiría a interpretar algunas observaciones experimentales, pero principalmente ayudaría en el estudio de su, hasta ahora desconocida, función biológica; por ejemplo, podría permitir analizar el posible sitio de unión de ligandos hidrofóbicos, la localización de los reemplazos de aminoácidos y la distribución de las cadenas de carbohidratos en su superficie molecular.

Para construir el modelo molecular de la variante A de la GPA, se utilizó una estación de trabajo IBM RISC 6000 modelo 320 con el paquete de cómputo BIOGRAF (Molecular Simulations, Inc., Waltham MA, E.U.A.). Para los cálculos de interacciones atómicas se usó un conjunto de parámetros, desarrollado específicamente para este paquete de cómputo, denominado *Dreiding* (Mayo et al., 1990). Esta parametrización incluye específicamente los siguientes tipos de interacción:

a) Distancia de enlaces. Se estima la dependencia de la energía potencial debida a fluctuaciones en la distancia de enlace con un potencial de tipo armónico:

$$E_{enl} = \frac{1}{2} \cdot K_b \cdot (R - R_0)^2 \quad (4-1)$$

en donde E_{enl} representa la contribución a la energía potencial total de los cambios en las distancias entre átomos enlazados, esto es, la energía que aleja a los átomos de su posición de equilibrio; K_b es la constante de fuerzas en unidades de $kcal \cdot mol^{-1} \cdot \text{Å}^{-2}$; R_0 la distancia de equilibrio entre los átomos que forman el enlace, en Å ; y R la distancia real entre los átomos, en Å .

b) Angulo de enlace. Este tipo de energía potencial cambia al modificarse el ángulo IJK en donde J está enlazado simultáneamente con I y con K (figura 4-1a). La expresión de esta energía, E_θ , es al igual que la anterior de tipo armónico:

$$E_\theta = \frac{1}{2} \cdot K_\theta \cdot (\theta - \theta_0)^2 \quad (4-2)$$

donde K_θ es, nuevamente, la constante de fuerzas; y θ y θ_0 son los valores real y de equilibrio del ángulo IJK.

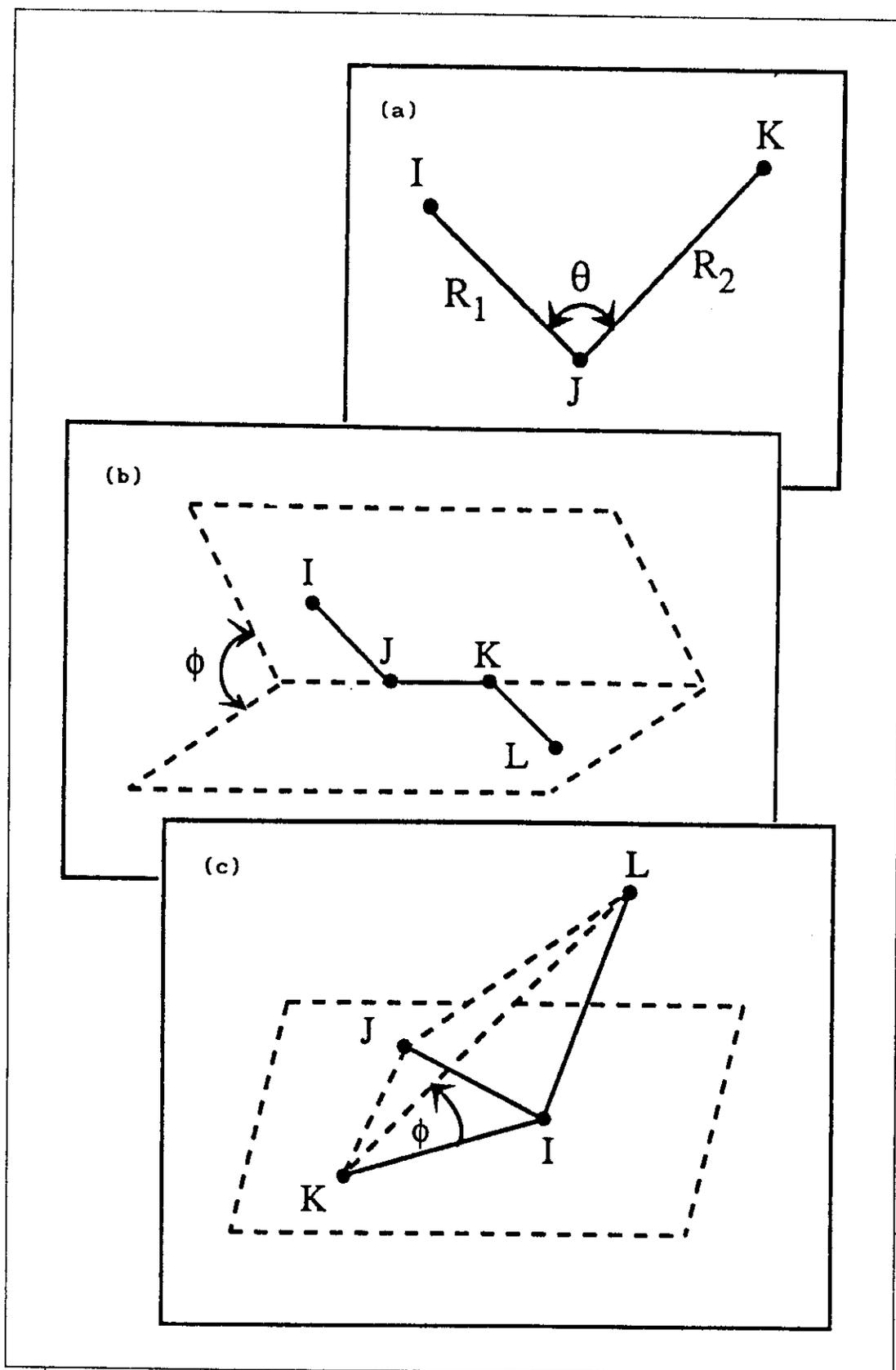


FIGURA 4-1. Definición de variables para estimar la energía potencial de los ángulos de enlace (a), de los ángulos dihédricos (b) y de los centros de inversión (c).

c) Torsión de ángulos dihédricos. Se define el ángulo ϕ (figura 4-1b), cuando dos enlaces IJ y KL están unidos a un enlace común JK, como el formado por los planos IJK y JKL. El valor de ϕ es positivo si, mirando sobre J hacia K, este ángulo gira en el sentido de las manecillas del reloj. De esta forma, $\phi = 0^\circ$ en la conformación *cis* y $\phi = 180^\circ$ en la *trans*. La energía torsional se define como:

$$E_\phi = \sum_{n=1}^6 \frac{1}{2} \cdot K_{\phi,n} \cdot [1 - d \cdot \cos(n\phi)] \quad (4-3)$$

$K_{\phi,n}$ es la barrera rotacional, n la periodicidad del potencial y $d = \pm 1$ el factor de fase. Un valor de $d = +1$ asigna el mínimo de energía a la configuración *cis* y un valor de $d = -1$ le asigna el máximo.

d) Inversión. Si un átomo I posee tres enlaces diferentes IJ, IK e IL (figura 4-1c), la energía potencial de la distorsión de los ángulos, E_i , es estimada como:

$$E_i = \frac{1}{2} \cdot K_i \cdot (\phi - \phi_0)^2 \quad (4-4)$$

en donde ϕ es el ángulo entre los planos IJK y LJK (a un átomo tetrahédrico con enlaces de igual longitud le corresponde un valor de ϕ de 35.264°), y K_i es la constante de fuerzas en $\text{kcal} \cdot \text{mol}^{-1} \cdot \text{rad}^{-2}$.

e) Electroestática. A cada átomo se le asigna una carga neta centrada en el núcleo, Q . Estas cargas son automáticamente calculadas en el caso de proteínas y ácidos nucleicos. La energía potencial debida a la interacción electrostática entre dos átomos (i, j) es:

$$E_{el} = \frac{332.0637 \cdot Q_i Q_j}{\epsilon \cdot R_{ij}} \quad (4-5)$$

en donde Q se encuentra en unidades de carga electrónica, R_{ij} es la distancia en Å entre los átomos i y j , ϵ es la constante dieléctrica, con un valor unitario para el vacío (ϵ_0). Para simular el efecto promedio del entorno, esta constante se calcula como $\epsilon = \epsilon_0 \cdot R_{ij}$. El factor numérico en la ecuación 4-5 realiza la conversión de unidades para que E_{el} resulte en $\text{kcal} \cdot \text{mol}^{-1}$.

f) Van der Waals. Este término se estima según la relación de Lennard-Jones con potencias 12 y 6:

$$E_{vdw} = (D_0)_{ij} \cdot \left\{ \left[\frac{R_0}{R} \right]^{12} - 2 \cdot \left[\frac{R_0}{R} \right]^6 \right\} \quad (4-6)$$

donde D_0 es la intensidad de la interacción, en $\text{kcal} \cdot \text{mol}^{-1}$ y R_0 la distancia de equilibrio en Å .

g) Puentes de hidrógeno. Estimado con una relación tipo Lennard-Jones con exponentes 12-10:

$$E_{ph} = (D_o)_{ad} \cdot \left\{ 5 \cdot \left[\frac{(R_o)_{ad}}{R_{ad}} \right]^{12} - 6 \cdot \left[\frac{(R_o)_{ad}}{R_{ad}} \right]^{10} \right\} \cdot \cos^2(\theta_{AHD}) \quad (4-7)$$

en donde $(D_o)_{ad}$ es la fuerza del enlace entre los átomos aceptor y donador en $kcal \cdot mol^{-1}$, $(R_o)_{ad}$ la distancia entre los átomos aceptor y donador y θ_{AHD} el ángulo formado por el aceptor del hidrógeno, el átomo de hidrógeno y el donador.

De esta manera, la energía potencial total es la suma de todas las contribuciones en una conformación particular de la molécula:

$$E = \underbrace{\sum E_{enl} + \sum E_{\theta} + \sum E_{\phi} + \sum E_l}_{\text{interacciones entre átomos enlazados}} + \underbrace{\sum E_{el} + \sum E_{vdW} + \sum E_{ph} + \sum E_r}_{\text{interacciones entre átomos no enlazados}} \quad (4-8)$$

En donde los E_r son términos adicionales de restricción conformacional que el usuario puede agregar, y las sumas se realizan sobre todos los enlaces, ángulos e interacciones presentes. El número de interacciones entre átomos no enlazados aumenta con el cuadrado del número de ellos. Para una proteína como la GPA, constituida por alrededor de 3000 átomos, son posibles más de cuatro millones de estas interacciones. Para reducir la capacidad de cómputo necesaria para simular la estructura de una proteína, en los cálculos se eliminan las interacciones entre átomos que se encuentran a distancias muy grandes. Se incluyen en la simulación solamente los términos de energía potencial que involucran átomos separados por una distancia (x) menor o igual a 9 \AA . Para evitar discontinuidades en el campo de fuerzas, se utilizó una función de corte gradual entre 8 y 8.5 \AA :

$$S = \begin{cases} 1 & \text{si } x \leq C \\ \frac{(F - x)^2 \cdot (F + 2x - 3C)}{(F - C)^3} & \text{si } C < x < F \\ 0 & \text{si } x \geq F \end{cases} \quad (4-9)$$

donde $C = 8$ y $F = 8.5$.

Una función similar se utiliza en el caso de los puentes de hidrógeno, para los ángulos de interacción (θ) entre los átomos aceptor, de hidrógeno y donador, en cuyo caso $x = \cos^2(180 - \theta)$, $C = \cos^2(180 - \theta_c)$ y $F = \cos^2(180 - \theta_f)$, con $\theta_c = 35^\circ$ y $\theta_f = 40^\circ$.

Todas las minimizaciones de la energía molecular se realizaron hasta que la diferencia de energía entre dos etapas consecutivas del cálculo alcanzara un valor menor a 0.01 kcal/mol. Las simulaciones de la dinámica molecular se realizaron empleando el algoritmo estándar de Verlet (1967), en donde se integran numéricamente las ecuaciones de movimiento de Newton. En ellas se describe la manera en que cambia la aceleración de los átomos en la estructura molecular conforme el tiempo transcurre, inducida por las fuerzas intermoleculares. Las relaciones fundamentales empleadas son:

$$F_{\alpha i} = \frac{\partial E}{\partial R_{\alpha i}} \quad \text{y} \quad m_i \cdot \frac{\partial^2 R_{\alpha i}}{\partial t^2} = F_{\alpha i} \quad (4-11)$$

donde m_i representa la masa de la partícula i , $R_{\alpha i}$ una de sus coordenadas espaciales ($\alpha = x, y, z$), t el tiempo y $F_{\alpha i}$ la componente de la fuerza en el eje α de esta partícula. Para simplificar la ecuación anterior, se definen:

$$F(R) = \frac{1}{m_i} \cdot F_{\alpha i} \quad \text{y} \quad \ddot{R} = \frac{\partial^2 R}{\partial t^2} \quad (4-12)$$

con lo que la ecuación 4-11 se convierte en:

$$\ddot{R} = F(R) \quad (4-13)$$

Utilizando diferencias finitas en la integración numérica de esta ecuación, podemos definir:

$$t_n = nh \quad (4-14)$$

donde t_n es el valor del tiempo en la n -ésima etapa del cálculo y h es el incremento del tiempo entre dos pasos consecutivos. Definiendo las variables discretas de la siguiente manera:

$$R_n = R(t_n), \quad \dot{R}_n = \dot{R}(t_n) \quad \text{y} \quad F_n = F(t_n) \quad (4-15)$$

las velocidades y aceleraciones pueden aproximarse por:

$$\dot{R} \approx \frac{R_{n+1} - R_{n-1}}{2 \cdot h} \quad \text{y} \quad \ddot{R}_n \approx \frac{R_{n+1} - 2 \cdot R_n + R_{n-1}}{h^2} \quad (4-16)$$

Usando 4-16 en 4-13 y despejando:

$$R_{n+1} = 2 \cdot R_n - R_{n-1} + h^2 \cdot F_n \quad (4-17)$$

Esta última relación, conocida como ecuación de Verlet (1967), permite estimar las coordenadas de la i -ésima partícula en el instante $n+1$ si se conoce su posición en las etapas n y $n-1$, además de la fuerza que sufre en el instante n . En este algoritmo, la velocidad está definida como:

$$v_n = \frac{R_{n+1} - R_{n-1}}{2 \cdot h} \approx \dot{R}_n \quad (4-18)$$

Al iniciar la aplicación del algoritmo se requieren las coordenadas y velocidades en el tiempo $n = 0$ (R_0 y v_0). Las velocidades iniciales son tomadas de una dinámica previa, o bien se asignan según una distribución aleatoria tipo Maxwell-Boltzman. Las ecuaciones 4-17 y 4-18 se transforman en:

$$R_1 = 2R_0 - R_{-1} + h^2 \cdot F_0 \quad (4-19)$$

$$v_0 = \frac{R_1 - R_{-1}}{2 \cdot h} \quad (4-20)$$

Pero dado que $R_{-1} = R_1 + 2 \cdot h \cdot v_0$, se tiene finalmente que:

$$R_1 = R_0 + h \cdot v_0 + \frac{1}{2} \cdot h^2 \cdot F_0 \quad (4-21)$$

De esta forma, el algoritmo estándar de Verlet utiliza la ecuación 4-21 para la primera etapa y la relación 4-17 en los pasos subsecuentes. A excepción de los errores de redondeo, 4-17 y 4-21 son ecuaciones conservativas de la energía, por lo que este tipo de cálculos son denominados *dinámica microcanónica*.

La temperatura del sistema se calculó en términos de la energía cinética promedio:

$$\frac{3}{2} \cdot nkT = \left\langle \sum \frac{1}{2} \cdot m_i \cdot v_i^2 \right\rangle_{\text{ciclo}} \quad (4-22)$$

donde los corchetes indican el promedio en un ciclo de equilibración, que en este estudio se estableció en 50 etapas de cálculo. Una vez completado este ciclo, la temperatura del sistema se comparó con la del baño térmico, T_0 (esto es, la establecida como constante), y el conjunto de velocidades atómicas se escaló por el siguiente factor:

$$\frac{v_{\text{nueva}}}{v_{\text{anterior}}} = \sqrt{\frac{T_0}{T}} \quad (4-23)$$

Debido a esta modificación en las velocidades, para reiniciar la dinámica debe utilizarse nuevamente la ecuación 4-21 en el primer paso del cálculo.

En este trabajo, se utilizó un incremento de tiempo entre cada cálculo, h , de 1 fs (el valor recomendado para evitar tensiones exageradas por el cambio demasiado grande en las posiciones de los átomos es $h \leq 2$ fs), por lo que la corrección del conjunto de velocidades atómicas, para mantener la temperatura virtual del sistema constante, se realizó cada 0.5 ps de simulación. En la etapa inicial del modelado, se incluyeron explícitamente los átomos de hidrógeno sólo en el caso en que éstos estuviesen unidos a átomos de oxígeno o nitrógeno; esto permitió disminuir el tiempo de cálculo sin despreñar la contribución que los posibles enlaces de hidrógeno pudieran aportar a la energía potencial total. Dada la complejidad del sistema a modelar, se evitó incluir al solvente explícitamente realizando todos los cálculos *in vacuo*. Esta decisión es justificada por el hecho de que la parametrización utilizada fue construida para incorporar el efecto del solvente en las interacciones intermoleculares.

El modelado de la estructura de la GPA se inició a partir de las coordenadas cristalográficas de la PTB (Huber *et al.*, 1987b), las cuales fueron generosamente proveídas por el Dr. Robert Huber del Instituto Max Planck. Partiendo de la PTB, se reemplazaron todos los aminoácidos necesarios para que la cadena polipeptídica de esta proteína se transformara en la de la GPA, de acuerdo con el alineamiento mostrado en la figura 3-13. Para evitar que las nuevas cadenas laterales de los aminoácidos de la GPA incluyeran tensiones exageradas en la estructura quimérica, cada mutación efectuada fue seguida de una minimización del nuevo grupo químico incluyendo las cadenas laterales en su microentorno (en un radio de 8 Å). Una vez sustituidos los nuevos aminoácidos y relajadas las cadenas laterales, se procedió a insertar los segmentos 12 a 13, 35 a 37, 66 a 68 y 148 a 149, ausentes en la molécula molde, y a eliminar los aminoácidos de la PTB que no se encuentran en la GPA: un total de 12, repartidos en cuatro segmentos que se localizan después de los residuos 85, 94, 104 y 117 (figura 3-13). Se realizó una nueva minimización local de energía en estas regiones y en aquéllas donde estuvieron involucrados en los cambios de aminoácidos residuos de prolina, permitiendo el reacomodo de los átomos tanto de las cadenas laterales como de la cadena principal. Después de estas modificaciones, se disminuyó la tensión de las cadenas laterales, que en esta etapa se encuentran interaccionando con un microambiente diferente, mediante una minimización de energía global, primero manteniendo la cadena polipeptídica fija y después relajando la molécula completa.

Durante el proceso anterior, y debido a las tensiones locales que no pudieron ser relajadas por la minimización, algunos enlaces peptídicos fueron forzados a adoptar la configuración *cis*, mientras que algunos carbonos α sufrieron inversiones, que convirtieron a los aminoácidos que los contenían en residuos de la serie *d*. La estereoquímica de estos grupos químicos se retornó a su geometría natural, seguida de la correspondiente minimización local de energía.

A continuación, se construyeron los enlaces disulfuro creando un enlace covalente entre los átomos de azufre involucrados. Es necesario señalar que la distancia entre las cadenas laterales de las cisteínas 72 y 165 inicialmente era demasiado grande para enlazarlas (alrededor de 10 Å), por lo que se manipuló su orientación y la de un segmento de cadena principal en conformación irregular (159-165); esto permitió acercar los grupos involucrados en este enlace disulfuro a una distancia razonable. La correspondiente relajación local involucró la minimización de energía de todos los átomos cercanos a ambos residuos de cisteína.

La siguiente etapa en el modelado de la GPA consistió en la remoción de todas las moléculas de agua presentes en la PTB y la minimización de energía por segmentos de 4 a 12 aminoácidos de toda la molécula. Las zonas de mayor flexibilidad, es decir, giros y regiones externas, podrían haber llegado a una conformación que correspondiera a un mínimo local pero que contuviera fuerte tensión molecular. Por tal motivo, se realizaron dinámicas moleculares locales en tales regiones, logrando que dichas zonas cambiaran este tipo de geometrías. Una minimización posterior condujo a estos fragmentos de la molécula a una conformación de baja energía.

Antes de iniciar con la dinámica molecular global del modelo de la GPA, se requirió de una nueva minimización de energía, esta vez sobre toda la glicoproteína para eliminar en lo posible las tensiones en ángulos, enlaces e interacciones no específicas y facilitar así la trayectoria de convergencia.

La última etapa del modelado incluyó dinámica molecular en una serie de pasos, cada uno de ellos se efectuó a temperatura constante. Conforme éstos fueron llevándose a cabo, la temperatura virtual de la dinámica fue descendiendo y las restricciones impuestas desapareciendo para que la última etapa consistiera de una dinámica libre a temperatura ambiente. A continuación se describen los diferentes pasos que constituyeron esta etapa. La primer

dinámica molecular se realizó a 600 K durante 50 ps con el ligando de la PTR en su posición original, es decir, con la biliverdina incluida como ligando de la GPA. La razón de haber llevado a cabo este cálculo fue la de permitir que las regiones más flexibles de la molécula convergieran a conformaciones con menor tensión. La temperatura virtual se estableció en un valor muy alto para asignar energías cinéticas altas a los átomos que les permitiera vencer las barreras de potencial entre el mínimo local de energía en que se encontraban y la posición de menor tensión. El hecho de no retirar el ligando original evitó un cambio repentino en el microambiente antes de este rearrreglo. Durante esta simulación se aplicaron algunas restricciones: las distancias entre donadores y aceptores de los enlaces de hidrógeno de la cadena principal se mantuvieron cerca de su valor original; además, se evitó que los enlaces peptídicos pudieran girar para abandonar su geometría *trans*. Ambas restricciones se agregaron al término de energía potencial total denotado por E_r en la ecuación 4-8, a través de un término armónico similar a los de distancias y ángulos de enlace.

Antes del segundo paso de la dinámica molecular, se sustituyó a la biliverdina por progesterona. Esta última molécula es el esteroide que se sabe que se une a la GPA con la mayor constante de afinidad (Westphal, 1985). A continuación se realizaron 50 ps de dinámica molecular a 500 K aplicando las mismas restricciones mencionadas en el paso anterior. Debido a que en esta simulación algunos carbonos α , localizados en las regiones más externas, sufrieron inversiones produciendo aminoácidos del tipo *d*, su geometría fue corregida y restringida al tipo *l*. Posteriormente, se realizaron dos nuevas etapas de dinámica molecular, simulando 30 ps en cada una de ellas, a temperaturas virtuales de 400 y 300 K. Finalmente, después de una minimización global de la energía molecular, un último paso se realizó simulando 150 ps a 300 K, en las mismas condiciones de los pasos anteriores, pero restringiendo únicamente la conformación *trans* de los enlaces peptídicos e incluyendo de manera explícita todos los átomos de hidrógeno de la GPA.

PREDICCIÓN DE LA ESTRUCTURA TRIDIMENSIONAL.

Como se mencionó en el capítulo 3, el estudio comparativo de las secuencias de la GPA y las lipocalinas, la semejanza estructural de los genes de algunas de estas proteínas y la evidencia experimental, permiten considerar que la glicoproteína posee la arquitectura común de las lipocalinas, y que su

conformación puede modelarse usando una de las estructuras cristalográficas disponibles de los miembros de esa familia. De esta manera, se ha modelado ya la conformación tridimensional de otra lipocalina, la cadena γ del complemento C8 (Haefliger *et al.*, 1991). Por otra parte, la evidencia experimental señala que para la predicción de la conformación molecular de la parte proteica de la GPA puede prescindirse de las cadenas de carbohidratos unidas a ella, y que la presencia de la progesterona o algún otro ligando unido a la cadena polipeptídica, no altera significativamente la estructura tridimensional (capítulo 2, página 27).

La parametrización seleccionada para efectuar los cálculos de mecánica molecular fue el *Dreiding*. Este campo de fuerzas fue desarrollado con el propósito de que fuera general, es decir, de que no poseyera una tendencia específica para predecir la estructura y dinámica de un tipo particular de moléculas. Esto contrasta con otras parametrizaciones ampliamente utilizadas, como AMBER (Weiner *et al.*, 1986) y CHARMM (Nillson y Karplus, 1986), específicos para proteínas y ácidos nucleicos.

Durante la construcción del modelo, se observó que todos los cambios de cadena lateral, necesarios para convertir la cadena de PTB a GPA, fueron complementarios o involucraron la sustitución de una cadena lateral por otra de menor tamaño. Esto es, los cambios de aminoácido que requerían un mayor volumen para la nueva cadena lateral, se acompañaban de otra sustitución que producía tal espacio; otros cambios de aminoácido se realizaron en la superficie de la molécula o no requirieron de volumen adicional. Por lo anterior, no existieron conflictos serios de espacio durante la construcción del primer modelo de la GPA. Sin embargo, en algunos casos fue necesaria una reorientación de las nuevas cadenas laterales para disminuir repulsiones excesivas. La región de los últimos 15 aminoácidos de la cadena polipeptídica de la GPA no fue incluida en el modelado por la falta de una estructura topológicamente equivalente en la PTB, la gran hidrofiliidad de ese segmento y su localización en el extremo de la proteína.

En todas las minimizaciones de energía que se llevaron a cabo en la GPA (a excepción de algunos casos extremos que involucraron inserciones o pérdidas de segmentos de cadena) la estructura molecular se modificó sólo marginalmente, alcanzando de manera rápida un mínimo energético local. Por otra parte, durante las simulaciones de dinámica molecular se observó una convergencia

rápida de la trayectoria que ni elevó la energía total, ni cambió sustancialmente la estructura. Ambos resultados muy probablemente, son consecuencia de que la estructura tuvo tensiones internas relativamente pequeñas, que no desviaron las trayectorias de simulación muy lejos de su camino a mínimos locales o a estructuras de equilibrio. A su vez, la baja tensión intermolecular fue debida a la ausencia de conflictos de espacio durante la sustitución de aminoácidos, como se describió en el párrafo anterior.

Cuando la molécula de progesterona fue colocada en lugar del ligando original de la PTB (biliverdina IX), no existía información alguna sobre la posible colocación de este esteroide. La progesterona posee un grupo carbonilo conjugado con un doble enlace carbono-carbono y un acetilo. Estos dos grupos se encuentran en posiciones opuestas de la molécula (figura 4-2) y son las regiones más polares de la misma. Se decidió colocar el grupo menos polar (el acetilo) apuntando hacia el interior de la GPA y el carbonilo al exterior. Es importante resaltar que los sitios de unión de ligandos del molde y de la glicoproteína son diferentes, y que la ubicación inicial de la progesterona en este lugar es sólo una suposición guiada principalmente por intuición química. Durante los primeros 30 ps de la dinámica molecular, este ligando mostró fuertes oscilaciones respecto a sus coordenadas iniciales. Después de este periodo, adoptó una orientación perpendicular a la original girando sobre el plano de la figura 4-2, manteniendo esta orientación durante el resto de los cálculos con sólo leves fluctuaciones.

Durante la dinámica molecular global (en la que estuvieron involucrados todos los átomos de hidrógeno del sistema), se incluyeron los 168 primeros aminoácidos de la porción polipeptídica de la AGP. Esto involucró un total de 2781 átomos, 5097 ángulos, 7550 torsiones y 1758 centros de inversión. La convergencia de la trayectoria de esta simulación se alcanzó a los 100 ps. Se muestra en la figura 4-3 la traza de la cadena principal (línea que une todos los carbonos α de la molécula) de la estructura final modelada. En esta misma figura se presenta, con fines comparativos, la traza de la cadena principal de la PTB.

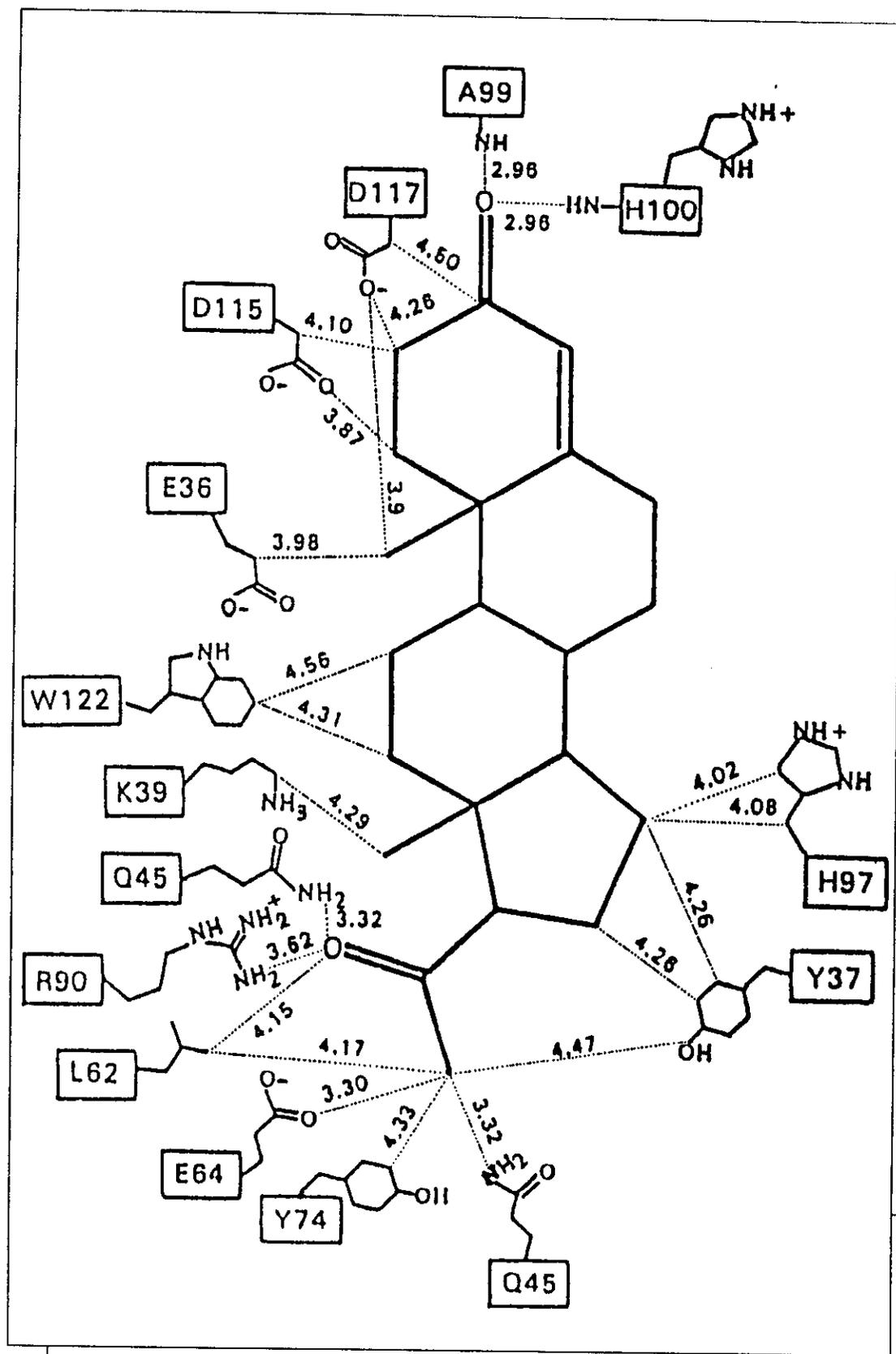


FIGURA 4-2. Representación esquemática de las interacciones entre la GPA y la progesterona. Las distancias se indican en Å.

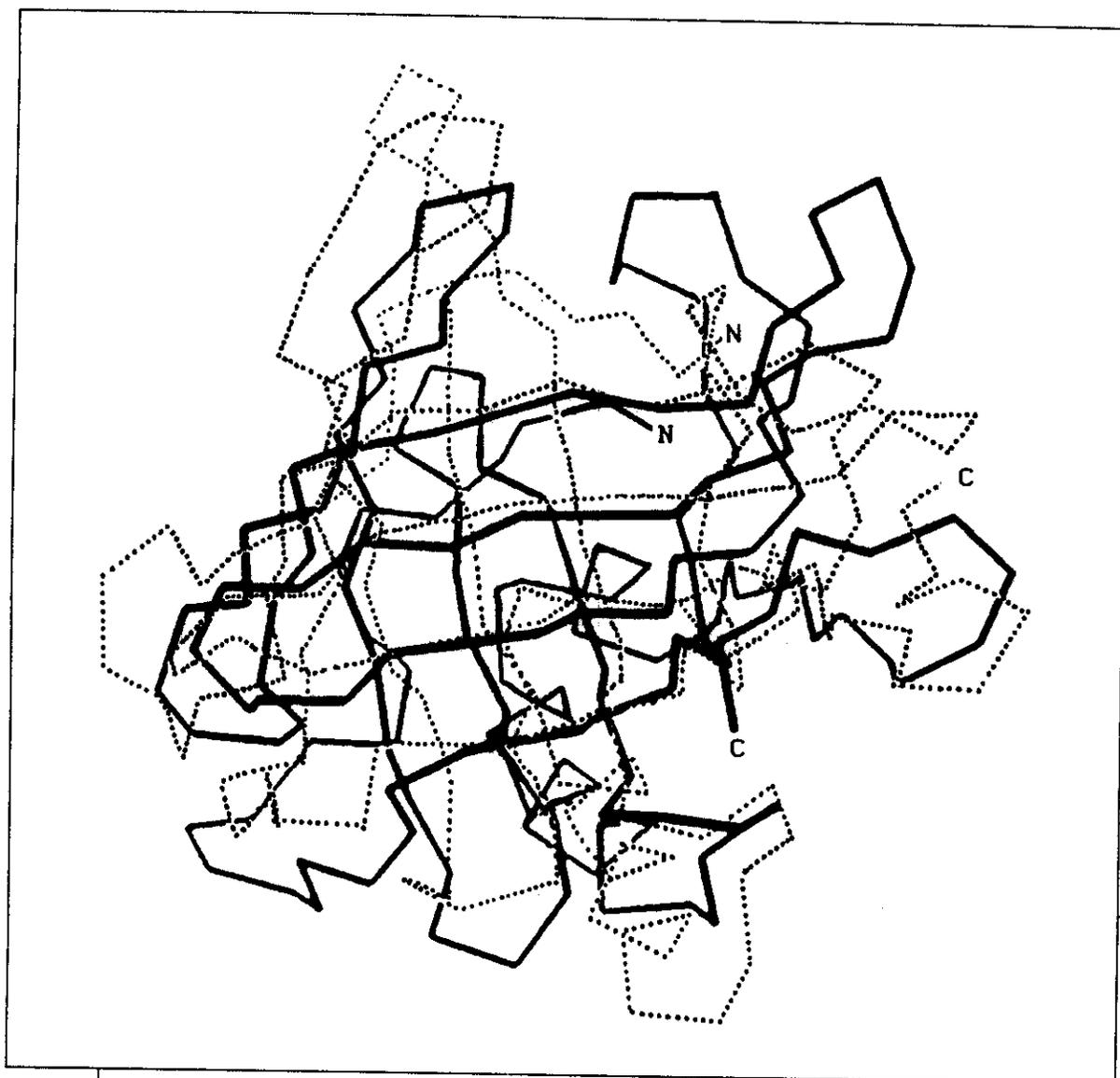


FIGURA 4-3. Traza de los carbonos α del modelo de la GPA (—) sobrepuesta a la del modelo inicial, la PTB (.....).

ANÁLISIS DEL MODELO.

El mapa de Ramachandran del modelo de la AGP (figura 4-4) muestra que la mayoría de los ángulos ϕ y ψ de su cadena principal se encuentran en la zona de baja energía. Esta región, delimitada con línea continua en la figura, contiene al 65% de los ángulos que pertenecen a aminoácidos diferentes a G o a P. De acuerdo con Morris *et al.* (1992), la estructura de una proteína, determinada por difracción de rayos X, que posea este porcentaje de ángulos en la zona marcada, correspondería aproximadamente a una resolución de 3.1 Å o bien a un factor R de 0.33. Todos los ángulos ϕ o ψ con valores fuera de las zonas de baja energía pertenecen a aminoácidos de las regiones externas de la estructura. La línea discontinua muestra las regiones clásicas preferidas de los ángulos ϕ y ψ observados en proteínas (Ramakrishnan y Ramachandran, 1965).

Al continuar el análisis de la estructura propuesta, se encuentra que todos los aminoácidos poseen la quiralidad correcta (configuración 1) y que todos los enlaces peptídicos tienen la conformación trans. Las cadenas laterales de I y T que poseen carbonos quirales, tienen la estereoquímica encontrada en los isómeros naturales. Los ángulos ω mantienen su planaridad en toda la cadena principal del modelo de la AGP con desviaciones menores a 30°. En el caso de los enlaces disulfuro, los valores de los ángulos dihédricos se encuentran dentro de los intervalos observados en las estructuras conocidas de proteínas (Morris *et al.*, 1992), y corresponden a 112 y -124° para los enlaces S5-S147 y S72-S165, respectivamente. Utilizando el algoritmo propuesto por Stouch y Jurs (1986), se estimó el volumen de la porción modelada de la AGP en 24510 Å³ y una superficie accesible de 7,800 Å², empleando un radio de prueba de 1.4 Å que simula al disolvente acuoso.

En la figura 4-5 se muestra la conectividad de los puentes de hidrógeno entre grupos de la cadena principal de la GPA, calculados con límites de longitud de 3.5 Å y ángulos de desviación menores a 35°. Puede observarse de esta figura que cada cadena de las hojas β es seguida y precedida por otras, en orientación antiparalela, que se desplazan unas respecto a otras. Este patrón es prácticamente idéntico al de las otras lipocalinas con estructura conocida (Cowan *et al.*, 1990), y puede especularse que este desplazamiento relativo puede ser muy importante para dirigir a la cadena a la conformación nativa durante el proceso de doblado. Las estructuras secundarias en la AGP determinadas del patrón de puentes de hidrógeno, se muestran con barras en la figura 3-13. Se incluyen en esta figura las estructuras secundarias presentes en la PTB, con el propósito de compararlas con las del modelo.

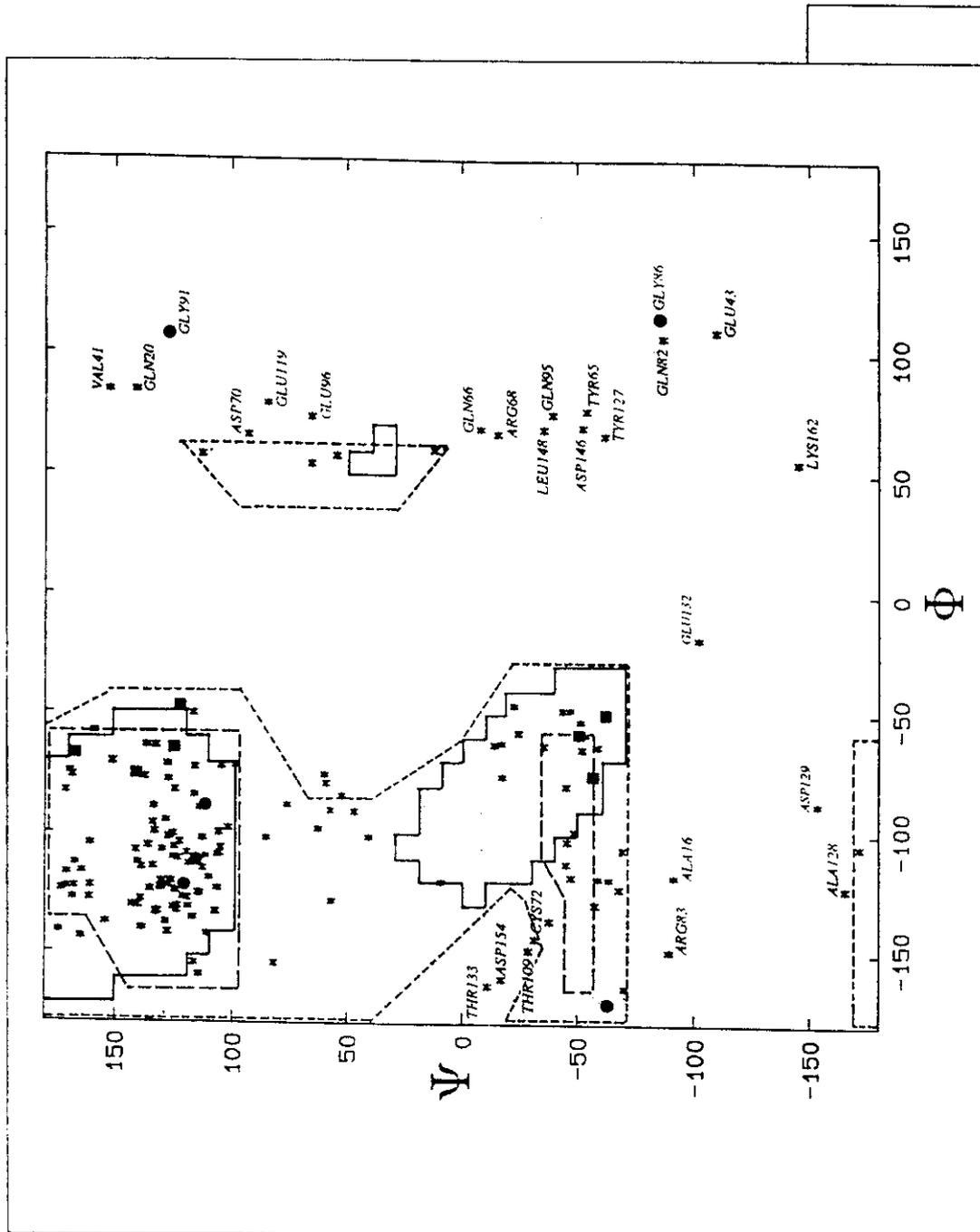


FIGURA 4-4. Distribución de los ángulos diédricos de la cadena principal de la GPA. Se indican por separado los residuos de G (●) y P (■). Se muestran además las regiones preferidas (---) y de baja energía (—)

Las cadenas laterales de los cinco residuos de asparagina que unen a las cadenas de carbohidratos en la glicoproteína, se encuentran expuestas al solvente. Los sitios N15, N54 y N85 se encuentran en los giros que unen el extremo amino con la cadena A y en los que unen a las cadenas B-C y D-E. La estructura de las lipocalinas posee un barril cuyas paredes son constituidas por hojas β , y su base es cerrada por los giros descritos arriba, además del que une las cadenas F-G y a la cadena H con la Hélice I. El N38 también se encuentra en un giro, entre las cadenas A-B, en el extremo del barril que interacciona con el ligando, mientras que el N75 se encuentra sobre la cadena D, cerca del extremo carboxilo de la molécula.

La progesterona tiene un número grande de interacciones con los grupos químicos de la proteína. En la figura 4-2 se representan las interacciones entre los átomos más cercanos. En particular, resaltan las distancias cortas, debidas a la formación de puentes de hidrógeno, entre los átomos de oxígeno del ligando y los aminoácidos Q45, R90, A99 y H100. En esta misma figura es notable el que las interacciones proteína-ligando se encuentran principalmente de un lado del esteroide. El lado opuesto de esta molécula se encuentra parcialmente expuesto al solvente, pero sin ser completamente accesible a él por la protección de los giros 34-38 y 94-97.

COMPARACION CON LOS RESULTADOS EXPERIMENTALES.

Es interesante comparar las características del modelo propuesto con la información experimental que se tiene sobre esta molécula. Estudios de reactividad química (Schmid, 1976) y perturbación de solvente (Zubillaga-Luna et al., 1991) han mostrado que solamente uno de los tres residuos de triptofano presentes en la AGP y que aproximadamente la mitad de las doce tirosinas se encuentran expuestos al solvente. En la estructura modelada de la AGP, observamos al W160 expuesto al solvente de manera casi total, mientras que los W25 y W122 se encuentran en el interior de la proteína, lejos de la superficie molecular (véase la figura 4-6B). Por otra parte, en el modelo pueden identificarse cinco cadenas laterales de tirosina externas (Y50, Y91, Y127, Y142 y Y157), tres completamente internas (Y27, Y78 y Y110), y tres con accesibilidad parcial al solvente (Y37, Y65 y Y74). En esta descripción se tienen sólo once tirosinas, número presente en la secuencia que se manejó en el modelo; la doceava tirosina reportada en el estudio de Schmid no existe en la secuencia de aminoácidos de la variante A de la GPA.

La posición de los triptofanos y tirosinas en el modelo corresponde con los estudios experimentales realizados en esta glicoproteína; sin embargo, Schmid *et al.* (1976) identificaron a las Y27 y Y78 como reactivas, en desacuerdo con la estructura que se propone en este trabajo.

Los grupos aromáticos no han sido los únicos cuya reactividad se ha estudiado en la GPA. Los mismos autores (Schmid *et al.*, 1976) encontraron que ambos enlaces disulfuro, las tres histidinas y la única metionina se encuentran parcialmente expuestas al solvente. Además, proponen que todas las lisinas y todos los ácidos carboxílicos, a excepción de dos, están localizados en el exterior de la molécula. Comparando la localización de los grupos químicos en el modelo de la AGP con estos resultados experimentales, puede observarse que el enlace disulfuro entre la C5 y la C147 se encuentra parcialmente protegido del solvente por las cadenas laterales de L4, L8, L145, D146 y L148, mientras que el otro disulfuro (72-165) aparentemente se encuentra expuesto al solvente, pero localizado en una región cercana a la segunda hélice y al sitio de glicosilación N75, porciones de la molécula que podrían ocultar parcialmente a este enlace disulfuro. Por otra parte, la H97 se encuentra interaccionando con el ligando; la H100 se encuentra en la superficie, pero cubierta parcialmente por las cadenas laterales de S89, F98 y V116; y la H171 pertenece a la zona de 15 aminoácidos no modelada en el extremo carboxilo de la AGP. Puede observarse que el modelo propuesto está de acuerdo con la localización de los grupos químicos propuesta por Schmid. Un caso que no concuerda es la metionina, que se encuentra en el interior del barril β , aparentemente aislada del exterior. Sin embargo, todas las lisinas y los grupos carboxílicos de los residuos de ácido aspártico y glutámico se encuentran expuestos al solvente, excepto el D58 que se localiza en el interior de la proteína. Puede concluirse que, en términos generales el modelo concuerda con la evidencia experimental en cuanto a la reactividad química concierne.

Respecto a otro tipo de estudios experimentales sobre la GPA, se ha observado por diferentes métodos la presencia de un intermediario estable en su desnaturalización térmica (Halsall y Kirley, 1981; Rojo-Domínguez *et al.*, 1990 y el capítulo 2 de este trabajo). Como se describió en la parte experimental, existe evidencia para proponer que este intermediario podría consistir en una especie con una porción de su estructura, que se encuentra plegada de manera poco compacta, parcialmente desdoblada. Esta región parece estar localizada cerca del extremo carboxilo terminal. Incluso, como se describió anteriormente, puede señalarse a un residuo de triptofano involucrado en este

cambio e inmerso en una zona con gran densidad de carga eléctrica. Utilizando el modelo que se propone para la GPA, puede observarse una región, que inicia en el aminoácido 128 y se extiende hasta el extremo de la cadena, que tiene un empaquetamiento poco compacto. Esta zona incluye a la hélice I y a una porción irregular y podría llegar a contener a los últimos 15 residuos no modelados entre los cuales hay una gran cantidad de aminoácidos cargados. Esta observación concuerda con los estudios de desnaturalización, sobre todo al notar que el triptofano 160 se encuentra en la superficie de la estructura modelada e interaccionando a una distancia menor de 4 Å con los siguientes residuos cargados: K24, K161, K164, D159 y E166.

Otra característica interesante es el hecho de que al observar la posición de los reemplazos de aminoácidos entre las variantes A (que es la modelada en este trabajo) y B, se encuentra que la región de la secuencia de mayor concentración de sustituciones está entre los aminoácidos 100 y 107, donde se presentan siete de los veintidos reemplazos. Es sorprendente notar que esta zona es la más cercana a la hélice I, pues esto es consistente con la propuesta de que la hélice forma parte de la porción poco empacada de la AGP al haber gran variabilidad en los aminoácidos en el barril β que interaccionan con ella. Sólo uno de los siete reemplazos mencionados en el párrafo anterior involucra cadenas laterales del interior del barril, consistiendo en el cambio de una alanina a una glicina. Sólo existen otras tres sustituciones en la región interna de la molécula, todas de naturaleza conservativa: I88→V, I103→F y V125→F). El resto de los reemplazos ocurren en regiones externas del modelo. Por otra parte, es importante hacer notar que un número alto de las sustituciones de aminoácidos (siete de las veintidos o el 32%) involucran cadenas laterales que interaccionan con la progesterona. Este hecho es remarcable por dos razones: primero, es un hecho que apoya la posición del ligando en el modelo y por otra parte, permite proponer que la localización de estas sustituciones no es casual. En efecto, el hecho de que las dos variantes de la AGP se sinteticen en proporción de tres a uno, y que el 32% de los cambios en estas especies se encuentren en el sitio de unión de ligandos, parece sugerir que las dos variantes poseen diferencias en su afinidad, o incluso en su especificidad, por ligandos hidrofóbicos. De esta manera, es posible que las dos variantes de GPA posean funciones biológicas complementarias. Aunque, hasta lo que sabemos, no se han realizado estudios de afinidad por ligandos en las variantes de esta glicoproteína por separado, el análisis del modelo propuesto para la GPA permite entrever la importancia de estos experimentos.

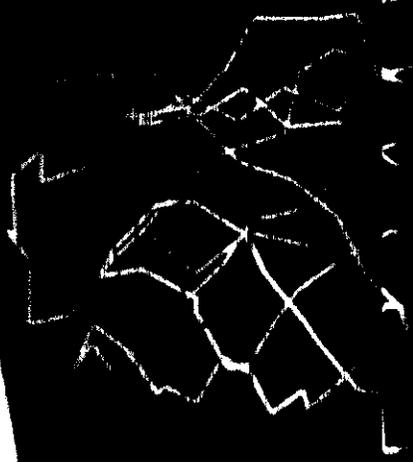
Finalmente, concluyendo la comparación de los resultados experimentales con el modelo de la glicoproteína, cabe mencionar que Kute y Westphal (1976) encontraron experimentalmente que un residuo de lisina, dos de tirosina y uno de triptofano están localizados en el sitio de unión de la progesterona. Incluso, lograron estimar la distancia entre el esteroide y el grupo indol del residuo de triptofano en el intervalo de 9.1 a 14.1 Å. La coincidencia de estos resultados con la estructura de la GPA construida en este trabajo es completa, pues se observan los residuos Y37, K39, Y74 y W122 interaccionar con el ligando (véase la figura 4-2), además de que la distancia entre los centros de la cadena lateral del W122 y la progesterona resulta ser de 9.7 Å.

En la figura 4-6 se muestran diferentes esquemas del modelo de la GPA. En el recuadro A se tiene la sobreposición de la traza de los aminoácidos 1 a 168 de la glicoproteína (amarillo) y la PTB (rojo). En la parte superior de la ilustración se encuentran las hélices del final de la cadena de estas proteínas. En el caso de la GPA se tiene esta estructura secundaria desplazada a la derecha respecto a la estructura molde. La posición de las hojas β permanece aproximadamente en su sitio, sin embargo, las regiones de giros que unen a las cadenas β sufrieron cambios estructurales mayores, sobre todo en el sitio de unión de ligandos, ubicada en la esquina inferior izquierda de este esquema. Esta figura es similar a la 4-3, aunque con una orientación diferente.

El siguiente recuadro (B) muestra la localización de las cadenas laterales de triptofano (amarillo) en la traza de la GPA (rojo). Con color azul se muestra a la progesterona unida a la glicoproteína. Pueden apreciarse la cercanía de este esteroide con el W122, la exposición al solvente del W160 (lado izquierdo) y la posición interna del W25. El extremo amino se localiza a la izquierda de la figura.

En el esquema C se muestra otra orientación de la cadena principal de la glicoproteína y la posición del esteroide (verde) en el barril formado por las hojas β . El eje de este barril coincide con la vertical de la figura. Los extremos amino y carboxilo se localizan en la parte media de la figura a la izquierda y a la derecha, respectivamente. Es interesante notar cómo los giros de la cadena principal que unen a los segmentos constituyentes de las hojas β se pliegan sobre la estructura cerrando la parte inferior del barril y se abren en la región superior formando el cáliz que alberga al ligando. La hélice I se observa en la parte posterior de la molécula (abajo a la

1271



A

B

C

D

E

FIGURA 4-6.

F

izquierda). La siguiente representación (D) incluye en color amarillo a las cadenas laterales de la GPA que interaccionan con el esteroide, en el análogo tridimensional de la figura 4-2. En esta orientación se aprecia la entrada al cáliz y cómo el espacio que aparentemente está vacío en las representaciones que sólo incluyen la traza de la cadena principal se llena con las cadenas laterales.

La figura 4-6E muestra a los dos enlace disulfuro (blanco) y a los puentes de hidrógeno entre átomos de la cadena principal (amarillo), como se representa en la figura 4-5. En este esquema se sustituyó la traza por el conjunto de átomos de la cadena principal. La hélice I se localiza en la esquina superior izquierda. La última representación (F) muestra una orientación más de la molécula que permite apreciar la cavidad interna que aloja a la progesterona (magenta). Esta característica es consistente con la alta accesibilidad al solvente de los grupos químicos de la GPA. Se incluyen además las cadenas polares y cargadas de la molécula, las cuales se localizan principalmente en el exterior.

En estos esquemas puede observarse la posición perpendicular de las cadenas que forman el barril (B, C, D y F) y que la región que incluye a la hélice I y el segmento que le sigue hasta el extremo carboxilo se encuentran empacadas deficientemente, esto es, que existe una distancia grande entre estos segmentos y el resto de la molécula (A, B, C, E, F).

CAPITULO 5.

CONCLUSIONES GENERALES.

En la introducción se mencionó el interés que actualmente tienen los estudios sobre la arquitectura de las proteínas para comprender mejor una gran cantidad de fenómenos biológicos y para su aplicación tecnológica. A pesar de la gran importancia que tiene el poder conocer la estructura de una proteína, y del aumento en la velocidad con la que se están determinando nuevas estructuras, esto es frecuentemente difícil para una proteína específica. Así, en el caso de la GPA no han sido posibles ni su cristalización apropiada ni sus estudios por resonancia magnética nuclear. Queda entonces la predicción de la estructura tridimensional modelada por medio de la mecánica molecular. Sin embargo, esta estimación requiere necesariamente de la convergencia de resultados derivados de una gran variedad de disciplinas diferentes. En este tipo de trabajos, es fundamental la complementariedad de los estudios teóricos y experimentales. Esta tesis es el primer esfuerzo en esa dirección en nuestro laboratorio, y posiblemente en el país.

El objetivo planteado en este trabajo, el estudio y predicción estructurales de la glicoproteína ácida α_1 , fue alcanzado al obtener un modelo de la arquitectura de esta macromolécula, el cual cumple simultáneamente con las características estereoquímicas de las proteínas y con la información estructural sobre ella que se ha determinado experimentalmente. Con el avance actual en el campo y la tecnología disponible, resulta aún imposible modelar una proteína si no se dispone de una estructura homóloga que pueda utilizarse como punto de partida en la simulación. En nuestro caso, pudo justificarse el clasificar a la GPA como un miembro de la familia de las lipocalinas; la evidencia para tal justificación proviene tanto de estudios experimentales (análisis de espectros de dicroísmo circular, determinación de su interacción con ligandos hidrofóbicos, tamaño de la porción peptídica, conservación de la secuencia en dos regiones y de algunas cisteínas involucradas en puentes disulfuro, homología en la estructura de los genes que las codifican, etc.) como teóricos (clasificación estructural con base en su composición de aminoácidos, significación de su homología, estimación de las posiciones de sus regiones antigénicas, predicción de estructuras secundarias, análisis comparativo de los núcleos hidrofóbicos, etc.).

Se determinó que el proceso de desnaturalización de la GPA es complejo. La transición térmica desde el estado nativo hasta el desnaturalizado ocurre a través de al menos un intermediario a concentraciones significativas. Este

intermediario no difiere sustancialmente de la estructura nativa, mantiene la disposición de sus grupos aromáticos respecto al solvente, además de que no altera su contenido de estructuras secundarias. El mayor cambio estructural parece encontrarse en una región que contiene un residuo de triptofano y se localiza cerca del extremo carboxilo de la molécula. Este resultado es corroborado en el modelo molecular, al encontrar a un grupo indol en un entorno con gran cantidad de cargas eléctricas en los últimos aminoácidos de la cadena. En la estructura propuesta para la GPA, se puede identificar una región cuyo empaquetamiento es menos compacto que en el resto de la molécula y que incluye al extremo carboxilo de la cadena, en buen acuerdo con los resultados experimentales. La semejanza de la especie intermediaria con la estructura nativa permitió utilizar el modelo de dos estados para estimar los parámetros termodinámicos de la estabilidad de la proteína. Los resultados de los estudios de desnaturalización inducida por hidrocioruro de guanidina sugieren que la proteína desnaturalizada térmicamente conserva algún tipo de estructura residual. De cualquier manera, la estructura nativa de la GPA parece tener gran accesibilidad al solvente. Esto coincide con el modelo propuesto de su estructura, dada la gran cavidad que posee para albergar al ligando, a la región de bajo empaquetamiento, a la gran cantidad de segmentos que sobresalen de la estructura globular y a la presencia del segmento, localizado al final de la cadena, que posee gran cantidad de cargas eléctricas.

Aunque el modelo de la GPA que se presenta en este trabajo no contiene la porción de carbohidratos de la molécula, ni a los últimos 15 aminoácidos de la cadena, es suficiente para mostrar la arquitectura general de la parte proteica y muchos de los detalles particulares. Se ha demostrado que la porción glicosídica no afecta sensiblemente la conformación de la cadena polipeptídica. Esta es consistente con la topología general de las lipocalinas, aunque es posible que posea un segmento helicoidal adicional altamente polar. Las variantes A y B parecen tener diferente afinidad e incluso distinta especificidad por su ligando natural, a juzgar por el alto número de reemplazos de aminoácidos involucrados en las interacciones con la progesterona. La posición de este esteroide en la cavidad de la GPA parece ser la adecuada, estabilizada por cuatro puentes de hidrógeno y una gran cantidad de interacciones tanto con cadenas laterales como con grupos químicos de la cadena polipeptídica. El estudio de la dinámica molecular de este ligando indica que después de los 30 ps iniciales, la progesterona encontró una

orientación muy favorecida, en un mínimo energético de gran profundidad que estabilizó fuertemente su posición. A pesar de que se ha reportado que al menos en una lipocalina el sitio de unión de moléculas hidrofóbicas se encuentra en otro lugar, en el caso de la GPA al igual que en la PTB y la PTR éste se encuentra en la cavidad formada por las hojas β , como se observa en el modelo.

La estructura propuesta para la glicoproteína depende en gran medida de la homología con las lipocalinas y en particular del alineamiento apropiado con la macromolécula que sirvió de molde. Sin embargo, basados en la coincidencia en la posición de las estructuras secundarias, en el análisis de núcleos hidrofóbicos y en los resultados favorables en la comparación del modelo con la información estructural disponible, puede considerarse que el alineamiento encontrado de la secuencia de la GPA con la de la PTB refleja de manera muy cercana la correspondencia entre las regiones topológicamente equivalentes en ambas macromoléculas, y que la predicción de la estructura de la glicoproteína ácida α_1 es adecuada.

Lamentablemente, no fue posible identificar el sitio de unión de drogas básicas a la GPA. Puede especularse que éste se encuentre en la porción de carbohidratos de la glicoproteína que tiene un alto número de residuos de ácido siálico; o bien, que la porción carboxilo terminal de gran cantidad de cargas eléctricas esté involucrada en este tipo de interacciones. En realidad, el modelo propuesto ha permitido responder algunas de las interrogantes sobre la conformación de esta macromolécula, pero ha generado otras que hacen del estudio estructural de la GPA y de las lipocalinas en general toda un línea de investigación. En el futuro se planea estudiar la desnaturalización de estas moléculas por medio de la calorimetría diferencial de barrido, y continuar con el modelado de algunas de ellas, con particular interés en el análisis de sus variantes y de sus interacciones con moléculas hidrofóbicas. Este trabajo es una contribución pequeña, pero entusiasta, a los estudios estructurales de proteínas, y en general, a la búsqueda de los fenómenos que dirigen el plegamiento de estas moléculas a la muy particular y funcionalmente activa estructura nativa.

APENDICE I.

LISTADO DEL PROGRAMA "TRANSICION".

```

10 '*****
20 '
30 '          TRANSICION (TR)
40 '
50 ' Este programa ajusta los puntos experimentales de una curva de
60 ' transición a una función doble exponencial definida en las
70 ' líneas 10000 a 10080, calculando los parámetros termodinámicos.
80 '
90 ' Arturo Rojo, Rafael Zubillaga y Andrés Hernández. Mayo 1989.
100 ' Departamento de Química. Laboratorio de Biofísicoquímica.
110 ' Universidad Autónoma Metropolitana-Iztapalapa.
120 '
130 '
140 DIM T(30), F(30), A(5), SIGMAY(30), SIGMAA(5), DELTAA(5), FTEO(30)
141 '*****
142 '
143 ' Cuerpo Principal del Programa.
144 '
149 KEY OFF
150 GOSUB 1000 : 'Lee puntos experimentales (N, Ti, Fi)
160 GOSUB 1500 : 'Valores iniciales, constantes y datos
170 GOSUB 2000 : 'Convierte temperatura de C a K
180 GOSUB 2500 : 'Calcula RMS Y SIGMAY
181 NGL = N-NVAR : GOSUB 50000 : 'Calcula CHI
182 TITULO$ = "Ajuste de parámetros termodinámicos" : GOSUB 60000
190 PRINT SPACES(21);"DELTA Cp   Tm      DELTA Hm"
191 PRINT SPACES(20);"(cal/molK) (C) (cal/mol) CHI CUAD. RMS"
192 PRINT:PRINT"VALORES INICIALES";TAB(21):PRINT USING"#####.##";A(1);
193 PRINT TAB(31) : PRINT USING " ##.###";A(2)-K;:PRINT TAB(39):PRINT
  USING " #####.##";A(3);
194 PRINT TAB(52) : PRINT USING " .#####^^^^";CHI;RMS
199 ITE = -1
200 GOSUB 20000 : 'Realiza la primera iteración ajustando variables
210 LOCATE 9,1 : PRINT "VAL. ITERACION ANT.";TAB(21) : PRINT USING
  "#####.##";A(1);
211 PRINT TAB(31):PRINT USING" ##.###";A(2)-K; :PRINT TAB(39) :PRINT
  USING " #####.##";A(3);
212 PRINT TAB(52) : PRINT USING " .#####^^^^";CHI;RMS
220 ANT = CHI
230 GOSUB 20000 : 'Continúa iteraciones
240 D = ABS(ANT-CHI)
245 D = D/ANT
246 LOCATE23,1:PRINT"VARIACION EN CHI RESPECTO A ULTIMA ITERACION:";
247 PRINT USING " ###.#### ";D*100;
248 PRINT "%."
250 IF D>.001 GOTO 210 : 'Si mejoró menos de 0.1% deja de iterar
260 GOSUB 2500 : GOSUB 50000

```

```

269 '*****
270 '
280 '           Resultados
290 '
300 TITULO$ = "Resultados"
310 GOSUB 60000
320 PRINT TAB(30);"Delta Cp           Tm           Delta Hm"
330 PRINT TAB(29);"(cal/molK)         (C)           (cal/mol)"
340 PRINT:PRINT"ValorCalculado";TAB(29):PRINT USING "#####.###";A(1);
350 PRINT USING "    ##.###";A(2)-K;:PRINT USING "    #####.###";A(3)
360 LOCATE 9,1:PRINT"Sigma";TAB(29):PRINT USING"#####.###";SIGMAA(1);
370 PRINT USING"    ##.###";SIGMAA(2);:PRINT USING "    #####.###";
    SIGMAA(3)
380 LOCATE 13,1 : PRINT "Chi cuadrada =";CHI,"           RMS =";RMS
390 LOCATE 16,1 : PRINT "Máxima estabilidad a ";
400 TX = A(2)*EXP(-A(3)/(A(2)*A(1)))
410 G = A(3)*(1-(TX/A(2))) + A(1)*(TX-A(2)-TX*LOG(TX/A(2)))
420 PRINT TX-K;"grados centígrados,"
430 LOCATE 18,22 : PRINT "con Delta G =";G;"cal/mol."
440 DH1 = A(3)+A(1)*(383.15-A(2))
450 LOCATE 21,1:PRINT "Delta H extrapolada a 110 grados centígrados =";
    DH1;" cal/mol."
460 PRINT : INPUT "Cont. ";R$
470 PRINT : PRINT : PRINT " #           T           FD           FTEO           DELTA G           DELTA
    G(CP)"
480 PRINT : FOR I = 1 TO N
490   PRINT USING "##           ";I; : PRINT USING "##.###           ";T(I)-K;F(I); :
492   DG = A(3)-T(I)*(A(3)/A(2))+A(1)*(T(I)-A(2)-T(I)*LOG(T(I)/A(2)))
494   DGC = -R*T*LOG(FTEO(I)/(1-FTEO(I)))
500   PRINT USING"##.####           ";FTEO(I);:PRINT USING"#####.###           ";DGC;DG
510 NEXT I
520 INPUT "Cont. ";R$
530 LPRINT TAB(30);"Delta Cp           Tm           Delta Hm"
540 LPRINT TAB(29);"(cal/molK)         (C)           (cal/mol)"
550 LPRINT:LPRINT"ValorCalculado";TAB(29):PRINT USING"#####.###";A(1);
560 LPRINT USING"    ##.###";A(2)-K;:LPRINT USING"    #####.###";A(3)
570 LPRINT :LPRINT : LPRINT "Sigma";TAB(29):LPRINT USING "#####.###";
    SIGMAA(1);
580 LPRINT USING "    ##.###";SIGMAA(2);:LPRINT USING"    #####.###";
    SIGMAA(3)
590 LPRINT : LPRINT : LPRINT "Chi cuadrada =";CHI,"           RMS =";RMS
600 LPRINT : LPRINT "Máxima estabilidad a ";
610 LPRINT TX-K;" grados centígrados,"
620 LPRINT "           con Delta G =";G;"cal/mol."
630 LPRINT : LPRINT "ΔH extrapolada a 110 grados centígrados=";DH1;"cal/mol."
640 LPRINT : LPRINT : LPRINT " #           T           FD           FTEO           DELTA G
    DELTA G(CP)"
650 LPRINT : FOR I = 1 TO N
660   LPRINT USING "##           ";I;:LPRINT USING "##.###           ";T(I)-K;F(I); :
662   DG = A(3)-T(I)*(A(3)/A(2))+A(1)*(T(I)-A(2)-T(I)*LOG(T(I)/A(2)))
664   DGC = -R*T*LOG(FTEO(I)/(1-FTEO(I)))
670   LPRINT USING"##.####           ";FTEO(I);:LPRINT USING"#####.###           ";DGC;DG
680 NEXT I

```

```

685 KEY ON
690 END
700 ' FIN DEL PROGRAMA

710 ' *****
1000 '
1010 '     Carga datos experimentales
1020 '
1030 TITULO$ = "Puntos Experimentales."
1040 GOSUB 60000
1200 DATA 6 : ' pH 5.25 con pts. suaves
1210 DATA 20,30,40,50,60,70
1220 DATA .2,.3,.4,.5,.6,.7
1300 READ N
1310 FOR I = 1 TO N
1320   READ T(I)
1330 NEXT I
1340 FOR I = 1 TO N
1350   READ F(I)
1360 NEXT I
1370 PRINT "      #      TEMP(C)      FD" : PRINT
1375 S = 1 : Q = 0
1380 FOR I = 1 TO N
1390   LOCATE I+6-Q,S : PRINT USING "      ##";I;
1400   PRINT USING "      ##.###";T(I);
1410   PRINT USING "      #.####";F(I)
1415   IF I = 15 THEN LOCATE 6,1 : S = 40 : Q = 15
1420 NEXT I
1430 LOCATE 23,1 : INPUT "Están todos los datos correctos? (S/N)";R$
1440 IF R$ = "s" OR R$ = "S" THEN RETURN
1450 IF R$ <> "n" AND R$ <> "N" THEN GOTO 1430
1460 PRINT:PRINT:PRINT"MODIFIQUE VALORES EN LAS SIGUIENTES LINEAS : "
1470 LIST 1200-1299
1480 END

1500 '
1510 '     Valores iniciales y constantes
1520 '
1530 A(1) = 13.9      : DELTAA(1) = 30      : 'DELTA Cp
1540 A(2) = 49.7      : DELTAA(2) = .1      : 'Tm
1550 A(3) = 8793.5    : DELTAA(3) = 20      : 'DELTA H EVALUADO A LA Tm
1560 K = 273.15      : R = 1.98717
1570 NVAR = 3
1590 RETURN

2000 '
2010 '     Convierte temperaturas centígradas a Kelvin
2020 '
2030 FOR I = 1 TO N
2040   T(I) = T(I)+K      : 'Temperaturas de los puntos exp.
2050 NEXT I
2060 A(2) = A(2)+K      : 'Valor inicial de Tm.
2070 RETURN

```

```

2500 '
2510 '   Calcula RMS y SIGMAY
2520 '
2530 RMS = 0
2540 SIGMAY = 0
2550 FOR I = 1 TO N
2560   T = T(I)
2570   GOSUB 10000
2580   RMS = RMS + (F(I)-FD)^2
2590 NEXT I
2600 SIGMAY = RMS/(N-NVAR)
2610 RMS = RMS/N
2620 RMS = SQR(RMS)
2630 FOR I = 1 TO N
2640   SIGMAY(I) = .05   : 'Fija incertidumbre en FD exp a 0.05
2650 NEXT I
2660 RETURN

10000 '
10010 '   Evalua la función doble exponencial calculando FTEO
10020 '
10030 CO = ((A(3)/(R*A(2))) - (A(1)/R))
10040 AR = CO*(1-(A(2)/T))
10050 NUM = AR+(A(1)/R)*LOG(T/A(2))
10060 FD = EXP(NUM)
10070 FD = FD/(1+FD)
10080 RETURN

20000 '
20010 '   Realiza la iteración ajustando las tres variables
20020 '
20030 LOCATE 15,1 : PRINT "DELTA"
20031 LOCATE 11,1 : PRINT "VALOR ACTUAL"
20032 LOCATE 17,1 : PRINT "AJUSTANDO"
20033 LOCATE 19,1 : PRINT "NUM. DE AJUSTES"
20034 LOCATE 21,1 : PRINT "ITERACIONES :";
20035 ITE = ITE+1 : PRINT USING "###";ITE
20040 CHI = 0
20050 IF NGL<=0 GOTO 20770
20060 FOR J = 1 TO NVAR
20064   F = 0
20065   LOCATE 17,19 : IF J = 1 THEN PRINT "DELTA Cp"
20066   IF J = 2 THEN PRINT SPACE$(15);"Tm"
20067   IF J = 3 THEN PRINT SPACE$(25);"DELTA Hm"
20070   FOR I = 1 TO N
20080     T = T(I)
20090     GOSUB 10000 : 'Evalúa FD
20100     FTEO(I) = FD
20110   NEXT I
20120   GOSUB 2500
20130   GOSUB 50000
20135   GOSUB 30000:'Muestra valores actuales del parámetro, CHI y RMS
20140   CHI1 = CHI

```

```

20150 F = F+1
20160 DELTA = DELTAA(J)
20162 LOCATE 15,18 : IF J = 2 THEN LOCATE 15,30
20163 IF J = 3 THEN LOCATE 15,42
20165 PRINT USING "##.###^";DELTA
20170 A(J) = A(J)+DELTA
20180 FOR I = 1 TO N
20190     T = T(I)
20200     GOSUB 10000
20210     FTEO(I) = FD
20220 NEXT I
20230 GOSUB 2500
20240 GOSUB 50000
20245 GOSUB 30000 : 'Valores actuales del parámetro, CHI y RMS
20250 CHI2 = CHI
20260 D = CHI1-CHI2
20270 IF D = 0 GOTO 20170
20280 IF D>0 GOTO 20450
20290 '
20300 ' Invierte dirección de búsqueda al aumentar CHI
20310 '
20320 DELTA = -DELTA
20325 LOCATE 15,(J+1)*10-4+J*2 : PRINT "-"
20330 A(J) = A(J)+DELTA
20340 FOR I = 1 TO N
20350     T = T(I)
20360     GOSUB 10000
20370     FTEO(I) = FD
20380 NEXT I
20390 TEMP = CHI1
20400 CHI1 = CHI2
20410 CHI2 = TEMP
20420 '
20430 ' Aumento de A(J) hasta que CHI aumente
20440 '
20450 F = F+1
20460 A(J) = A(J)+DELTA
20470 FOR I = 1 TO N
20480     T = T(I)
20490     GOSUB 10000
20500     FTEO(I) = FD
20510 NEXT I
20520 GOSUB 2500
20530 GOSUB 50000
20535 GOSUB 30000 : 'Valores actuales del parámetro, CHI y RMS
20540 CHI3 = CHI
20550 D = CHI3-CHI2
20560 IF D>=0 OR F>15 GOTO 20600
20570 CHI1 = CHI2
20580 CHI2 = CHI3
20590 GOTO 20450
20600 '
20610 ' Busca el mínimo de la parábola con tres puntos

```

```

20620 '
20630 DELTA = DELTA*(1/(1+(CHI1-CHI2)/(CHI3-CHI2))+.5)
20640 A(J) = A(J)-DELTA
20645 SIGMAA(J) = DELTAA(J)*SQR(2/(NGL*(CHI3-2*CHI2+CHI1)))
20650 DELTAA(J) = DELTAA(J)*(F-1)/3
20660 NEXT J
20670 '
20680 ' Calcula valores con los nuevos parámetros
20690 '
20700 FOR I = 1 TO N
20710 T = T(I)
20720 GOSUB 10000
20730 FTEO(I) = FD
20740 NEXT I
20750 GOSUB 2500
20760 GOSUB 50000
20770 RETURN

30000 '
30010 'Valores actuales de los parámetros a la pantalla
30020 '
30030 IF J = 1 THEN LOCATE 11,21 : PRINT USING "#####.##";A(1)
30040 IF J = 2 THEN LOCATE 11,31 : PRINT USING " ##.###";A(2)-K
30050 IF J = 3 THEN LOCATE 11,39 : PRINT USING " #####.##";A(3)
30060 LOCATE 11,52 : PRINT USING " .#####^???" ;CHI;RMS
30070 LOCATE 19,(J+1)*10+2*J : PRINT USING "##";F+1
30080 RETURN

50000 '
50010 ' Evalúa CHI CUADRADA
50020 '
50030 CHI = 0
50040 IF NGL<=0 GOTO 50100
50050 FOR I = 1 TO N
50060 PESO = 1/(SIGMAA(I)^2)
50070 CHI = CHI + PESO*(F(I)-FTEO(I))^2
50080 NEXT I
50090 CHI = CHI/NGL
50100 RETURN

60000 '
60010 ' Subrutina de título centrado en video inverso
60020 '
60030 CLS : COLTEXT = 5
60040 COLOR 0,COLTEXT
60050 LOCATE 1,1
60060 PRINT SPACE$(79);
60070 LOCATE 1,((79-(LEN(TITULO$)))/2)
60080 PRINT TITULO$
60090 LOCATE 4,1
60100 COLOR COLTEXT,0
60110 RETURN

```

APENDICE II.

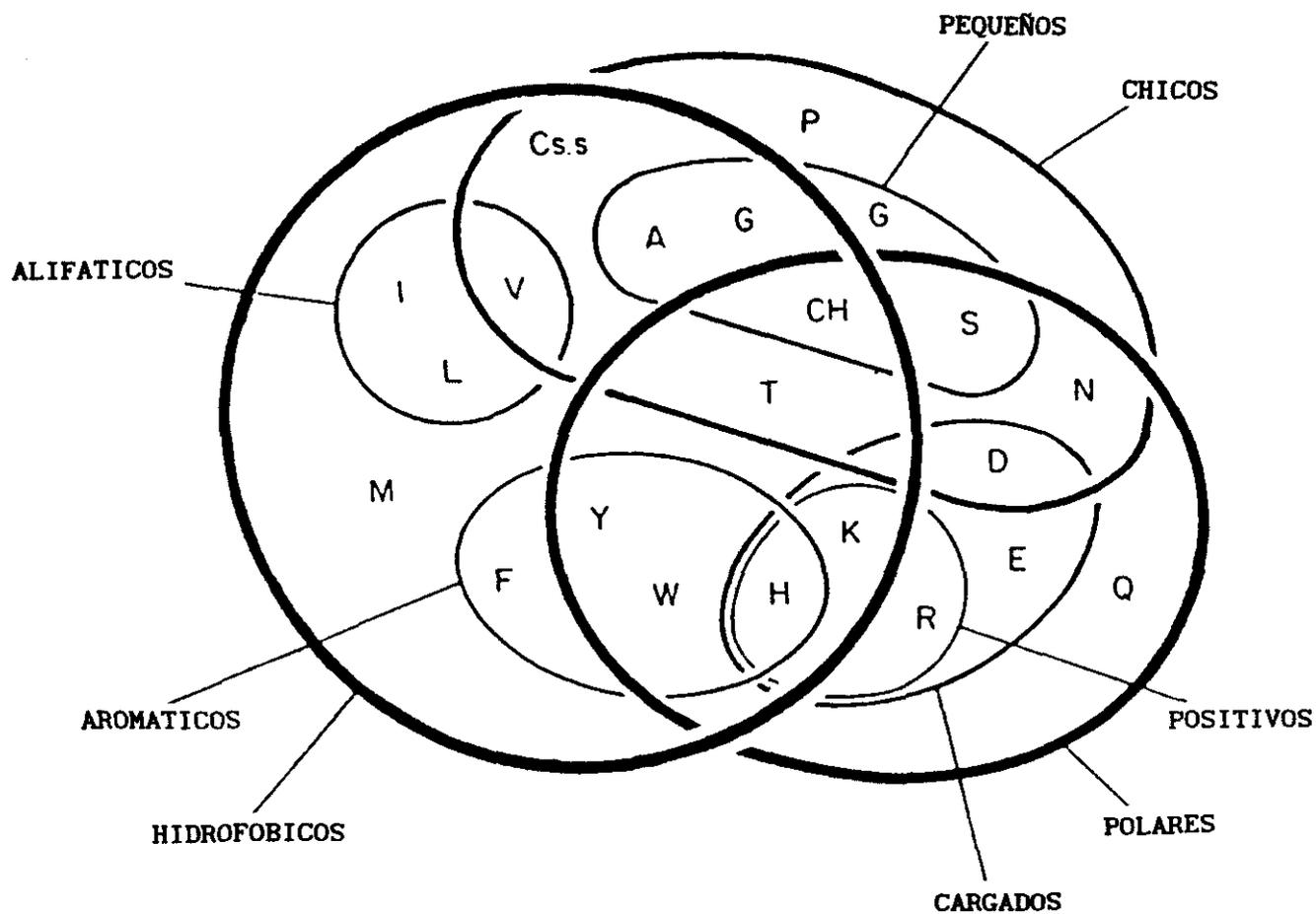
ABREVIATURAS Y SIMBOLOS.

Lipocalinas mencionadas en este trabajo

Abreviatura	Nombre	Referencia
GPA	Glicoproteína ácida α_1 humana u orosomucoide Human alpha-1-acid glycoprotein or orosomucoid	Dente <i>et al.</i> , 1987
PTB	Proteína transportadora de bilina Bilin binding protein	Huber <i>et al.</i> , 1987
INS	Insecticianina de <i>Manduca sexta</i> Insecticyanin from <i>Manduca sexta</i>	Holden <i>et al.</i> , 1987
BLG	β -Lactoglobulina β -Lactoglobulin	Papiz <i>et al.</i> , 1986
PTR	Proteína transportadora de retinol Retinol binding protein	Cowan <i>et al.</i> , 1990
AFR	Afrodisina Aphrodisin	Henzel <i>et al.</i> , 1988
A1M	Alfa-1 microglobulina Alpha-1 microglobulin, HC protein	Vetr y Gebhard, 1990
A2M	Alfa-2u globulina Alpha-2u-globulin	Unterman <i>et al.</i> , 1981
ALD	Apolipoproteína D Apolipoprotein D	Drayna <i>et al.</i> , 1986
C8G	Cadena γ del complemento 8 Eighth component of human complement, γ chain	Ng <i>et al.</i> , 1987
P14	Proteína 14 Placental protein 14	Julkunen <i>et al.</i> , 1988
PDA	Proteína dependiente de andrógeno Rat epididymal androgen dependent protein	Girotti <i>et al.</i> , 1992
POR	Proteína olfatoria de rana Frog olfactory protein, Bowman's gland protein	Lee <i>et al.</i> , 1987
POL	Proteína olfatoria de rata Rat olfactory binding protein	Dear <i>et al.</i> , 1991
PRO	Proteína de orina de ratón Mouse major urinary protein	Clark <i>et al.</i> , 1984
PTP	Proteína transportadora de pirazina Pyrazine binding protein	Tirindelli <i>et al.</i> , 1989
PUR	Purpurina Purpurin	Berman <i>et al.</i> , 1987

Abreviaturas de los aminoácidos

Aminoácido	Símbolos	Aminoácido	Símbolos
Alanina	ALA A	Asparagina	ASN N
Cisteína	CYS C	Prolina	PRO P
Acido aspártico	ASP D	Glutamina	GLN Q
Acido glutámico	GLU E	Arginina	ARG R
Fenilalanina	PHE F	Serina	SER S
Glicina	GLY G	Treonina	THR T
Histidina	HIS H	Valina	VAL V
Isoleucina	ILE I	Triptofano	TRP W
Lisina	LYS K	Tirosina	TYR Y
Leucina	LEU L	Acido	
Metionina	MET M	pirrolidoncarboxílico	PCA X



BIBLIOGRAFIA.

- Aqvist J, Sandblom P, Jones TA, Newcomer ME, van Gunsteren WF y Tapia O. (1986) *Molecular dynamics simulations of the holo and apo forms of retinol binding protein. Structural and dynamical changes induced by retinol removal.* J. Mol. Biol. 192: 593-604.
- Aki H y Yamamoto M. (1989) *Thermodynamics of the binding of phenothiazines to human plasma, human serum albumin and α_1 -acid glycoprotein: a calorimetric study.* J. Pharm. Pharmacol. 41: 674-679.
- Ali S y Clark AJ. (1988) *Characterization of the gene encoding ovine beta-lactoglobulin. Similarity to the genes for retinol binding protein and other secretory proteins.* J. Biol. Chem. 199: 415-426.
- Ananthanarayanan VS y Bigelow CC. (1969a) *Unusual difference spectra of proteins containing tryptophan. I. Studies with model compounds.* Biochemistry 8: 3717-3723.
- Ananthanarayanan VS y Bigelow CC. (1969b) *Unusual difference spectra of proteins containing tryptophan. II. Studies with proteins.* Biochemistry 8: 3723-3728.
- Andersen P, Andersen S y Brosstad F. (1980) *Reduced anti-heparin effect of desialyzed α_1 -acid glycoprotein.* Thromb. Res. 17: 865-887.
- Andersen P y Eika C. (1980) *Scand. J. Haematol.* 25: 202-204.
- Andrews LJ y Forster LS. (1972) *Protein difference spectra. Effect of solvent and charge on tryptophan.* Biochemistry 11: 1875-1879.
- Anfinsen CB. (1973) *Principles that govern the folding of protein chains.* Science 181: 223-230.
- Aubert JP y Locheux-Lefebvre MH. (1976) *Conformational study of α_1 -acid glycoprotein.* Arch. Biochem. Biophys. 175: 400-409.
- Baldwin RL. (1986) *Seeding protein folding.* Trends Biochem. Sci. 11: 6-9.
- Baron M, Norman DG y Campbell ID. (1991) *Protein modules.* Trends Biochem. Sci. 16: 13-17.
- Bauman P, Müller WE, Eap CB y Tillement J-P eds. (1989) *Alpha-1-acid glycoprotein: genetics, biochemistry, physiological functions and pharmacology.* Alan R. Liss, Inc, Nueva York.
- Bax A. (1989) *Two dimensional NMR and protein structure.* Annu. Rev. Biochem. 58: 223-256.
- Bell SC y Keyte JW. (1987) *Amino-acid sequence similarities.* Nature 329: 296.
- Bennett M y Schmid K. (1980) *Immunosuppression of human plasma α_1 -acid glycoprotein: Importance of the carbohydrate moiety.* Proc. Natl. Acad. Sci. 77: 6109-6113.
- Berman P, Gray P, Chen E, Keyser K, Ehrlich D, Karten H, LaCorbiere M, Esch F y Schubert D. (1987) *Sequence analysis, cellular localization, and expression of a neuroretina adhesion and cell survival molecule.* Cell 51: 135-142.

- Bernstein FC, Koetzle TF, Williams GJ, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T y Tasumi M. (1977) *The Protein Data Bank: a computer-based archival file for macromolecular structures*. J. Mol. Biol. 112: 535-542.
- Bevington PR (1969) *Data reduction and error analysis for the physical sciences*, pp 204-246, McGraw-Hill, Nueva York.
- Bezkorovainy A y Winzler RJ. (1961) *Isolation and characterization of an acidic α -glycoprotein*. Biochim. Biophys. Acta 49: 559-565.
- Black JA y Dixon GH. (1968) *Amino acid sequence of alpha chains of human haptoglobins*. Nature 218: 736-741.
- Boguski MS. (1992) *Computational sequence analysis revisited: new databases, software tools, and the research opportunities they engender*. J. Lipid Res. 33: 957-974.
- Brooks DE, Means AR, Wright EJ, Singh SP y Tiver KK. (1986) *Molecular cloning of the cDNA for two major androgen-dependent secretory proteins of 18.5 kilodaltons synthesized by the rat epididymis*. J. Biol. Chem. 26(11): 4956-4961.
- Busetta B y Hospital M (1982) *An analysis of the prediction of secondary structures*. Biochim. Biophys. Acta 701: 111-118.
- Cantor CR y Schimmel PR. (1980b) en *Biophysical Chemistry*, Parte III, pp 1075-1088, W.H. Freeman and Co., Nueva York.
- Cavaggioni A, Sorbi RT, Keen JN, Pappin DJC y Findlay JBC. (1987) *Homology between the pyrazine-binding protein from nasal mucosa and major urinary proteins*. FEBS Letters 212(2): 225-228.
- Chou PY y Fasman GD. (1974) *Prediction of protein conformation*. Biochemistry 13: 222-224.
- Chou PY y Fasman GD. (1978) *Prediction of the secondary structure of proteins from the amino acid sequence*. Adv. Enzymol. 47: 45-148.
- Chou PY y Fasman GD. (1979) *Predictions of β -turns*. Biophys. J. 26: 367-384.
- Chou YH y Hayman MJ. (1991) *Characterization of a member of the immunoglobulin gene superfamily that possibly represents an additional class of growth factor receptor*. Proc. Natl. Acad. Sci. USA 88: 4897-4901.
- Chothia C. (1984) *Principles that determine the structure of proteins*. Annu. Rev. Biochem. 53: 537-572.
- Chotia C. (1992) *One thousand families for the molecular biologist*. Nature 357: 543-544.
- Clark AJ, Clissolk PM, Al Shawi R, Beattie P y Bishop J. (1984) *Structure of mouse major urinary protein genes: different splicing configurations in the 3'-non-coding region*. EMBO J. 3: 1045-1052.
- Clure GM y Gronenborn AM. (1991) *Structures of larger proteins in solution: three and four dimensional heteronuclear NMR*. Science 252: 1390-1399.
- Cohn EJ, Strong LE, Hughes WL, Jr, Mulford DJ, Ashyworth JN, Melin M y Taylor HL. (1946) J. Am. Chem. Soc. 68: 459.

- Colantuoni V, Romano V, Bensi G, Santoro C, Constanzo F, Raugei G y Cortese R. (1983) *Cloning and sequencing of a full length cDNA coding for human retinol binding protein*. *Nucleic Acid Res.* 11(22): 7769-7776.
- Cooper R y Papaconstantinou J. (1986) *Evidence for the existence of multiple α_1 -acid glycoprotein genes in the mouse*. *J. Biol. Chem.* 261: 1849-1853.
- Cowan SW, Newcomer ME, Jones TA. (1990) *Crystallographic refinement of human serum retinol binding protein at 2Å resolution*. *Proteins Struct. Funct. Genet.* 8: 44-61.
- Creighton TE. (1983) *Proteins: Structures and molecular principles*. W. H. Freeman and Co. Nueva York.
- Crick FHC, Barnett L, Brenner S y Watts-Tobin RJ. (1961) *General nature of the genetic code for proteins*. *Nature* 192: 1227-1232.
- Davies DR, Padlan EA y Segal DM. (1975) *Three-dimensional structure of immunoglobulins*. *Annu. Rev. Biochem.* 44: 639-667.
- Dayhoff MO. (1978) *En Atlas of protein sequence and structure.*, 5 suppl. 3: 1-8. Nat. Biomed. Res. Found., Washington D.C.
- Dear TN, Campbell K y Rabbitts TH. (1991) *Molecular cloning of putative odorant-binding and odorant-metabolizing proteins*. *Biochemistry* 30: 10376-10382.
- DeLeve LD y Piafzky KM. (1981) *Clinical significance of plasma binding of basic drugs*. *Trends Pharmacol. Sci.* 2: 283-284.
- DeVos AM, Ultsch M y Kossiakoff AA. (1992) *Human growth hormone and extracellular domain of its receptor: crystal structure of the complex*. *Science.* 255: 306-312.
- Dente L, Ciliberto G y Cortese R. (1985) *Structure of the human α_1 -acid glycoprotein gene: sequence homology with other human acute phase protein genes*. *Nucleic Acids Res.* 13: 3941-3952.
- Dente L, Pizza MG, Metspalu A y Cortese R (1987) *Structure and expression of the genes coding for human α_1 -acid glycoprotein*. *EMBO J.* 6: 2289-2296.
- Doi E y Jirgensons B. (1970) *Circular dichroism studies on the acid denaturation of γ -immunoglobulin G and its fragments*. *Biochemistry* 9(5): 1066-1073.
- Donovan JW. (1969) *Changes in ultraviolet absorption produced by alteration of protein conformation*. *J. Biol. Chem.* 244: 1961-1967.
- Doolittle RF. (1981) *Similar amino acid sequences: chance or common ancestry?* *Science* 214: 149-159
- Doolittle RF. (1986) *Of URFS and ORFS. A primer on how to analyse derived amino acid sequences*. University Science Books. Mill Hill, California.
- Doolittle RF. (1989) *Redundancies in protein sequences*. En Prediction of protein structure and the principles of protein conformation. Fasman GD. Plenum Press, Nueva York.
- Doolittle RF, ed. (1990) *Molecular evolution: Computer analysis of protein and nucleic acid sequences*. *Methods Enzymol.* 183.

- Drayna D, Fielding C, McLean J, Baer B, Castro G, Chen E, Comstock L, Henzel W, Kohr W, Rhee L, Wion K y Lawn R. (1986) *Cloning and expression of human apolipoprotein D cDNA*. J. Biol. Chem. 261(35): 16535-16539.
- Dufour E, Marden MC y Haertlé T. (1990) *β -lactoglobulin binds retinol and protoporphyrin IX at two different binding sites*. FEBS Letters 277(1): 223-226.
- Edelman GM. (1970) *The covalent structure of a human γ G-immunoglobulin. XI. Functional Implications*. Biochemistry 9(16): 3197-3205.
- Edelman GM. (1970) *Antibody structure and molecular immunology*. Science 180: 830-840.
- Edmundson AB, Ely KR, Girling RL, Abola EE, Schiffer M, Westholm FA, Fausch MD y Deutsch HF. (1974) *Binding of 2,4-dinitrophenil compounds and other small molecules to a crystalline λ -type Bence-Jones dimer*. Biochemistry 13(18): 3816-3827.
- Eisenberg D, Weiss RM y Terwilliger TC. (1982) *The helical hydrophobic moment: a measure of the amphiphilicity of a helix*. Nature 299: 371-374.
- Elwell ML y Schellman JA. (1977) *Stability of phage T4 lysozymes I. Native properties and thermal stability of wild type and two mutant lysozymes*. Biochim. Biophys. Acta 494: 367-383.
- Emura J, Ikeneka T, Collins JH y Schmid K. (1971) *The constant and variable regions of the carboxyl-terminal CNBr fragment of α_1 -acid glycoprotein*. J. Biol. Chem. 246: 7821-7823.
- Epp O, Lattman EE, Schiffer M, Huber R y Palm W. (1975) *The molecular structure of a dimer composed of the variable portions of the Bence-Jones protein REI refined at 2.0Å resolution*. Biochemistry 14(22): 4943-4952.
- Fasman GD. (1989) *Prediction of protein structure and the principles of protein conformation*. Plenum Press. Nueva York.
- Feng DF, Johnson MS y Doolittle RF. (1985) *Aligning amino acid sequences: comparison of commonly used methods*. J. Mol. Evol. 21: 112-125.
- Friedman ML, Schlueter KT, Kirley TL y Halsall HB. (1985) *Fluorescence quenching of human orosomucoid. Accessibility to drugs and small quenching agents*. Biochem. J. 232: 863-867.
- Franzblau C, Schmid K, Faris B, Beldekas J, Garvin P, Kagan HM y Baum BJ. (1976) *The interaction of collagen with α_1 -acid glycoprotein*. Biochim. Biophys. Acta 427: 302-314.
- Gaboriaud C, Bissery V, Benchetrit T y Mornon JP. (1987) *Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acid sequences*. FEBS Lett. 224: 149-155.
- Gahmberg CG y Andersson LC. (1978) *Leukocyte surface origin of human α_1 -acid glycoprotein (orosomucoid)*. J. Exp. Med. 148: 507-521.

- Gahmberg CG y Andersson LC. (1982) *GP52: A membrane form of α 1-acid glycoprotein (orosomuroid) on both T and B cells*. En *The glycoconjugates III*. p 255. Academic Press, Nueva York.
- Ganguly M, Carnighan RH y Westphal U. (1976) *Steroid-protein interactions. XIV. Interaction between human α 1-acid glycoprotein and progesterone*. *Biochemistry* 6(9): 2803-2814.
- Garnier J, Osguthorpe DJ y Robson B. (1978) *Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins*. *J. Mol. Biol.* 120: 97-120.
- Ghèlis C y Yon J. (1982) *Protein Folding*. Academic Press, Nueva York.
- Gierasch LM y King J. (1991) *Protein folding. Desciphering the second half of the genetic code*. American Asociation for the Advancement of Science. Washington, D.C.
- Girotti M, Jones R, Emery DC, Chia W y Hall L. (1992) *Structure and expression of the rat epididymal secretory protein I gene*. *Biochem. J.* 281: 203-210.
- Godovac-Zimmermann J. (1988) *The structural motif of β -lactoglobulin and retinol-binding protein: a basic framework for binding and transport of small hydrophobic molecules?* *TIBS* 13: 64-66.
- Goldberg ME. (1985) *The second traslation of the genetic message: protein folding and assemmbly*. *Trends Biochem. Sci.* 10: 388-391.
- Goolkasian DL, Slaughter RL, Edwards DJ y Lalka D. (1983) *Displacement of lidocaine from α 1-acid glycoprotein binding sites by basic drugs*. *Eur. J. Clin. Pharmacol.* 25: 413-417.
- Goto Y y Hamaguchi K. (1987) *Role of amino-terminal residues in the folding of the constant fragment of the immunoglobulin light chain*. *Biochemistry* 26: 1879-1884.
- Green RF y Pace CN. (1974) *Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, α -chymotrypsin and β -lactoglobulin*. *J. Biol. Chem.* 249: 5388-5393.
- Gu W y Brady JW. (1992) *Molecular dynamics simulations of the whey protein β -lactoglobulin*. *Protein Eng.* 5(1): 17-27.
- Haefliger JA, Peitsch MC, Jenne DE y Tschopp J. (1991) *Structural and functional characterization of complement C8 γ , a member of the lipocalin protein family*. *Molecular Immunology* 28: 123-131.
- Halsall HB y Kirley TL. (1981) *Observations on the thermostability of orosomuroid*. *Biochem. Biophys. Res. Commun.* 102(2): 761-765.
- Halsall HB y Kirley TL. (1982) *The denaturation of orosomuroid*. *Arch. Biochem. Biophys.* 216: 392-399.
- Hellerstein MK, Sasak V y Munro HN. (1985) *Isolation of alpha-1-acid glycoprotein from human plasma using high-performance liquid chromatography*. *Anal. Biochem.* 146: 360-371.
- Hennessey JP y Johnson WC, Jr. (1981) *Information content in the circular dichroism of proteins*. *Biochemistry* 20: 1085-1094.

- Hennessey JP y Johnson WC, Jr. (1982) *Experimental errors and their effect on analyzing circular dichroism spectra of proteins*. Anal. Biochem. 125: 177-188.
- Henzel WJ, Rodríguez H, Singer AG, Stults Jt, Macrides F, Agosta WC y Niall H. (1988) *The primary structure of aphrodisin*. J. Biol. Chem. 263(32): 16682-16687.
- Herskovits TT y Sorensen M. (1968a) *Studies of the location of tyrosyl and tryptophyl residues in proteins. I. Solvent perturbation data of model compounds*. Biochemistry 7: 2523-2532.
- Herskovitz TT y Sorensen M. (1968b) *Studies of the location of tyrosyl and tryptophyl residues in proteins. II. Applications of model data to solvent perturbation studies of proteins rich in both tyrosine and triptophan*. Biochemistry 7: 2533-2542.
- Hill RL, Delaney R, Fellows RE y Lebovitz HE. (1966) *The evolutionary origins of the immunoglobulins*. Proc. Natl. Acad. Sci. USA 56: 1762-1769.
- Higgins DG y Sharp PM. (1988) *CLUSTAL: a package for performing multiple sequence alignment on a microcomputer*. Gene 73: 237-244.
- Higgins DG y Sharp PM. (1989) *CABIOS 5*: 151-153.
- Holden HM, Rypniewski WR, Law JH y Rayment I. (1987) *The molecular structure of insecticyanin from the tobacco hornworm Manduca sexta L. at 2.6 Å resolution*. EMBO Journal 6(6): 1565-1570.
- Hood L y Prahl J. (1971) *The immune system: A model for differentiation in higher organisms*. Adv. Immunol. 14: 291-351.
- Hoop TP y Woods KR. (1981) *Prediction of protein antigenicity determined from amino acid sequence*. Proc. Natl. Acad. Sci. USA 78: 3824-3828.
- Huber R, Schneider M, Epp O, Mayr I, Messerschmidt A, Pflugrath J y Kayser H. (1987a) *Crystallization, crystal structure analysis and preliminary molecular model of the bilin binding protein from the insect Pieris brassicae*. J. Mol. Biol. 195: 423-434.
- Huber R, Schneider M, Mayr I, Müller R, Deutzmann R, Suter F, Zuber H, Falk H y Kayser H. (1987b) *Molecular structure of the bilin binding protein (BBP) from Pieris brassicae after refinement at 2.0 Å resolution* J. Mol. Biol. 198: 499-513.
- Ikenaka T, Ishiguro M, Emura J, Kaufmann H, Isemura S, Bauer W y Schmid K. (1972) *Isolation and partial characterization of the cyanogen bromide fragments of α 1-acid glycoprotein and the elucidation of the amino acid sequence of the carboxyl-terminal cyanogen bromide fragment*. Biochemistry 11(20): 3817-3829.
- Jeanloz RW. (1972) en *Glycoproteins* (A. Gottschalk, ed.), Part A p. 607. American Elsevier, New York.
- Jerne NK. (1967) *Antibodies and learning: selection versus instruction*. en *The neurosciences: A study program*. (Quarton GC. et al., eds.) Rockefeller University Press.

- Jukes TH. (1978) *The amino acid code*. Adv. Enzymol. 47: 375-432.
- Julkunen M, Seppälä M y Jänne OA. (1988) Complete amino acid sequence of human placental protein 14: A progesterone-regulated uterine protein homologous to β -lactoglobulins. Proc. Natl. Acad. Sci. USA 85: 8845-8849.
- Kabsch W y Sander C. (1983) How good are predictions of protein secondary structure? FEBS Letters 155: 179-182.
- Kálal P y Kalous V. (1984) The tyrosyl and tryptophyl residues of orosomucoid studied by difference and derivative spectrophotometry. Coll. Czech. Chem. Commun. 49: 165-169.
- Kaminogawa S, Shimizu M, Ametani A, Hattori M, Ando O, Hachimura S, Nakamura Y, Totsuka M y Yamauchi K. (1989) Monoclonal antibodies as probes for monitoring the denaturation process of bovine β -lactoglobulin. Biochim. Biophys. Acta 998: 50-56.
- Karplus PA y Schulz GE. (1985) Prediction of chain flexibility in proteins. Naturewissenschaften 72: 212-213.
- Kirley TL, Sprague ED y Halsall HB. (1982) The binding of spin-labeled propranolol and spin-labeled progesterone by orosomucoid. Biophys. Chem. 15: 209-216.
- Kim PS y Baldwin RL. (1982) Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. Annu. Rev. Biochem. 51: 459-489.
- Kuriyan J, Krishna TSR, Wong L, Guenther B, Pahler A, Williams Jr. CH y Model P. (1991) Convergent evolution of similar function in two structurally divergent enzymes. Science 352: 172-174.
- Kurosky A, Barnett DR, Lee TH, Touchstone B, Hay RE, Arnott MS, Bowman BH y Fitch WM. (1980) Covalent structure of human haptoglobin: a serine protease homolog. Proc. Natl. Acad. Sci. USA. 77(6): 3388-3392.
- Kute T y Westphal U. (1976) Steroid-protein interactions. XXXIV. Chemical modification of α 1-acid glycoprotein for characterization of the progesterone binding site. Biochim. Biophys. Acta 420: 195-213.
- Kyte J y Doolittle RF. (1982) A simple method for displaying the hydrophobic character of a protein. J. Mol. Biol. 157: 105-132.
- Labhardt AM. (1982) Secondary structure in ribonuclease. I. Equilibrium folding transitions seen by amide circular dichroism. J. Mol. Biol. 157: 331-355.
- Lapanje S, Poklar N (1989) Calorimetric and circular dichroic studies of the thermal denaturation of β -lactoglobulin. Biophys. Chem. 34: 155-162.
- Laurent BC, Nilson MHL, Båvik CO, Jones TA, Sundelin J y Peterson PA. (1985) Characterization of the rat retinol-binding protein gene and its comparison to the three-dimensional structure of the protein. J. Biol. Chem. 260(21): 11476-11480.
- Law JH y Wells MA. (1989) Insects as biochemical models. J. Biol. Chem. 264(28): 16335-16338.

- Lee K-H, Wells RG y Reed RR. (1987) *Isolation of an olfactory cDNA: Similarity to retinol-binding protein suggests a role in olfaction.* Science 235: 1053-1056.
- Levitt M. (1978) Biochemistry 17: 4277-4285.
- Levitt M y Chothia C. (1976) *Structural patterns in globular proteins.* Nature 261: 552-558.
- Liao Y-CJ, Taylor JM, Vannice JL, Clawson GA y Smuckler EA. (1985) *Structure of the rat α -acid glycoprotein gene.* Molec. Cel. Biol. 5: 3634-3639.
- Lipman DJ y Pearson WR. (1985) *Rapid and sensitive protein similarity searches.* Science 227: 1435-1441.
- López C, Grubb A, Soriano F y Méndez E. (1981) *The complete amino acid sequence of human complex forming glycoprotein heterogeneous in charge (Protein HC).* Biochem. Biophys. Res. Commun. 103(3): 919-925.
- Lüthy R, Bowie JU y Eisenberg D. (1992) *Assessment of protein models with three dimensional profiles.* Nature 356: 83-885
- Manavalan P y Johnson C. (1983) *Sensitivity of circular dichroism to protein tertiary structure class.* Nature 305: 831-832.
- Malinowski FR y Howery DG. (1980) *Factor Analysis in Chemistry*, pp 23-58, Wiley, Nueva York.
- Marchalonis JJ, Vasta GR, Warr GW y Barker WC. (1984) *Probing the boundaries of the extended immunoglobulin family of recognition molecules: jumping domains, convergence and minigenes.* Immunology Today 5(5): 133-142.
- Marquart M, Deisenhofer, Huber R y Palm W. (1980) *Crystallographic refinement and atomic models of the intact immunoglobulin molecule Kol and its antigen-binding fragment at 3.0 Å and 1.9 Å resolution.* J. Mol. Biol. 141: 369-391.
- Mayo SL, Olafson BD y Goddard III WA. (1990) *Dreiding: A generic force field for molecular simulations.* J. Phys. Chem. 94: 8897-8909.
- McPherson A, Friedman ML y Halsall HB. (1984) *Crystallization of α -acid glycoprotein.* Biochem. Biophys. Res. Commun. 124(2): 619-624.
- Monaco HL y Zanotti G. (1992) *Three dimensional structure and active site of three hydrophobic molecule binding proteins with significant amino acid sequence similarity.* Biopolymers 32: 457-465.
- Monaco HL, Zanotti G, Spadon P, Bolognesi M, Sawyer L y Eliopoulos EE. (1987) *Crystal structure of the trigonal form of bovine beta-lactoglobulin and of its complex with retinol at 2.5 Å resolution.* J. Mol. Biol. 197: 695-706.
- Morris AL, McArthur MW, Hutchinson EG y Thornton JM. (1992) *Stereochemical quality of protein structure coordinates.* Proteins Struct. Funct. Genet. 12: 345-364.
- Myers EW y Miller W. (1988) CABIOS 4: 11-17.

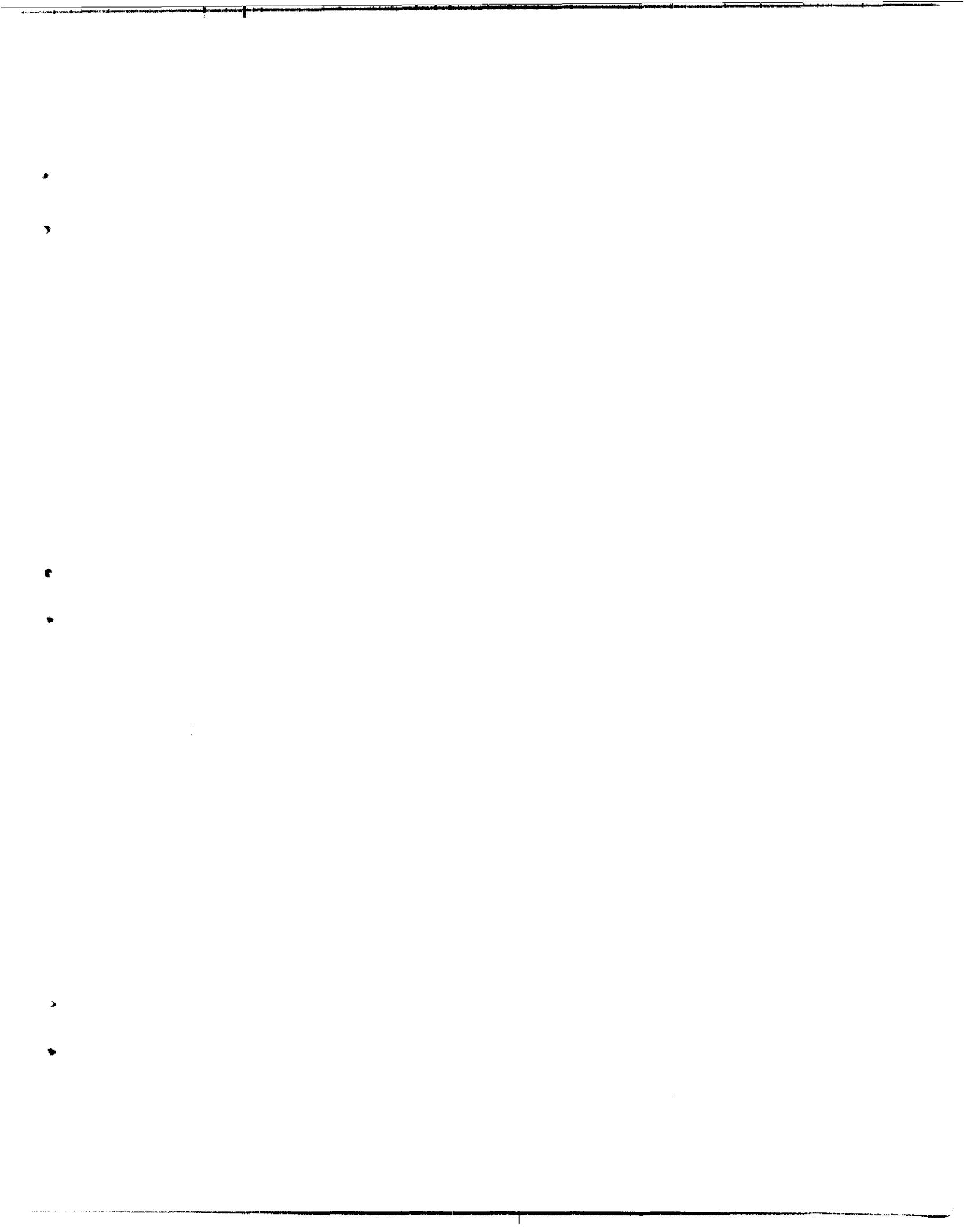
- Nagata A, Suzuki Y, Igarashi M, Eguchi N, Toh H, Urade Y y Hayaishi O. (1991) *Human brain prostaglandin D synthase has been evolutionary differentiated from lipophilic-ligand carrier proteins*. Proc. Natl. Acad. Sci. USA. 88: 4020-4024.
- Nakashima H, Nishikawa K y Ooi T. (1986) *The folding type of a protein is relevant to the amino acid composition*. J. Biochem. 99: 153-162.
- Needleman SB y Wunsch CD. (1970) *A general method applicable to the search for similarities in the amino acid sequence of two proteins*. J. Mol. Biol. 48: 443-453.
- Newcomer ME, Liljas A, Sundelin J, Rask L y Peterson PA. (1983) *Crystallization of and preliminary X-ray data for the plasma retinol-binding protein*. J. Biol. Chem. 259(8): 5230-5231.
- Ng S, Rao AG, Howard OMZ y Sodetz JM. (1987) *The eighth component of human complement: Evidence that it is an oligomeric serum protein assembled from products of three different genes*. Biochemistry 26: 5229-5233.
- Nillson L y Karplus M (1986) J. Comput. Chem 7: 591.
- Nishikawa K y Ooi T. (1986) *Radial locations of amino acid residues in a globular protein: correlation with the sequence*. J. Biochem. 100: 1043-1047.
- Novotny J y Auffray C. (1984) *A program for prediction of protein secondary structure from nucleic sequence data: Application to histocompatibility antigens*. Nucleic Acids Res. 12: 243-255.
- Nozaki Y. (1972) *The preparation of guanidine hydrochloride*. Methods Enzymol. 27: 43-50.
- Pace CN. (1975) *The stability of globular proteins*. CRC Crit. Rev. Biochem. 3: 1-43.
- Papiz MZ, Sawyer L, Eliopoulos EE, North ACT, Findlay JBC, Sivaprasadarao R, Jones TA, Newcomer ME y Kraulis PJ. (1986) *The structure of β -lactoglobulin and its similarity to plasma retinol-binding protein*. Nature 324: 383-385.
- Perczel A, Hollosi M, Tusnady G y Fasman GD. (1991) *Convex constraint analysis: a natural deconvolution of circular dichroism curves of proteins*. Protein Eng. 4: 669-679.
- Perczel A, Park K y Fasman GD. (1992) *Deconvolution of the circular dichroism spectra of proteins: The circular dichroism spectra of the antiparallel β -sheet in proteins*. Proteins. Struct. Funct. Genet. 13: 57-69.
- Pervaiz S y Brew K. (1985) *Homology of β -lactoglobulin, serum retinol-binding protein, and protein HC*. Science 228: 335-337.
- Pervaiz S y Brew K. (1987) *Homology and structure-function correlations between α_1 -acid glycoprotein and serum retinol-binding protein and its relatives*. FASEB J. 1: 209-214.
- Petsko GA. (1991) *Enzyme evolution. Déjà vu all over again*. Science 352: 104-105.

- Pevsner J, Reed RR, Feinstein PG y Snyder SH. (1988) *Molecular cloning of odorant-binding protein: member of a ligand carrier family*. Science 241: 336-339.
- Porter RR. (1973) *Structural studies of immunoglobulins*. Science 180: 713-716.
- Privalov PL. (1979) *Stability of proteins*. Adv. Protein Chem. 33: 167-241.
- Privalov PL y Gill SJ. (1988) *Stability of protein structure and hydrophobic interaction*. Adv. Protein Chem. 39: 191-234.
- Privalov PL. (1989) *Thermodynamic problems of protein structure*. Annu. Rev. Biophys. Biophys. Chem. 18: 47-69.
- Ptitsyn OB. (1987) *Protein folding: Hypotheses and experiments* J. Protein Chem. 6: 273-293.
- Ramakrishnan C y Ramachandran GN (1965) Biophys. J. 5: 909-933.
- Ray BK y Ray A. (1991) *Molecular cloning and nucleotide sequence of complementary DNA encoding rabbit α_1 -acid glycoprotein*. Biochem. Biophys. Res. Commun. 178: 507-513.
- Ray BK y Ray A. (1992) *Cloning and structural characterization of a rabbit genomic DNA for α_1 -acid glycoprotein*. Biochem. Biophys. Res. Commun. 185: 69-77.
- Reinke R y Feigelson P. (1985) *Rat α_1 -acid glycoprotein. Gene sequence and regulation by glucocorticoids in transfected L-cells*. J. Biol. Chem. 260(7): 4397-4403.
- Ricca GA y Taylor JM. (1981) *Nucleotide sequence of rat α_1 -acid glycoprotein messenger RNA*. J. Biol. Chem. 256(21): 11199-11202.
- Riley CT, Barbeau BK, Keim PS, Kézdy FJ, Heinrikson RL y Law JH. (1984) *The covalent protein structure of insecticyanin, a blue biliprotein from the hemolymph of the tobacco hornworm, Manduca sexta L*. J. Biol Chem. 259(21): 13159-13165.
- Richardson JS. (1977) *β -Sheet topology and the relatedness of proteins*. Nature (London) 268: 495-500.
- Richardson JS. (1981) *The anatomy and taxonomy of protein structure*. Adv. Protein Chem. 34: 167-339.
- Rojo-Domínguez A. (1988) *Determinación de los parámetros termodinámicos de la estabilidad de la glicoproteína ácida- α_1 y su implicación en las características estructurales de la molécula*. Tesis de maestría. Universidad Autónoma Metropolitana. Unidad Iztapalapa, México.
- Rojo-Domínguez A, Zubillaga-Luna RA y Hernández-Arana A. (1989) *Manual, descripción, listado y detalles técnicos del programa Transición*. Universidad Autónoma Metropolitana, México D.F.
- Rojo-Domínguez A, Zubillaga-Luna RA y Hernández-Arana A. (1990) *Unfolding behavior of human α_1 -acid glycoprotein is compatible with a loosely folded region in its polypeptide chain*. Biochemistry 29: 8689-8695.
- Rose GD y Roy S. (1980) Proc. Natl. Acad. Sci. USA 77: 4643-4647.

- Rowe ES y Tanford C. (1973) *Equilibrium and kinetics of the denaturation of a homogeneous human immunoglobulin light chain*. *Biochemistry* 12: 4822-4827.
- Sawyer L, Papiz MZ, North ACT y Eliopoulos EE. (1985) *Structure and function of bovine β -lactoglobulin*. *Biochem. Soc. Trans.* 13: 265-266.
- Sawyer L. (1987) *One fold among many*. *Nature* 327:659.
- Scheer H y Kayser H. (1988) *Conformational studies of biliproteins from the insects *Pieris brassicae* and *Cerura vinula**. *Z. Naturforsch.* 43c: 84-90.
- Schiffer M y Edmundson AB. (1967) *Use of helical wheels to represent the structures of proteins and to identify segments with helical potential*. *Biophys. J.* 7: 121-135.
- Schiffer M, Girling RL, Ely KR y Edmundson AB. (1973) *Structure of a λ -type Bence-Jones protein at 3.5Å resolution*. *Biochemistry* 12(23): 4620-4631.
- Schmid FX. (1992) *The mechanism of protein folding*. *Curr. Op. in Struct. Biol.* 2: 21-25.
- Schmid K. (1953). *Preparation and properties of serum and plasma proteins. XXIX. Separation from human plasma of polysaccharides, peptides and proteins of low molecular weight. Crystallization of an acid glycoprotein*. *J. Amer. Chem. Soc.* 75: 60-78.
- Schmid K. (1972) *Recent advances in the study of glycoproteins*. *Chimia* 26(8): 405-414
- Schmid K. (1975) *α_1 -Acid glycoprotein*. en *The Plasma Proteins. Structure Function and Genetic Control. Vol 1* (Putnam F, ed.) 183-228. Springer-Verlag, Nueva York.
- Schmid K, Bencze WL, Nussbaumer T y Wehrmüller JO. (1959) *Studies on the structure of α_1 -acid glycoprotein*. *J. Mol. Biol.* 234: 529-533.
- Schmid K, Burlingame RW, Paulson JC y Hill RL. (1971) *The relationship between the carbohydrate units and the secondary structure of α_1 -acid glycoprotein*. *Fed. Proc.* 37(6): 1298.
- Schmid K, Emura J, Kaufmann H, Isemura S, Nimberg RB, Ishiguro M e Ikenaka T. (1971) *Homologies between α_1 -acid glycoprotein and the immunoglobulins and homologies within the amino acid sequence of α_1 -acid glycoprotein*. *Fed. Proc.* 31: 465.
- Schmid K, Chen KH, Occhino JC, Foster JA y Sperandio K. (1976) *Topography of human plasma α_1 -acid glycoprotein*. *Biochemistry* 15: 2245-2254.
- Schmid K, Kaufmann H, Isemura S, Bauer F, Emura J, Motoyama T, Ishiguro M y Nanno S. (1973) *Structure of α_1 -acid glycoprotein. The complete amino acid sequence, multiple amino acid substitutions, and homology with the immunoglobulins*. *Biochemistry* 12(14): 2711-2724.
- Schuber D, LaCorbiere M y Esch F. (1986) *A chick neural retina adhesion and survival molecule is a retinol-binding protein*. *J. Cell Biol.* 102: 2295-2301.
- Schwartz RM y Dayhoff MO. (1978) en *Atlas of Protein Sequence and Structure*, 5 suppl. *Nat. Biomed. Res. Found.*, Washington D.C. 3: 353-358.

- Shaw PH, Held WA y Hastie ND. (1983) *The gene family for mayor urinary proteins: Expression in several secretory tissues of the mouse.* Cell 32: 755-761.
- Silverton EW, Navia MA y Davies DR. (1977) *Three-dimensional structure of an intact human immunoglobulin.* Proc. Natl. Acad. Sci. USA 74(11): 5140-5144.
- Stouch TR y Jurs PC. (1986) J. Chem. Inform. Comput. Sci. 26: 4-12.
- Sundelin J, Laurent BC, Anundi H, Trágårdh L, Larhammar D, Björck L, Eriksson U, Åkerström B, Jones A, Newcomer M, Peterson PA y Rask L. (1985) *Amino acid sequence homologies between rabbit, rat and human serum retinol-binding proteins.* J. Biol. Chem. 260(10): 6472-6480.
- Svobodová X, Karpenko V y Kalous V. (1977) *On the alkaline denaturation of orosomucoid.* Collect. Czech. Chem. Commun. 42: 1742-1753.
- Takagi T, Takagi K y Kawai T. (1981) *Complete amino acid sequence of human α_1 -microglobulin.* Biochem. Biophys. Res. Commun. 98(4): 997-1001.
- Tanford C. (1969) *Protein denaturation.* Adv. Protein Chem. 23: 121-282.
- Tanford C. (1970) *Protein denaturation. Part C. Theoretical models for the mechanism of denaturation.* Adv. Protein Chem. 24: 1-95.
- Tirindelli R, Keen JN, Cavaggioni A, Eliopoulos EE y Findlay JBC. (1989) *Complete amino acid sequence of pyrazine-binding protein from cow nasal mucosa.* Eur. J. Biochem. 185: 569-572.
- Toh H, Hayashida H, Kikuno R, Yasunaga T y Miyata T. (1985) *Sequence similarity between EGF receptor and α_1 -acid glycoprotein.* Nature 314: 199.
- Treuheit MJ, Costello CE y Halsall HB. (1992) *Analysis of the five glycosilation sites of human α_1 -acid glycoprotein.* Biochem. J. 283: 105-112.
- Ulrich A, Coussens L, Hayflick JS, Dull TJ, Gray A, Tam AW, Lee J, Yarden Y, Libermann TA, Schlessinger J, Downward J, Mayes ELV, Whittle N, Waterfield MD y Seeburg PH. (1988) *Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells.* Nature 309: 418-425.
- Unterman RD, Lynch KR, Nakhasi HL, Dolan KP, Hamilton JW, Cohn DV y Feigelson P. (1981) *Cloning and sequence of several α_2 -globulin cDNAs.* Proc. Natl. Acad. Sci. USA 78(6): 3478-3482.
- Verlet L. (1967) *Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules.* Phys. Rev. 159: 98-103.
- Vetr H y Gebhard W. (1990) *Structure of the human α_1 -microglobulin-bikunin gene.* Biol. Chem. Hoppe-Seyler. 371: 1185-1196.
- Von Däniken E. (1990) *El día que llegaron los dioses.* Diana, México D.F.
- Waterman MS, Smith TE y Beyer WA. (1976) *Some biological sequence metrics.* Adv. Math. 20: 367-387

- Weidinger S, Müller T, Schwarzfischer F y Cleve H. (1987) *Three new orosomucoid (ORM) variants revealed by isoelectric focusing and print immunofixation.* Hum. Genet. 77: 286-288.
- Westphal U. (1971) *XIII. Interaction of progesterone and other steroids with α_1 -acid glycoprotein.* en Steroid-protein interactions. Monographs on endocrinology; v.4. Springer-Verlag, Berlin.
- Westphal U. (1985) *α_1 -acid glycoprotein (AAG, orosomucoid).* en Steroid-protein interactions II. Monographs on endocrinology; v.27. Springer-Verlag, Berlin.
- Weiner SJ, Kollmann PA, Nguyen DT y Case DA (1986) *Comput. Chem.* 7: 230.
- Wilson AC, Carlson SS y White TJ. (1977) *Biochemical evolution.* Ann. Rev. Biochem. 46: 573-639.
- Wong AKL y Hsia JC. (1983) *In vitro binding of propranolol and progesterone to native and desialylated human orosomucoid.* Can. J. Biochem. Cell. Biol. 61: 1114-1116.
- Yamagami K y Schmid K. (1967) *Conformational transitions associated with the release of the buried tryptophan residues of α_1 -acid glycoprotein.* J. Biol. Chem. 242(18): 4176-4181.
- Yamagami K, Labat J, Pandey RS y Schmid K. (1968) *The free and buried tyrosyl residues of α_1 -acid glycoprotein.* Biochemistry 7(8): 2873-2879.
- Zubillaga-Luna RA, Rodríguez-Muñoz I, Rojo-Domínguez A y Hernández-Arana A. (1991) *Análisis de residuos aromáticos superficiales en la glicoproteína α_1 de plasma humano.* Rev. Latinoamer. Quím. 22(1): 9-12.





Casa abierta al tiempo

UNIVERSIDAD AUTONOMA METROPOLITANA

AAP-CBI-24.92.

15 de Diciembre de 1992.

A QUIEN CORRESPONDA

Muy atentamente me permito certificar que el Dr. ANDRES HERNANDEZ ARANA, Profesor del Departamento de Química de la U.A.M.- Iztapalapa, participó como miembro del jurado que : 1) revisó la Tesis Doctoral, 2) opinó acerca de los artículos con arbitraje internacional producto de esa investigación y 3) entrevistó para la defensa técnica correspondiente a Arturo Rojo Domínguez quien presentó la Tesis Doctoral "Estudio conformacional y modelado de la glicoproteína ácida α 1 humana " la cual fue defendida ante ese Jurado y aprobada el 15 de diciembre de 1992.

Asimismo el Dr. Andrés Hernández fue el Secretario del Jurado durante el Examen de Grado.

ATENTAMENTE
"CASA ABIERTA AL TIEMPO"

DR. FERNANDO ROJAS GONZALEZ
COORDINADOR DEL DOCTORADO EN CIENCIAS
DIVISION DE CIENCIAS BASICAS E INGENIERIA
TEL: 724 26 72 FAX: 612 24 79

FRG'jv1



Casa abierta al tiempo

UNIVERSIDAD AUTONOMA METROPOLITANA

AAP-CBI-016/92.

A QUIEN CORRESPONDA

Muy atentamente me permito certificar que el **Dr. ANDRES HERNANDEZ ARANA** fue el Director de la Tesis Doctoral de **Arturo Rojo Domínguez: "Estudio conformacional y modelado molecular de la glicoproteína ácida α 1 humana"**, la cual fue defendida y aprobada en esta Universidad el 15 de diciembre de 1992.

Se extiende la presente para los fines que sean pertinentes en México, Distrito Federal a los quince días del mes de diciembre de mil novecientos noventa y dos..

A T E N T A M E N T E
"CASA ABIERTA AL TIEMPO"

DR. FERNANDO ROJAS GONZALEZ
COORD. DEL DOCTORADO EN CIENCIAS
C.B.I.

*Este trabajo fue desarrollado bajo la asesoría
del Dr. Andrés Hernández Arana en el Area de
Biofisicoquímica del Departamento de Química de
la Universidad Autónoma Metropolitana-
Iztapalapa.*

La materia, desde el instante de su creación, se ha transformado
hasta dar origen a sistemas vivientes.

Ha continuado su evolución generando pensamiento y
autoconciencia y ahora ese polvo de estrellas, materia
extraordinariamente organizada, empieza a mirar hacia
atrás; se asombra al descubrir algunos esbozos de la
gran complejidad y precisión de su propia
arquitectura, y se atreve a intentar describir,
clasificar y simular minúsculas porciones del vasto
conjunto de fenómenos que la estructuran ...
...y tal vez guían su destino.

Este año de 1992 ha estimulado el encuentro con nuestras raíces, al conmemorarse 500 años de la destrucción, con la cruz y la espada, de las culturas precolombinas. Abajo se muestra una porción del Códice Borbónico. Puede apreciarse una serie de 20 símbolos que representan los días en los que se basan los calendarios azteca y maya; o bien, dado el carácter plurisignificativo de estos símbolos, podrían interpretarse como los 20 aminoácidos principales que constituyen a las proteínas. Mas aún, esta figura podría describir a una célula completa, rodeada de su membrana celular y conteniendo un núcleo. En el interior de éste se encuentra una doble hélice que simboliza a los ácidos nucleicos; en el centro del núcleo parece representarse la separación de las cadenas que los constituyen y cómo los nucleótidos se alejan entre sí rompiendo tres puentes de hidrógeno que los mantenían unidos. En el interior de la membrana, cerca de las cuatro esquinas, los ribosomas simbolizados antropomórficamente forman un collar con los aminoácidos libres al ensartarlos con un bastón ...



En realidad, esta interpretación es una versión modificada de la que presenta von Däniken (1990) en uno de sus libros. Es interesante notar la importancia que empiezan a cobrar algunos de los descubrimientos de la Biología Molecular, la Biofísicoquímica y la Biotecnología fuera de los medios académicos. Esto habla de lo extraordinario de sus conclusiones y de lo prometedor de sus especulaciones. Cada vez, tenemos más herramientas para escudriñar con mayor detalle las características geométricas de las macromoléculas. Poder comprender el sutil balance que mantiene a la conformación de las biomoléculas en el estrecho intervalo que les concede su actividad biológica, y su regulación, es uno de los sueños (y de las metas) de muchos grupos de investigación. Puede anticiparse que en la próxima década se logrará la síntesis de las primeras enzimas artificiales, pues actualmente ya se están generando proteínas diseñadas *de novo*, aunque todavía sin actividad biológica. Nuestra generación es testigo del surgimiento de una nueva revolución tecnológica que, tal vez, sea de mayor envergadura que la producida en este siglo por la electrónica.