

T-P 21/10/98

T-10-135

UNIVERSIDAD DE GRANADA  
FACULTAD DE CIENCIAS

DEPARTAMENTO DE QUÍMICA FÍSICA

UNIVERSIDAD DE GRANADA
Facultad de Ciencias
Fecha ...22-4-98...
ENTRADA NUM. ...1609...

*Cálculos termodinámico-estructurales en proteínas:  
Predicción de afinidades de unión y diseño molecular  
de ligandos peptídicos*

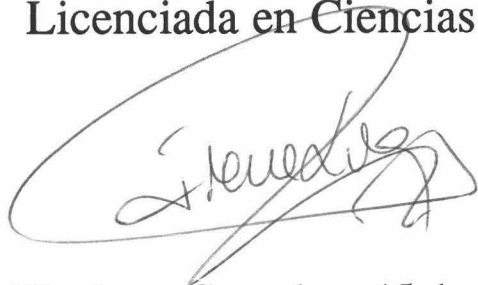
*Irene Luque Fernández*

Tesis Doctoral  
1998



Memoria presentada para aspirar al grado de  
Doctor en Ciencias Químicas

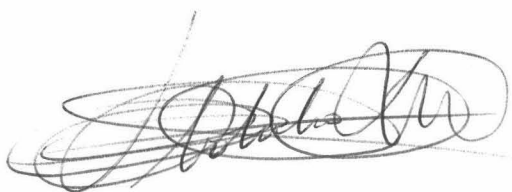
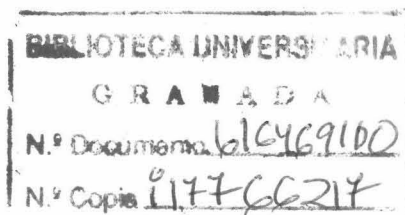
Firmado: Irene Luque Fernández  
Licenciada en Ciencias Químicas.



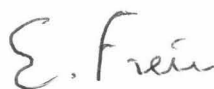
Visado en Granada, a 15 de Abril de 1998

DIRECTORES DE LA TESIS

Firmado:



Obdulio López Mayorga,  
Catedrático de Química Física  
de la Universidad de Granada



Ernesto Freire  
Profesor de la  
Universidad Johns Hopkins



## INDICE GENERAL

I.-INTRODUCCIÓN.....	1
II.- ANTECEDENTES BIBLIOGRÁFICOS	
II.A.-ESTABILIDAD DE HÉLICES ALFA. ESCALAS DE PROPENSIÓN HELICOIDAL	
<b>II.A.1.-Introducción.</b>	
II.A.1.1.-Importancia del estudio de los elementos de estructura secundaria: Hélices alfa.....	7
II.A.1.2.-Hélices alfa: modelo estructural.....	8
II.A.1.3.-Propensiones helicoidales.....	9
<b>II.A.2.-Modelos para la descripción termodinámico-estadística de la transición hélice-ovillo estadístico.</b>	
II.A.2.1.-Modelo general para la formación de una hélice.....	14
II.A.2.2.-Principales tratamientos mecanico-estadísticos.....	15
II.A.2.2.1.-Modelo de Zimm y Bragg.	
II.A.2.2.2.-Modelo de Lifson y Roig.	
II.A.2.2.3.-Comparación de ambas teorías.	
II.A.2.3.-Modificaciones de los modelos originales para la incorporación de distintas interacciones.....	18
<b>II.A.3.-Escala de propensión helicoidal.</b>	
II.A.3.1.-Correlaciones estadísticas: Frecuencias de aparición en hélices proteicas.....	19
II.A.3.1.1.-Análisis estadístico de Chou y Fasman.	
II.A.3.1.2.-Otros estudios estadísticos: <i>Consideración de las propensiones N-cap y C-cap</i>	
II.A.3.2.-Estudios con copolímeros sintéticos.....	21
II.A.3.3.-Estudios con péptidos monoméricos de bajo peso molecular.....	22
II.A.3.3.1.-Interés del estudio de pequeños péptidos monoméricos.....	22
II.A.3.3.2.-Necesidad de la aplicación de teorías mecanico-estadísticas al análisis de transiciones hélice-ovillo estadístico en pequeños péptidos.....	23
II.A.3.3.3.-Principales sistemas peptídicos utilizados para el estudio de propensiones helicoidales.....	24
II.A.3.4.-Estudios con proteínas y hélices enrolladas (coiled-coils).....	25
II.A.3.4.1.-Consideraciones generales.....	25
II.A.3.4.2.-Diferentes sistemas proteicos utilizados para el estudio de propensiones helicoidales.....	27
II.A.3.4.3.-Hélices enrolladas ("Coiled coils").....	28
II.A.3.5.-Estudios teóricos y semiempíricos.....	28

#### **II.A.4.-Principales conclusiones de los estudios experimentales.**

II.A.4.1.-Intervalo de propensiones helicoidales.....	29
II.A.4.2.-Comparación entre los resultados de análisis estadísticos y estudios experimentales.....	32
II.A.4.3.-Comparación de las propensiones helicoidales en péptidos y proteínas.....	34
II.A.4.4.-Escala de propensiones helicoidales consenso.....	35
II.A.4.5.- Interacciones adicionales de carácter estabilizador.....	36
II.A.4.5.1.- <i>Interacciones N-cap y C-cap</i> .....	36
II.A.4.5.2.- <i>Interacciones entre grupos cargados y el dipolo de la hélice</i> .....	39
II.A.4.5.3.- <i>Interacciones específicas entre cadenas laterales</i> .....	40
a.- <i>Motivos estructurales característicos de los extremos de la hélice</i> .....	40
b.- <i>Otras interacciones estabilizadoras entre cadenas laterales</i> .....	42

#### **II.B.-PROTEASAS ASPARTICAS.**

##### **II.B.1.- Enzimas proteolíticas.....45**

##### **II.B.2.-Proteasas Aspárticas.**

II.B.2.1.-Características Generales.....	45
II.B.2.2.-Estructura Tridimensional.....	47
II.B.2.3.-Estado de ionización del sitio activo.....	48
II.B.2.4.-Mecanismo catalítico.....	49
II.B.2.5.-Especificidad de sustratos.....	51
II.B.2.6.-Estudios de inhibición.....	52
II.B.2.6.1.- <i>Inhibidores que contienen Estatina</i> .....	53
II.B.2.6.2.- <i>Reducción del enlace peptídico hidrolizable</i> .....	54
II.B.2.6.3.- <i>Otros inhibidores</i> .....	55
II.B.2.6.4.- <i>Estructuras tridimensionales de los complejos</i> .....	55

##### **II.B.3.-Endothiapepsina.**

II.B.3.1.-Origen y secuencia.....	56
II.B.3.2.-Estructura tridimensional.....	57
II.B.3.3.-Estructura del centro activo.....	61
II.B.3.4.-Actividad hidrolítica.....	62
II.B.3.5.-Estudios de inhibición.....	63
II.B.3.5.1.- <i>Estructura Tridimensional</i> .....	63
II.B.3.5.2.- <i>Tipos de inhibidores</i> .....	65
a.- <i>Pepstatina A.</i>	
b.- <i>Otros inhibidores.</i>	

##### **II.B.4.-Proteasa del virus del S.I.D.A. tipo I.**

II.B.4.1.-Consideraciones generales.....	67
II.B.4.1.1.- <i>Proteasas retrovíricas</i> .....	67
II.B.4.1.2.- <i>Proteasa del virus del S.I.D.A. como objetivo terapéutico</i> .....	69

II.B.4.2.-Estructura tridimensional.....	70
II.B.4.3.-Actividad hidrolítica.....	75
II.B.4.3.1.-Especificidad de sustratos.....	75
a.-Sustratos oligopeptídicos.....	76
b.-Sustratos proteicos.....	77
II.B.4.3.2.-Influencia del pH y fuerza iónica.....	78
a.-Parámetros cinéticos del proceso de hidrólisis.....	78
b.-pKa de los aspárticos catalíticos en los complejos con distintos inhibidores.....	79
II.B.4.3.3.-Inhibición por los productos de la hidrólisis.....	80
II.B.4.4.-Estudios de inhibición.....	81
II.B.4.4.1.-Inhibidores peptídicos.....	81
a.-Inhibidores basados en sustratos peptídicos	
b.-Inhibidores de otras proteasas aspárticas.	
II.B.4.4.2.-Inhibidores no peptídicos.....	83
a.-Inhibidores simétricos.....	83
b.-Diseño de novo de inhibidores no peptídicos.....	85
II.B.4.5.-Desarrollo de resistencia a fármacos.....	86
II.B.4.5.1.-Localización y naturaleza de las mutaciones que confieren resistencia.....	85
II.B.4.5.2.-Estudios cinéticos y estructurales.....	87
II.B.4.5.3.-Resistencia múltiple.....	88

### III.-METODOLOGÍAS DE CÁLCULO DE ENERGÍA LIBRE.

#### III.1.-CONSIDERACIONES GENERALES.

III.1.1.- <u>Interés del cálculo de Energías libres</u> .....	91
III.1.2.- <u>Cálculo mecánico-estadístico de magnitudes termodinámicas</u> .....	92
III.1.2.1.- Muestreo configuracional.....	93
III.1.2.2.-Funciones de energía potencial.....	94
III.1.3.- <u>Aproximaciones semiempíricas al cálculo de energías libres</u> .....	95
III.1.3.1.-Ciclos termodinámicos clásicos.....	95
III.1.3.1.1.-Hidratación: Hipótesis de aditividad de grupos.....	96
III.1.3.1.2.-Estimación de las contribuciones a la hidratación a través de accesibilidad al disolvente.....	97
III.1.3.1.3.-Problemas de estos enfoques.....	98
a.-Fenómenos compensatorios entre magnitudes termodinámicas.	
b.-Comportamiento no aditivo de las contribuciones individuales.	

## III.2.-PARAMETRIZACIÓN ESTRUCTURAL DE LA ENERGÉTICA DE PLEGAMIENTO DE PROTEÍNAS Y UNIÓN DE LIGANDOS.

### III.2.1.-Características generales.

III.2.1.1.-Parametrización independiente de las distintas magnitudes termodinámicas.....	100
III.2.1.2.-Parametrización basada en datos termodinámicos de proteínas.....	100
III.2.1.3.-La parametrización utilizada representa un enfoque fenomenológico.....	101

### III.2.2.- Energía libre de Gibbs.

III.2.2.1.-Partición general de la energía libre de Gibbs.....	101
III.2.2.2.-Energía de Gibbs genérica.....	102
III.2.2.3.-Cálculo de las áreas accesibles al disolvente.....	103
III.2.2.3.1.-Algoritmo de Lee y Richards.....	104
III.2.2.4.2.-Áreas accesibles de conformaciones no estructuradas.....	105
III.2.2.4.-Base de datos empleada para la parametrización.....	106

### III.2.3.-Contribuciones genéricas a la energía de Gibbs.

#### III.2.3.1.-Entalpía.

III.2.3.1.1.-Consideraciones generales.....	107
III.2.3.1.2.-Parametrización inicial.....	108
III.2.3.1.3.-Temperaturas de referencia.....	108
III.2.3.1.4.-Parametrización actual: consideración del grado de empaquetamiento.....	109
a.-Distribución de distancias interatómicas en proteínas.....	110
b.-Introducción del grado de empaquetamiento.....	111
b.1.-Interacciones de van der Waals.....	111
b.2.-Parametrización conjunta.....	112

#### III.2.3.2.-Entropía.

III.2.3.2.1.-Consideraciones generales.....	113
III.2.3.2.2.-Entropía de hidratación.....	114
III.2.3.2.3.-Entropía conformacional.....	115
a.-Parametrización inicial.....	115
b.-Parametrización actual: evaluación independiente de los distintos términos entrópicos.....	116
b.1.-Entropía conformacional de las cadenas laterales.....	117
b.1.1.- $\Delta S_{bu \rightarrow ex}$ .....	117
b.1.2.- $\Delta S_{ex \rightarrow u}$ .....	118
b.2.-Entropía conformacional del esqueleto peptídico.....	119
b.2.1.-Efecto de la longitud de la cadena.....	120

#### III.2.3.3.- Capacidad calorífica.

III.2.3.3.1.-Consideraciones generales.....	121
---	-----

III.2.3.3.2.-Parametrización inicial.....	123
III.2.3.3.3.-Parametrización actual:	
dependencia con la temperatura.....	124
a.-Contribuciones a la capacidad calorífica de proteínas.....	124
b.-Parametrización global de la capacidad calorífica.....	125
b.1.-Contribuciones covalentes.....	126
b.2.-Contribución de las interacciones internas.....	127
b.3.-Hidratación.....	127
b.4.-Efectos de protonación.....	128
b.5.-Ajuste global de la capacidad calorífica.....	129
<b>III.2.4.-Contribuciones electrostáticas a la energía de Gibbs.</b>	
III.2.4.1.-Contribuciones electrostáticas.....	131
III.2.4.2.-Efectos de protonación.....	133
<b>III.2.5.- Contribuciones traslacionales a la energía de Gibbs.</b>	134
<b>III.2.6.-Diversas aplicaciones.</b>	135

## IV.-RESULTADOS Y DISCUSIÓN.

### IV.A.-OBTENCIÓN DE UNA ESCALA TERMODINÁMICO-ESTRUCTURAL DE PROPENSIÓN HELICOIDAL. OPTIMIZACIÓN DE LA PARAMETRIZACIÓN ESTRUCTURAL DE LA ENERGÉTICA DE PLEGAMIENTO DE PROTEÍNAS Y UNIÓN DE LIGANDOS..

<b>IV.A.1.-Consideraciones generales.</b>	137
<b>IV.A.2.-Métodos</b>	
IV.A.2.1.-Cálculo de áreas accesibles promedio en conformación helicoidal.	
IV.A.2.1.1.-Construcción de las estructuras helicoidales.....	139
IV.A.2.1.2.-Obtención de los perfiles de accesibilidad.....	140
IV.A.2.1.3.-Obtención de los perfiles de probabilidad.....	140
IV.A.2.1.4.-Cálculo de los promedios de Boltzmann para las áreas accesibles en conformación helicoidal.....	141
IV.A.2.1.5.-Influencia de cadenas laterales vecinas en la accesibilidad de un aminoácido determinado.....	142
IV.A.2.2.-Áreas accesibles al disolvente en conformaciones desplegadas.....	142
IV.A.2.2.1.-Construcción de los tripéptidos.....	143
IV.A.2.2.2.-Cálculo de perfiles de accesibilidad en el estado desplegado.....	143
<b>IV.A.3.-Resultados y discusión.</b>	
IV.A.3.1.-Cálculo de áreas accesibles promedio en conformación helicoidal.....	144
IV.A.3.1.1.-Perfiles de accesibilidad y probabilidad..	

IV.A.3.1.2.-Áreas accesibles promedio..	
IV.A.3.2.-Cálculo de áreas accesibles promedio en conformaciones desplegadas.....	149
IV.A.3.3.-Escala de propensión helicoidal.....	155
IV.A.3.3.1.-Cálculo de propensiones helicoidales.	
Refinado de los valores de accesibilidad en el estado desplegado.....	157
IV.A.3.3.2.-Escala de propensión helicoidal.	
Comparación con otros trabajos.....	159
IV.A.3.4.-Análisis termodinámico-estructural de la formación de una hélice alfa.....	163
IV.A.3.4.1.-Energía libre de Gibbs.....	163
IV.A.3.4.2.-Entalpía.....	164
IV.A.3.4.3.-Entropía.....	165

**IV.B.-APLICACIÓN DE LA PARAMETRIZACIÓN ESTRUCTURAL DE LA ENERGÉTICA DE PLEGAMIENTO DE PROTEÍNAS Y UNIÓN DE LIGANDOS AL DISEÑO MOLECULAR. DISEÑO DE LIGANDOS PEPTÍDICOS DE PROTEASAS ASPÁRTICAS.**

<b>IV.B.1.-Consideraciones generales.....</b>	<b>167</b>
---	------------

**IV.B.2.-Métodos.**

IV.B.2.1.-Descripción general del enfoque termodinámico-estructural del diseño molecular	
IV.B.2.1.1.-Consideraciones generales.....	169
IV.B.2.1.2.-Estrategia general para el diseño termodinámico-estructural de péptidos.....	170
a.-Análisis termodinámico-estructural del complejo original.....	171
b.-Modificación del péptido original.....	171
b.1.-Modificación de la secuencia: Análisis mutacional.	
b.2.-Modificación de la longitud.	
IV.B.2.1.3.-Minimización de conformaciones.....	173
a.-Descripción general.....	173
b.-Selección de conformaciones permitidas.....	174
c.-Función potencial.....	176
IV.B.2.2.-Minimización de conformaciones helicoidales.	
IV.B.2.2.1.-Construcción de las hélices.....	177
IV.B.2.2.2.-Generación de conformaciones.....	177
IV.B.2.3.-Estudio termodinámico-estructural de la energética de unión de inhibidores a la proteasa del virus V.I.H.-1.	
IV.B.2.3.1.-Cálculo de la Energía libre de unión.....	177
a.-Situaciones especiales que deben ser consideradas.....	181

a.1.-Presencia de aminoácidos y grupos poco comunes..	
a.2.-Estructuras incompletas o con conformaciones múltiples.	
b.-Cálculo de accesibilidades.....	182
c.-Cálculo de entropía conformacional para los inhibidores no peptídicos.....	183
d.-Contribuciones de los efectos de ionización.....	185
e.-Contribución de la entropía translacional.....	186
IV.B.2.3.2.-Localización estructural de las contribuciones a la energética de unión.....	186
IV.B.2.3.3.-Cálculo de las constantes de estabilidad por resto.....	187
a.-Constantes de estabilidad de restos individuales.....	187
b.-Aplicación al estudio de la proteasa del virus V.I.H.-1.....	188
IV.B.2.4.-Modelado de ligandos peptídicos para proteasas aspárticas.	
IV.B.2.4.1.-Diseño de inhibidores peptídicos para Endothiapepsina relacionados con Pepstatina A.....	188
a.-Identificación de posiciones de mutación.....	188
b.-Generación de los inhibidores mutantes.....	189
c.-Optimización energética de la conformación de los inhibidores.....	190
d.-Cálculo de las afinidades de unión.....	190
IV.B.2.4.2.-Modelado de sustratos de la proteasa el virus V.I.H.-1.....	191
a.-Sustrato cromogénico.....	191
b.-Sustrato natural.....	193
c.-Optimización de la conformación de los restos mutados.....	193
d.-Cálculo de las afinidades de unión.....	195
<b>IV.B.3.-Resultados y discusión.</b>	
IV.B.3.1.-Minimización de conformaciones helicoidales.....	195
IV.B.3.2.-Estudio termodinámico-estructural de la unión de inhibidores a la proteasa del virus V.I.H.-1.	
IV.B.3.2.1.-Consideraciones generales.....	209
IV.B.3.2.2.-Inhibidores no peptídicos.....	211
a.-Predicción de afinidades de unión.....	211
b.-Diseción de la contribución de las distintas magnitudes termodinámicas a la afinidad de unión.....	212
c.-Localización estructural de los restos de la proteasa que contribuyen más notablemente a la afinidad de unión.....	214
d.-Análisis de la estabilidad estructural de los restos de la proteasa que contribuyen más notablemente a la afinidad de unión.....	215

e.- <i>Bases moleculares de la resistencia de la proteasa     frente a distintos fármacos</i> .....	221
IV.B.3.2.3.- <i>Inhibidores peptídicos: Acetil-pepstatina</i> .....	222
a.- <i>Análisis de la energética del proceso de unión</i> .....	222
b.- <i>Identificación de los restos que aportan las principales     contribuciones a la energética de unión</i> .....	225
 IV.B.3.3.- <i>Modelado de ligandos peptídicos para proteasas aspárticas</i>	
IV.B.3.3.1.- <i>Diseño de inhibidores peptídicos para Endothiapepsina</i> .....	226
a.- <i>Análisis mutacional en la posición 5 de Pepstatina A</i> .....	226
b.- <i>Adición de un resto Glutámico en el extremo C-terminal</i> .....	231
c.- <i>Localización estructural de las contribuciones</i>	
a <i>las afinidades de unión</i> .....	231
c.1.- <i>Pepstatina A</i> .....	231
c.2.- <i>Mutante A5F</i> .....	235
c.3.- <i>Mutante E7</i> .....	237
IV.B.3.3.2.- <i>Unión de sustratos a la proteasa del virus V.I.H.-1</i> .....	239
a.- <i>Consideraciones generales</i> .....	239
b.- <i>Descripción de los modelos estructurales</i> .....	240
c.- <i>Análisis termodinámico-estructural del proceso de unión</i> .....	240
c.1.- <i>Localización estructural de las principales contribuciones             la energética de unión</i> .....	241
c.2.- <i>Especificidad de los sustratos</i> .....	247
IV.B.3.3.3.- <i>Bases moleculares de la resistencia a fármacos     en la proteasa del virus H.I.V.-1</i> .....	248
a.- <i>Análisis comparativo de la distribución estructural de las         contribuciones a la energética de unión de sustratos peptídicos         e inhibidores orgánicos</i> .....	248
b.- <i>Diferencias fundamentales en la energética de unión</i> .....	249
c.- <i>Hipótesis sobre el origen molecular de la resistencia</i> .....	250
 <b>V.-APÉNDICES</b> .....	253
<b>VI.-CONCLUSIONES</b> .....	257
<b>VII.-BIBLIOGRAFÍA</b> .....	263



## INDICE DE FIGURAS

FIGURA	PÁGINA
II.A.1.....	10
II.A.2.....	17
II.B.1.....	48
II.B.2.....	50
II.B.3.....	53
II.B.4.....	58
II.B.5.....	64
II.B.6.....	68
II.B.7.....	72
II.B.8.....	84
III.1.....	96
IV.A.1.....	146
IV.A.2.....	149
IV.A.3.....	151
IV.A.4.....	153
IV.A.5.....	158
IV.A.6.....	160
IV.A.7.....	162
IV.A.8.....	164
IV.B.1.....	172
IV.B.2.....	175
IV.B.3.....	178
IV.B.4.....	184
IV.B.5.....	189
IV.B.6.....	192
IV.B.7.....	194
IV.B.8.....	196
IV.B.9.....	206
IV.B.10.....	213
IV.B.11.....	216
IV.B.12.....	219
IV.B.13.....	222
IV.B.14.....	224
IV.B.15.....	228
IV.B.16.....	227
IV.B.17.....	232
IV.B.18.....	236
IV.B.19.....	238
IV.B.20.....	242
IV.B.21.....	245
IV.B.22.....	246
IV.B.23.....	249

## INDICE DE TABLAS

TABLA	PAGINA
II.A.1.....	30
II.A.2.....	33
II.A.3.....	35
II.A.4.....	37
II.B.1.....	76
III.1.....	105
III.2.....	119
III.3.....	130
IV.A.1.....	140
IV.A.2.....	143
IV.A.3.....	145
IV.A.4.....	156
IV.B.1.....	176
IV.B.2.....	213
IV.B.3.....	215
IV.B.4.....	225
IV.B.5.....	227
IV.B.6.....	241



## **I.-INTRODUCCIÓN**

## I.-INTRODUCCIÓN.

El reconocimiento molecular constituye una de las etapas esenciales en numerosos procesos de interés biológico, de modo que el estudio de estos fenómenos, que pueden tener lugar tanto intermolecularmente (como es el caso de la unión de cofactores, sustratos e inhibidores a enzimas, de la identificación de antígenos por sus correspondientes anticuerpos, de la interacción entre hormonas y receptores, o de las asociaciones entre proteínas y ácidos nucleicos) como de forma intramolecular, de modo que el proceso de plegamiento de una proteína constituye un ejemplo notable por su complejidad, se ha convertido en una cuestión de extraordinaria relevancia.

La elucidación de los factores que determinan estos procesos de reconocimiento posee un indudable interés académico y enormes repercusiones biotecnológicas y terapéuticas, ya que la comprensión de estos fenómenos facilitaría el desarrollo de estrategias racionales que harían posible el diseño de nuevas proteínas con las funciones y estabilidades deseadas, o fármacos específicos que posean una afinidad predeterminada por sus receptores.

Además, el análisis de las relaciones existentes entre la estructura, función y estabilidad de las macromoléculas biológicas, que cobran especial relevancia en este contexto, se ha visto facilitado por el notable incremento de la cantidad y calidad de la información estructural disponible para los distintos sistemas biológicos como consecuencia del desarrollo experimentado en los últimos años por las distintas técnicas de análisis estructural. Tanto la estabilidad estructural de un sistema macromolecular, como las afinidades de unión de sus ligandos, están determinadas por la energía libre de Gibbs, por lo que el principal objetivo de cualquier metodología de diseño molecular basada en consideraciones estructurales debe ser la predicción de su magnitud a partir de las coordenadas atómicas y otros parámetros estructurales de los sistemas bajo estudio.

Sin embargo, hasta el momento, la termodinámica tradicional no ha desempeñado un papel relevante en el desarrollo de las técnicas de diseño molecular debido a la inexistencia de una base teórica que relacione de forma precisa las características estructurales del sistema con su energía libre de Gibbs. Aunque la termodinámica estadística proporciona la conexión entre la descripción microscópica del sistema y las magnitudes termodinámicas de naturaleza macroscópica, la complejidad de las interacciones, fundamentalmente de carácter no covalente, presentes en estas macromoléculas biológicas y la gran influencia que sus interacciones con el disolvente tienen en la determinación de sus magnitudes termodinámicas dificultan sobremanera el proceso de cálculo.

En este contexto, la utilización de funciones potenciales aproximadas de carácter semiempírico que, aún a pesar de no aportar información acerca de la fuerzas particulares que determinan la energética de los procesos bajo estudio, permitan predecir de forma cuantitativa la energía libre de Gibbs a partir de datos estructurales, aparece como una alternativa de gran atractivo. Algunos grupos de investigación han desarrollado formalismos semiempíricos que

permiten el cálculo de las distintas magnitudes termodinámicas correspondientes a procesos que implican proteínas a partir de los cambios en sus superficies accesibles al disolvente como consecuencia del proceso bajo estudio (Makhatadze & Privalov, 1995; Oobatake & Ooi, 1992), utilizando los parámetros de proporcionalidad obtenidos a partir del estudio experimental de pequeños compuestos modelo (Fauchère & Pliska, 1983; Kauzmann, 1959; Nozaki & Tanford, 1971; Tanford, 1962; Wolfenden et al., 1981). La utilización de estas metodologías para el estudio de macromoléculas presenta serios problemas, ya que la extrapolación del comportamiento de los compuestos modelo, generalmente moléculas de pequeño tamaño y de gran simplicidad estructural, a la situación encontrada en una proteína requiere la suposición de un comportamiento aditivo para las contribuciones de los grupos individuales. Aunque la hipótesis de aditividad de grupos parece ser adecuada para la descripción de las contribuciones de los grupos apolares, que se encuentran implicados fundamentalmente en interacciones de corto alcance, presenta grandes limitaciones a la hora de estimar las contribuciones polares a las magnitudes termodinámicas, para las que se han descrito serias desviaciones del comportamiento aditivo (Lazaridis et al., 1995; Cabani et al., 1981).

En este contexto, la parametrización estructural de la energética de plegamiento de proteínas y unión de ligandos, desarrollada durante los últimos años en el laboratorio del Dr. Freire, que considera la utilización de datos experimentales procedentes del estudio de proteínas para el establecimiento de las correlaciones energía/estructura como una aproximación más adecuada al problema de la predicción cuantitativa de la energética de estos sistemas, constituye una interesante alternativa. Así, esta metodología está basada en un juego de ecuaciones empíricas que permiten relacionar las distintas magnitudes termodinámicas características de los procesos de plegamiento de proteínas y unión de sus ligandos con los correspondientes cambios en las áreas polares y apolares accesibles al disolvente y otras características estructurales como la distribución de distancias interatómicas. Este formalismo se ha aplicado con éxito a la estimación de los parámetros termodinámicos de procesos de plegamiento/desplegamiento de proteínas (Freire et al., 1993; Freire & Murphy, 1991; Murphy et al., 1992; Murphy & Freire, 1992; Xie et al., 1993) y ha hecho posible llevar a cabo el análisis de las bases moleculares de la cooperatividad en estos sistemas, permitiendo identificar las distintas propiedades, tanto termodinámicas como estructurales, que caracterizan los posibles intermedios termodinámicos en el proceso de plegamiento (Freire, 1995; Freire & Xie, 1994; Xie et al., 1994; Xie & Freire, 1994b). Asimismo, esta metodología se ha aplicado al estudio de procesos unión, como es el caso de la asociación Angiotensina con anticuerpos (Murphy et al., 1993) o la unión del inhibidor Pepstatina A (Gómez & Freire, 1995) y otros inhibidores de renina a la proteasa aspártica Endothiapepsina (Gómez et al., 1997).

En esta memoria se ha abordado la aplicación de esta metodología a procesos de diseño molecular, para lo cual se requiere necesariamente la capacidad de evaluar los efectos de pequeños cambios estructurales en la molécula, y que frecuentemente se encuentran asociados a la sustitución de aminoácidos individuales. Concretamente, se ha abordado el estudio de las pansiones helicoidales de los distintos aminoácidos, para lo cual es necesario evaluar con la mayor precisión posible los pequeños cambios en accesibilidades que se producen como

consecuencia de la sustitución de un aminoácido por otro. Como se describe en la sección IV.A. de esta memoria, se han obtenidos los promedios de Boltzmann para las áreas accesibles al disolvente de los distintos aminoácidos en posiciones centrales de una hélice alfa expuesta al disolvente. La evaluación de las áreas accesibles al disolvente en conformaciones desestructuradas es una cuestión de más difícil solución, en este trabajo se ha obtenido un juego de valores optimizados para estas áreas, que supuestamente reflejan con mayor fidelidad la situación existente en cadenas polipeptídicas desplegadas. La utilización de estos nuevos valores para las áreas accesibles al disolvente permite calcular las propensiones helicoidales de los distintos aminoácidos con un error de  $0.2 \text{ kcal}\cdot\text{mol}^{-1}$  y desarrollar una escala de propensión helicoidal que se encuentran en excelente correspondencia con las distintas escala existentes en la literatura (Luque et al., 1996).

Una vez comprobado que la metodología de cálculo es suficientemente sensible como para cuantificar los efectos de simples mutaciones, es necesario, dado que cualquier estrategia de diseño molecular implica una etapa de minimización estructural, comprobar asimismo que la parametrización estructural de la energética utilizada en este trabajo posee la resolución necesaria para identificar los mínimos energéticos en el espacio conformacional. Para ello se ha procedido a la optimización energética de las conformaciones adoptadas por las cadenas laterales de los distintos aminoácidos en las posiciones centrales de una hélice alfa, obteniéndose perfiles energéticos que concuerdan con los existentes en la literatura y que han sido calculados a partir de funciones potenciales estándar. En este punto, queda asegurado que la parametrización estructural de la energética reúne los requisitos necesarios para su aplicación con éxito al desarrollo de procesos de diseño molecular, que será tratado en la segunda parte de esta memoria (IV.B).

Los sistemas escogidos para la aplicación de esta metodología pertenecen al grupo de las proteasas aspárticas, que constituyen una gran familia de enzimas proteolíticas, algunos de cuyos miembros han adquirido recientemente gran relevancia médica. En primer lugar nuestra atención se ha centrado en el estudio de la interacción de inhibidores y sustratos de diversa naturaleza con Endothiapepsina, que es una de las proteasas aspárticas más profundamente estudiadas, por lo que constituye un buen modelo para iniciar el proceso de diseño de inhibidores.

Una vez comprobada la capacidad de la parametrización estructural de la energética para reproducir con precisión los resultados experimentales obtenidos con estos sistemas (con lo que queda asegurado que la metodología empleada posee la sensibilidad necesaria para captar el delicado balance de fuerzas que determinan las afinidades de unión de ligandos a esta enzima) se procedió a la aplicación de las estrategias de diseño molecular, que incorporan esta parametrización estructural como función potencial, al diseño de ligandos peptídicos (Luque & Freire, 1998).

En concreto, se ha llevado a cabo el análisis termodinámico estructural de la unión de diversas variantes del inhibidor universal de las proteasas aspárticas Pepstatina A a Endothiapepsina. Los nuevos ligandos se generaron, bien mediante la sustitución de las cadenas laterales en distintas posiciones del pentapéptido, o bien modificando su longitud. De este modo, ha sido posible establecer una escala de afinidades de unión para los posibles mutantes en una



posición determinada en la secuencia e identificar los restos de la enzima y del inhibidor que contribuyen más notablemente a la energética de unión (Luque et al., 1997).

La segunda enzima considerada en este trabajo, también perteneciente al grupo de las proteasas aspárticas, es la proteasa del virus V.I.H.-1. Esta enzima retrovírica, que recientemente ha adquirido una gran relevancia médica debido a su papel central en el proceso de maduración del virus del S.I.D.A., constituye uno de los principales objetivos en el desarrollo de terapias efectivas contra la enfermedad. Aunque los tratamientos clínicos que incorporan algunos de los inhibidores de esta proteína recientemente desarrollados han producido resultados muy prometedores, la aparición de cepas víricas que muestran resistencia a muchos de estos fármacos ha constituido un revés importante en el tratamiento de la infección (Condra et al., 1995; Ho et al., 1994; Kaplan et al., 1994; Roberts, 1995; Tisdale, 1996). La pérdida de sensibilidad hacia los inhibidores tiene su origen en el desarrollo de virus mutantes que codifican moléculas de proteasa con mutaciones específicas que tienen como resultado la disminución drástica de la afinidad de la enzima por los inhibidores pero que, sin embargo, mantienen tanto su afinidad por los sustratos, como su actividad catalítica a niveles suficientes para la supervivencia del virus.

Los orígenes de la resistencia aún permanecen oscuros, y se han observado, tanto mutaciones localizadas en las cercanías del sitio de unión, como otras situadas en regiones muy alejadas del centro catalítico y de efectos desconocidos. Además, los distintos inhibidores producen patrones mutacionales diferentes de modo que, a pesar de que todos ellos han sido diseñados con el mismo centro activo como objetivo, se han observado patrones de resistencia múltiple bastante variados. El desarrollo de una nueva generación de inhibidores de la proteasa que resuelva de modo efectivo el problema de la resistencia requiere necesariamente la comprensión de las interacciones implicadas en la determinación de la afinidad de la enzima, tanto por los inhibidores, como por los sustratos.

En este contexto, se ha llevado a cabo el análisis termodinámico-estructural de la unión de trece inhibidores de naturaleza no peptídica, algunos de los cuales se encuentran actualmente en pruebas clínicas, y para cuyos complejos con la enzima se dispone de estructuras tridimensionales de alta resolución (Abdel-Meguid et al., 1994; Erickson et al., 1990; Fassler et al., 1993; Hoog et al., 1995; Kim et al., 1995; Lin et al., 1995; Madhusoodan et al., 1994; Thaisrivongs et al., 1995; Thompson et al., 1994). Este análisis proporciona una descripción detallada de la energética de unión de ambos tipos de inhibidores así como la localización de los restos de la enzima y del ligando que contribuyen especialmente a la determinación de la afinidad de unión. Con objeto de profundizar en el estudio de los determinantes del fenómeno de la resistencia se ha abordado el estudio de los efectos que una de las mutaciones más frecuentemente observadas, V82A tiene sobre las interacciones establecidas con el inhibidor no peptídico A-77003.

De forma complementaria, el estudio de las características termodinámicas del proceso de unión de los sustratos de estas enzimas permitiría la identificación de sus rasgos diferenciales con respecto a los inhibidores, que posibilitan la aparición del fenómeno de la resistencia. Como un primer paso, el análisis termodinámico estructural de inhibidores peptídicos, diseñados como análogos de los sustratos aunque con el enlace peptídico hidrolizable sustituido, puede proporcionar información interesante. En este sentido, se ha realizado el estudio del proceso de

unión a la proteasa del virus V.I.H.-1 del inhibidor peptídico Acetil-pepstatina. Asimismo, se ha abordado el proceso de modelado de los complejos de la proteasa del virus con diversos sustratos peptídicos con objeto de elucidar las bases moleculares de la resistencia. Este análisis ha permitido identificar las diferencias fundamentales en la energética de unión entre inhibidores no peptídicos y los sustratos, por lo que ha sido posible sugerir una posible explicación del origen molecular de la resistencia a fármacos en este sistema (Luque et al., 1998; Luque & Freire, 1998).





## **II.-ANTECEDENTES BIBLIOGRÁFICOS**

## II.A.-ESTABILIDAD DE HÉLICES ALFA. ESCALAS DE PROPENSIÓN HELICOIDAL.

### II.A.1.-Introducción.

#### II.A.1.1.-Interés del estudio de los elementos de estructura secundaria: Hélices alfa.

El desarrollo de estrategias de diseño racional e ingeniería de proteínas que posibiliten, tanto la modificación de las características (en especial estabilidad y función) de proteínas conocidas, como la creación de enzimas artificiales con propiedades específicas, es uno de los problemas biológicos de mayor relevancia en la actualidad. En este contexto es especialmente interesante profundizar en el estudio de los factores determinantes del proceso de plegamiento, tanto desde un punto de vista cinético como termodinámico.

A pesar de haberse demostrado que en numerosas ocasiones las proteínas se repliegan espontáneamente *in vitro*, originando su conformación nativa, lo que implica que toda la información necesaria acerca de su estructura tridimensional biológicamente activa se encuentra contenida en su secuencia aminoacídica, aún no se ha alcanzado una solución satisfactoria al problema del plegamiento. Desde un punto de vista termodinámico, la predicción de la estructura adoptada por una cadena polipeptídica bajo unas condiciones determinadas requeriría el conocimiento cuantitativo de la contribución energética de cada resto a su estabilidad, así como de las preferencias conformacionales de los distintos aminoácidos. En este sentido, el estudio de los factores que estabilizan los distintos elementos de estructura secundaria adquiere especial relevancia.

Desde una perspectiva dinámica, el corto intervalo de tiempo necesario para completar el proceso de plegamiento, en ocasiones del orden de 100  $\mu$ s, anula la posibilidad de que el plegamiento transcurra mediante el muestreo aleatorio de todas las posibilidades configuracionales. Deben existir, por tanto, uno o más "camino" o mecanismos de plegamiento que limiten de forma significativa la búsqueda conformacional. En este sentido, aunque todavía es motivo de controversia, los principales modelos actuales para el mecanismo de plegamiento de proteínas (Dill et al., 1995; Karplus & Weaver, 1994) atribuyen un papel crítico a las interacciones de carácter local en la determinación y definición de las etapas más tempranas del proceso de plegamiento. Por tanto, la identificación de motivos estructurales de carácter local, con estabilidad suficiente para reducir la flexibilidad conformacional en ciertos puntos de la cadena polipeptídica, presenta un interés especial para la simplificación y comprensión del modo en que se pliega una proteína.

Existen evidencias experimentales que otorgan una especial importancia a la formación de ciertos elementos de estructura secundaria en las etapas más tempranas del proceso de plegamiento, con anterioridad al establecimiento de la estructura terciaria, que aparecería como consecuencia de las interacciones establecidas entre las distintas unidades independientes de plegamiento (Mathews, 1993). Asimismo, la tendencia mostrada por pequeños péptidos aislados a

poblar significativamente la misma conformación que presentan en el estado nativo de la proteína de la cual proceden (Dyson & Wright, 1993) ha llevado a postular un papel de gran importancia para los elementos de estructura secundaria en la conducción de la búsqueda conformacional (Dill et al., 1995; Karplus & Weaver, 1994; Kim & Baldwin, 1990; Wright et al., 1988).

Por tanto, el estudio de los factores que afectan a la formación, localización y estabilización de dichos elementos estructurales presenta un extraordinario interés para la comprensión del mecanismo de plegamiento de una proteína. En este contexto, el estudio de las hélices alfa presenta un interés excepcional, dado que no sólo constituyen uno de los elementos secundarios más abundantes en proteínas globulares, una cuarta parte de cuyos aminoácidos se encuentran en conformación helicoidal, sino que se encuentran estabilizadas fundamentalmente por interacciones de tipo local, lo que incrementa su potencial como elementos iniciadores del proceso de plegamiento.

#### **II.A.1.2.-Hélices alfa: modelo estructural.**

La característica fundamental del modelo estructural para la hélice alfa, propuesto inicialmente por Pauling, Corey y Branson (Pauling et al., 1951), y corroborado por numerosos resultados experimentales posteriores, es su carácter repetitivo, de modo que los ángulos diedros del esqueleto polipeptídico ( $\phi$  y  $\psi$ ) adoptan valores constantes, cercanos a los valores canónicos de  $-60^\circ$  y  $-40^\circ$ , respectivamente (Richardson, 1981). Estos valores, corresponden a un mínimo energético relativamente marcado en el espacio conformacional, aún ignorando la contribución estabilizadora de los enlaces de hidrógeno.

Una vez que los ángulos diedros de un segmento de la cadena adoptan los valores helicoidales, los grupos polares del esqueleto se encuentran automáticamente situados con la orientación adecuada, paralela al eje de la hélice, para la formación de enlaces de hidrógeno con otros grupos complementarios. Como consecuencia, cada grupo donador [-N-H] de la hélice se encuentra implicado en un puente de hidrógeno con el grupo aceptor [-C-O] perteneciente al resto situado cuatro posiciones más adelante. La estructura resultante satisface por completo el potencial formador de enlaces de hidrógeno de los grupos polares consecutivos del esqueleto peptídico, con la excepción de los primeros cuatro grupos de los extremos inicial y final, que se encuentran descompensados. De acuerdo con la nomenclatura clásica para la descripción de hélices polipeptídicas, la hélice alfa se conoce como una hélice  $3.6_{13}$ , siendo 3.6 el número de restos por vuelta y 13 el número de átomos incluido en el segmento comprendido en un enlace de hidrógeno. El avance por resto en la hélice es de  $1.5 \text{ \AA}$ , por lo que cada vuelta se encuentra separada verticalmente de la anterior por  $5.4 \text{ \AA}$ .

El modelo propuesto por Pauling (mostrado en la figura II.A.1) fue confirmado experimentalmente a raíz del estudio mediante reflexión de rayos X de cristales de hemoglobina y de fibras de queratina, habiendo sido la obtención de la estructura de alta resolución de

mioglobina la prueba concluyente. Dado que cada vuelta de hélice contiene 3.6 aminoácidos, sus cadenas laterales se encuentran dispuestas casi perpendicularmente al eje de la hélice y espaciadas aproximadamente cada  $100^\circ$ , por lo que las posibilidades de establecimiento de interacciones específicas entre ellas son máximas en las posiciones  $(i, i+4)$  e  $(i, i+3)$ . Las hélices en proteínas suelen encontrarse en regiones próximas a la superficie, por lo que puede observarse cierta tendencia a alternar el carácter hidrofílico e hidrofóbico de las cadenas laterales con una periodicidad de aproximadamente tres o cuatro restos, característica que dota a la estructura de un marcado carácter anfipático (Eisenberg et al., 1982).

A pesar de que estas hélices son una de las estructuras más regulares encontradas en proteínas, pueden observarse ciertos defectos o deformaciones que se traducen en pequeñas desviaciones del eje de la hélice de hasta  $20^\circ$ , pero que no alcanzan a romper el patrón característico de enlaces de hidrógeno. La presencia de restos de prolina en el interior de la hélice es la causa más comúnmente observada de este tipo de desviaciones, ya que, aunque es capaz de adaptarse correctamente a la primera vuelta de la hélice, el anillo establecido entre su cadena lateral y el esqueleto peptídico introduce especiales restricciones estéricas y anula su potencial como donador en un enlace de hidrógeno.

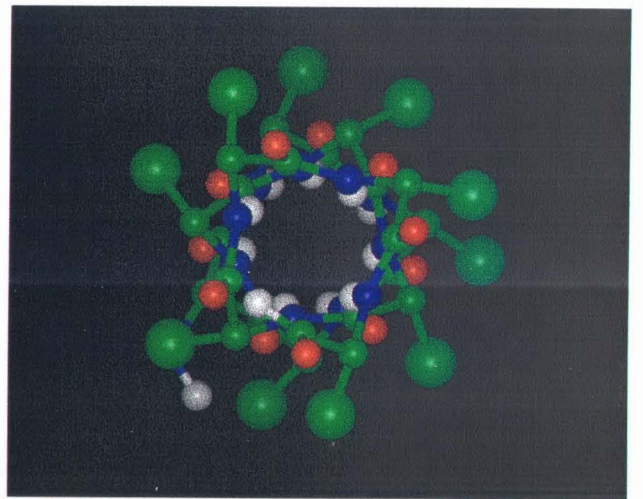
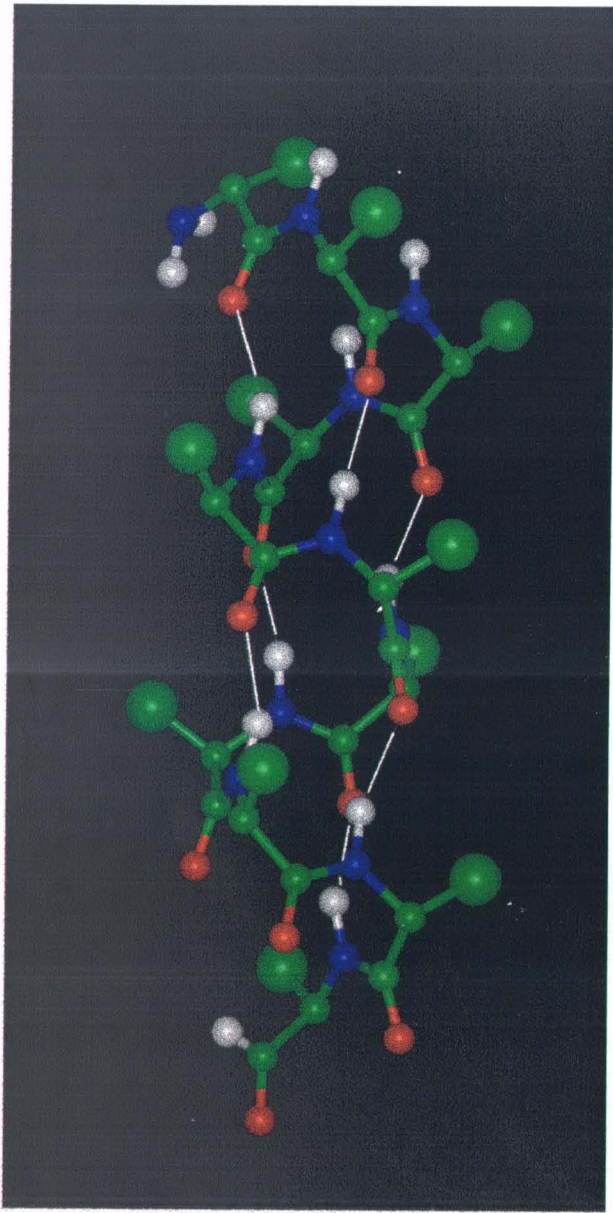
Debido a la elevada regularidad de la estructura, los dipolos de las unidades peptídicas se encuentran orientados en la misma dirección, a lo largo del eje de la hélice y con sus regiones positivas apuntando hacia el extremo amino-terminal. El efecto global de esta disposición es la existencia, incluso a pesar de los distintos efectos apantalladores, de un dipolo neto en la estructura helicoidal (Hol et al., 1978; Wada, 1976), que le confiere propiedades características, como la tendencia a unir especies cargadas. Diversas medidas de los momentos dipolares de cadenas polipeptídicas no cargadas, formadoras de hélices, en función de su longitud demuestra la existencia de dicho macrodipolo. El incremento del momento dipolar por aminoácido se estima en 3.2-3.4 D (Wada, 1976).

El carácter local de las interacciones establecidas en la hélice contrasta con la situación encontrada para la lámina beta, estructura secundaria también de carácter repetitivo en la que los puentes de hidrógeno del esqueleto se establecen entre grupos complementarios pertenecientes a hebras adyacentes, que pueden encontrarse muy lejanos en la secuencia. La frecuencia con que aparecen en proteínas, aproximadamente una cuarta parte de los restos de proteínas globulares de estructura conocida se encuentran en conformación helicoidal (Kabsch & Sander, 1983), así como el hecho de que sus enlaces de hidrógeno se encuentren confinados en el interior del segmento helicoidal, sugiere que las hélices alfa pueden actuar como unidades autónomas de plegamiento en proteínas. Esta hipótesis se ve reforzada por un número cada vez mayor de evidencias experimentales que demuestran la estabilidad en disolución acuosa de la estructura helicoidal tanto de fragmentos aislados de proteínas como de pequeños péptidos sintéticos.

#### II.A.1.3.-Propensiones helicoidales.

**Figura II.A.1: Modelo estructural de una hélice alfa.** Los átomos se han coloreado de acuerdo con su naturaleza química, de modo que las esferas verdes representan átomos de carbono, las azules de nitrógeno, las rojas de oxígeno, y las blancas los hidrógenos amida. El panel de la izquierda muestra una vista frontal de la hélice, donde las líneas blancas representan los enlaces de hidrógeno establecidos entre los átomos del esqueleto peptídico y las esferas verdes de mayor tamaño las posiciones de las cadenas laterales. El panel de la derecha muestra la hélice vista desde el extremo carboxilo-terminal, a lo largo de su eje principal.





Tanto para la predicción de estructuras secundarias a partir de la secuencia aminoacídica, como para el estudio del mecanismo de plegamiento de proteínas, es importante determinar si efectivamente existen diferencias en cuanto a la tendencia de cada uno de los aminoácidos a adoptar conformaciones helicoidales, así como las consecuencias que las posibles diferencias pudieran tener en la determinación de la estructura nativa. En este sentido, tanto los diversos análisis estadísticos de la base de datos de estructuras tridimensionales de proteínas (que han revelado la existencia de diferencias en las frecuencias con que los distintos aminoácidos aparecen formando parte de una hélice alfa) como los numerosos estudios experimentales llevados a cabo con polipéptidos y copolímeros sintéticos (que muestran la estabilización diferencial de las hélices por los distintos tipos de aminoácidos) apuntan hacia la existencia efectiva de diferentes tendencias a adoptar conformaciones helicoidales entre los distintos aminoácidos.

En general, con la excepción de prolina, cuyo nitrógeno amida presenta características especiales al encontrarse formando parte del anillo (por lo que carece de capacidad para establecer los enlaces de hidrógeno y adoptar los ángulos diedros característicos de la hélice) todos los aminoácidos son equivalentes en cuanto a su potencial para establecer enlaces de hidrógeno entre los grupos de su cadena principal. El origen de sus contribuciones diferenciales a la estabilidad de una hélice alfa debe radicar, pues, en la naturaleza de sus cadenas laterales. Por tanto, puede definirse la propensión helicoidal intrínseca de un aminoácido cualquiera como la medida de la influencia que su cadena lateral ejerce sobre la conformación del esqueleto peptídico como consecuencia de las interacciones de corto alcance establecidas entre dicha cadena lateral y los grupos del esqueleto, y de sus respectivas interacciones con el disolvente.

Las propensiones helicoidales intrínsecas de los distintos aminoácidos se han estimado utilizando muy distintos enfoques, que incluyen estudios estadísticos (Chou & Fasman, 1978; Williams et al., 1987), análisis "host-guest" (Wójcik et al., 1990), estudios sobre péptidos modelo (Lyu et al., 1990; Merutka et al., 1990; O'Neil & DeGrado, 1990; Padmanabhan et al., 1990), mutagénesis dirigida en proteínas (Horovitz et al., 1992), aplicación de redes neuronales (Hirst & Sternberg, 1992; Holley & Karplus, 1989; Kneller et al., 1990), y cálculos de dinámica molecular (Creamer & Rose, 1992; Hermans et al., 1992). La propensión helicoidal de un aminoácido determinado puede expresarse como su frecuencia de aparición en una posición determinada en hélices de proteínas, como la constante de equilibrio correspondiente a la adición de dicho aminoácido al final de una hélice previamente iniciada (equivalente a los parámetros de propagación de los modelos termodinámico-estadísticos<sup>a</sup>), o bien como la diferencia relativa en energía libre resultante de la sustitución de dicho aminoácido en una posición determinada de una proteína ( $\Delta\Delta G$ ).

---

<sup>a</sup>Ver sección XX donde se describen los formalismos termodinámico-estadísticos para la transición hélice-ovillo estadístico.

En general, además de las propensiones helicoidales intrínsecas de los aminoácidos que constituyen su secuencia, existen diversos factores que determinan la estabilidad de una hélice, siendo los más importantes las interacciones establecidas entre ciertas cadenas laterales y los grupos polares descompensados de los extremos amino y carboxilo de la hélice (interacciones de "capping") y las existentes entre las cadenas laterales. Entre estas últimas cabe destacar, los contactos de naturaleza hidrofóbica, las interacciones electrostáticas (especialmente las existentes entre las cadenas cargadas y el dipolo de la hélice), y otras interacciones específicas como la formación de pares iónicos o puentes de hidrógeno. En el caso de una hélice no aislada, inmersa en el contexto de una proteína, las interacciones terciarias de largo alcance con el resto de la estructura contribuyen notablemente a su estabilización. En resumen, la propensión de una cadena a adoptar una estructura secundaria determinada, que depende específicamente de su secuencia y surge a partir de interacciones de corto alcance, contribuye, junto con los contactos terciarios, a la estabilización de la estructura final de una proteína.

#### **II.A.2.-Modelos para la descripción termodinámico-estadística de la transición hélice-ovillo estadístico.**

A finales de los años cincuenta y principio de los sesenta se desarrollaron varios modelos, similares entre sí en cuanto a sus características principales, con objeto de explicar la transición de hélice-ovillo estadístico en cadenas polipeptídicas aisladas, habiendo sido los desarrollados por Zimm y Bragg (Zimm & Bragg, 1959) y Lifson y Roig (Lifson & Roig, 1961) los más ampliamente utilizados hasta el momento. Estos formalismos están basados en un modelo mecánico-estadístico para la transición, según el cual cada resto puede existir únicamente en dos conformaciones diferentes: hélice alfa y ovillo estadístico. Sin embargo, la transición global, que implica a toda la molécula, no consiste en un proceso de dos estados en el que únicamente intervienen conformaciones completamente helicoidales y ovillos estadísticos. Por el contrario, se trata de una transición entre un estado caracterizado por una distribución de moléculas, en su mayor parte helicoidales aunque fuertemente distorsionadas en sus extremos, y otro, en el que las moléculas se encuentran predominantemente constituidas por regiones desplegadas.

##### **II.A.2.1.-Modelo general para la formación de una hélice.**

La termodinámica de formación de una hélice alfa se basa fundamentalmente en su geometría, según la cual el grupo carbonilo de un enlace peptídico  $i$  se encuentra unido mediante un enlace de hidrógeno al nitrógeno amida del grupo peptídico en posición  $i+3$ . De este modo, el establecimiento de uno de estos enlaces requiere la inmovilización de cuatro unidades peptídicas consecutivas o, lo que es equivalente, la restricción de sus correspondientes ángulos  $\phi$  y  $\psi$  a las regiones helicoidales del espacio conformacional.



Desde un punto de vista energético, la adición de una unidad peptídica a una hélice ya existente implica una contribución a la energía libre procedente de la formación de un enlace de hidrógeno, una contribución entrópica negativa originada por la congelación de los grados de libertad del esqueleto peptídico, así como el cambio en energía libre correspondiente a su cadena lateral. Este último incluye las interacciones de la cadena lateral con el resto de la hélice así como el costo entrópico asociado a la reducción de su libertad conformacional, como consecuencia de la adopción de los valores helicoidales para los diedros del esqueleto peptídico. No se consideran, sin embargo, al menos en los tratamientos originales, las posibles interacciones específicas entre dos cadenas laterales. En resumen, la adición de un nuevo resto a la hélice lleva asociado un cambio en energía libre ( $\Delta G_s$ ) así como una constante de equilibrio ( $s$ ) que dependen de la naturaleza de la cadena lateral y de la temperatura..

La iniciación o nucleación de la hélice, es decir, la formación del primer enlace de hidrógeno, es especial, ya que implica la restricción conformacional de tres restos aminoacídicos (o cuatro grupos peptídicos). Como consecuencia, el cambio de energía libre asociado a esta etapa contiene, además de la energía libre correspondiente a la propagación  $\Delta G_s$ , un término adicional,  $\Delta G_{nuc}$ , que refleja la existencia de un costo entrópico adicional que no se encuentra compensado por el establecimiento de los enlaces de hidrógeno correspondientes. Esta energía libre de exceso, considerada generalmente de naturaleza fundamentalmente entrópica, se representa como el parámetro de nucleación  $\sigma$ , de modo que la constante de equilibrio que describe la formación de una hélice de  $n_r$  restos tiene la forma  $K = \sigma s^{n_r - 2}$ . Este parámetro de nucleación es el responsable de la cooperatividad de la transición y se considera generalmente independiente de la temperatura y de la naturaleza de los aminoácidos implicados.

#### II.A.2.2.-Principales tratamientos mecanico-estadísticos.

En general, cabe distinguir entre dos tipos de modelos teóricos, que se diferencian en el criterio escogido para la definición de los estados de referencia. Por un lado tenemos las teorías que se sirven de la cuantificación de los enlaces de hidrógeno establecidos en cada estado para la asignación de sus pesos estadísticos (Gibbs & DiMarzio, 1959; Peller, 1959; Schellman, 1955a; Zimm & Bragg, 1959), en las que el estado de referencia consiste en un resto cuyo grupo amida no se encuentre implicado en la formación de ningún enlace. Por otro, aquéllas en las que se definen los distintos estados según su localización en el espacio conformacional (Lifson & Roig, 1961; Nagai, 1960).

##### II.A.2.2.1.-Modelo de Zimm y Bragg.

Para el desarrollo de este modelo los autores consideran el grupo peptídico (-CONH-) como unidad básica para la formación de la hélice. Ésta se inicia cuando el oxígeno carbonílico de la unidad peptídica ( $i$ ) establece un enlace de hidrógeno con el nitrógeno amida del resto en

posición  $(i+3)$ , de modo que los grupos carbonilo de los restos intermedios  $(i+1)$  e  $(i+2)$  quedan sin emparejar. El peso estadístico de esta etapa de nucleación es  $\sigma s$ , donde  $\sigma$  representa el parámetro de nucleación y  $s$  el de propagación.

En las etapas de propagación posteriores, el grupo carbonilo de la unidad peptídica  $(i+1)$  participa en el establecimiento de un enlace de hidrógeno con el grupo amida de la unidad peptídica  $(i+4)$ , siendo  $\sigma s^2$  el peso estadístico del estado con dos puentes de hidrógeno, y así sucesivamente.

#### II.A.2.2.2.-Modelo de Lifson y Roig.

En este caso, la teoría considera un aminoácido en conformación helicoidal o no helicoidal en función de los valores adoptados por los ángulos  $\phi$  y  $\psi$  de su esqueleto peptídico. En cierto modo, este modelo se encuentra mejor adaptado para el análisis de la formación de hélices peptídicas ya que, en este caso, la unidad básica en el proceso corresponde a cada uno de los restos aminoacídicos. Esto implica que el parámetro de propagación  $w$  se encuentre asociado a cada aminoácido o cadena lateral en lugar de a la formación de un enlace de hidrógeno entre dos restos.

En este modelo, la nucleación de la hélice requiere la existencia de tres restos contiguos en conformación helicoidal. El papel de nucleación de la hélice se asigna a ambos extremos, correspondiendo a cada uno de ellos un peso estadístico  $v$  y siendo la constante total de nucleación  $v^2$ .

#### II.A.2.2.3.-Comparación de ambas teorías.

En la figura II.A.2 se muestran esquemáticamente los códigos y pesos estadísticos asignados por los distintos modelos a cada posición en la hélice, de modo que pueden observarse las similitudes y diferencias entre ambas teorías. Como consecuencia de los distintas definiciones de los estados, los valores de los parámetros de propagación no son numéricamente equivalentes en ambos modelos, aunque es posible la comparación directa de sus parámetros (Lifson y Roig ( $w$  y  $v$ ) y Zimm y Bragg ( $s$  y  $\sigma$ )) mediante el empleo de las siguientes expresiones (Qian & Schellman, 1992):

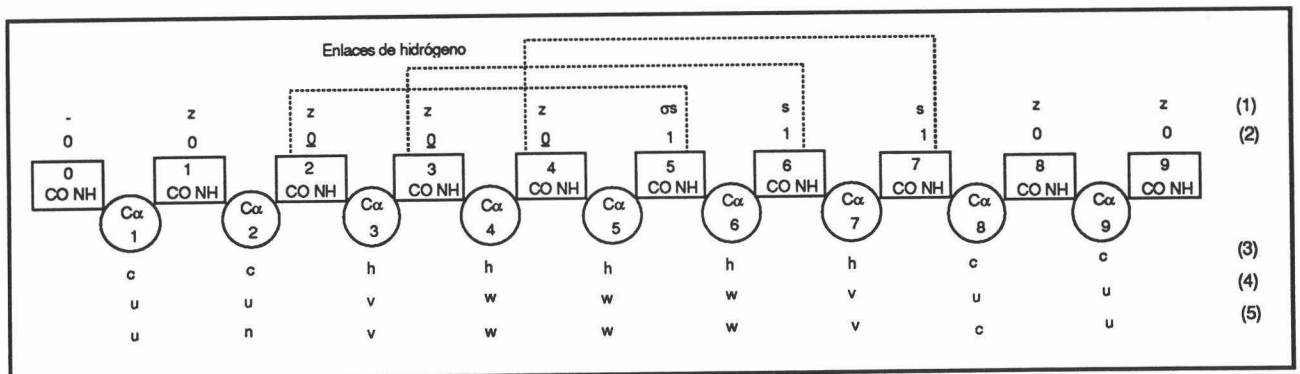
$$s = w / (1 + v) \quad (\text{II.A.1})$$

$$\sigma = v^2 / (1 + v)^4 \quad (\text{II.A.2})$$

Ambos formalismos difieren asimismo en su definición de la longitud de la cadena y de la nucleación de la hélice. En el modelo de Lifson y Roig, la longitud de la cadena se encuentra determinada por el número de restos que poseen enlaces peptídicos a ambos lados, mientras que el

modelo de Zimm y Bragg la define como el número de unidades amida de la cadena. Como consecuencia, en aquellos casos en que los péptidos se encuentran bloqueados en sus extremos amino y carboxilo terminal, el número de unidades amida excede en uno al número de restos. En el modelo de Zimm y Bragg, la iniciación de la hélice implica la formación del primer puente de hidrógeno entre los restos ( $i, i+3$ ), mientras que, por el contrario, la definición de Lifson y Roig implica que tres restos contiguos se encuentren en el espacio conformacional correspondiente a la hélice.

Sin embargo, a pesar de estos matices que los diferencian, ambas teorías coinciden en muchos aspectos como la definición de la función de partición para la transición hélice-ovillo estadístico mediante el empleo de matrices de correlación, procedimiento que permite el análisis de heteropolímeros que contienen restos con diferentes propensiones helicoidales (valores de los parámetros  $s$  ó  $w$ ). En ocasiones, aunque las teorías completas permiten la consideración de distintas regiones helicoidales en una única cadena, es frecuente recurrir a varias simplificaciones, como son el tratamiento de la cadena peptídica como un homopolímero y la consideración de un único segmento helicoidal. Estas aproximaciones no demasiado drásticas si se trabaja con cadenas polipeptídicas de pequeña longitud, en las que la etapa de nucleación es muy desfavorable.



**Figura II.A.2:** Secuencia de diez unidades peptídicas o nueve restos aminoacídicos que contiene un pequeño fragmento helicoidal, donde las unidades peptídicas se muestran como rectángulos y los restos aminoacídicos, centrados en los carbonos alfa, se representan como círculos. Las líneas punteadas muestran el patrón de puentes de hidrógeno característico de las hélices alfa. Como puede observarse, existen seis unidades peptídicas o cinco restos inmovilizados en la hélice. Se muestran asimismo los códigos (1) y (3) asignados por las teorías de Zimm y Bragg y de Lifson y Roig, respectivamente, a cada posición de la cadena, y los correspondientes pesos estadísticos (2) y (4). La última línea representa los pesos estadísticos de acuerdo con la teoría de Lifson y Roig modificada para introducir los efectos de "capping".

Una vez asignados los pesos estadísticos a cada una de las posiciones en la hélice, la función de partición (definida como la sumatoria de todos ellos) permite obtener de forma sencilla la fracción helicoidal, lo cual posibilita la comparación de los resultados teóricos con datos

experimentales, obtenidos mediante la aplicación de técnicas estructurales como difracción circular o resonancia magnética nuclear al estudio de las transiciones conformacionales de péptidos helicoidales.

Ambos formalismos han demostrado su capacidad para describir adecuadamente las transiciones hélice-ovillo estadístico, tanto en el caso de homopolipéptidos de larga longitud en disolventes orgánicos (a los que fueron aplicadas originariamente), como de disoluciones acuosas de pequeños péptidos, aún a pesar de la influencia de los extremos de la cadena y de la posible dependencia de los parámetros con la naturaleza de los aminoácidos adyacentes. El éxito de ambas teorías justifica su utilización para la determinación experimental de las propensiones helicoidales de los distintos aminoácidos a partir del análisis de los resultados obtenidos con pequeños péptidos en disolución acuosa.

#### **II.A.2.3.-Modificaciones de los modelos originales para la incorporación de distintas interacciones.**

El estudio de la estabilidad estructural de pequeños péptidos de secuencias variadas, bien características de fragmentos helicoidales en proteínas, o bien diseñadas con el propósito de analizar la contribución de interacciones específicas a la estabilidad de la estructura helicoidal, presenta un especial interés. Sin embargo, la aplicación de los modelos mecánico-estadísticos, desarrollados en un principio para el tratamiento de homopolipéptidos de gran longitud, al estudio de pequeños péptidos helicoidales de secuencias variadas requiere la introducción de ciertas modificaciones en el formalismo.

En primer lugar, la interpretación de los resultados obtenidos de los experimentos de sustitución en pequeños péptidos de secuencia conocida requiere la extensión de las teorías para su empleo en heteropolímeros. En este sentido, el formalismo desarrollado por Qian (Qian, 1993) permite la estimación del efecto que la sustitución de un resto determinado produce en un homopolímero de longitud finita, llegando a la conclusión de que la estabilidad de un péptido cualquiera con respecto a la sustitución de uno de sus restos se encuentra determinada fundamentalmente por dos factores: la propensión helicoidal intrínseca del sustituyente y la susceptibilidad del sitio en el que se realiza la sustitución, dependiente de la naturaleza de los restos vecinos a éste.

En los dos modelos estadísticos originales, los restos en posiciones N-cap y C-cap (definidas con detalle en la sección II.A.3.1.2), que se encuentran integrados sólo parcialmente en la hélice, se consideran en conformación desplegada, por lo que no contribuyen a la función de partición. Sin embargo, se ha observado experimentalmente que la naturaleza de los restos situados en estas posiciones contribuye notablemente a la modulación de la estabilidad de la hélice, por lo que la incorporación de este fenómeno al formalismo de Lifson y Roig permitiría cuantificar las variaciones en el contenido helicoidal asociadas a estas sustituciones. Doig y

colaboradores (Doig et al., 1994) demostraron que la teoría de Lifson y Roig puede modificarse de forma sencilla para permitir la consideración de estos efectos mediante la inclusión de los aminoácidos terminales (o de los grupos acetilo o amida, que en ocasiones se encuentran bloqueando los extremos) en la definición de la función de partición sin más que asignar a estos restos sus correspondientes pesos estadísticos. Esta operación podría interpretarse como la existencia de un proceso de nucleación de mayor complejidad que el considerado inicialmente.

Las teorías originales ignoran otros aspectos relevantes para la formación de hélices aisladas como pueden ser la interacción de grupos cargados con el macrodipolo de la hélice o las interacciones específicas establecidas entre cadenas laterales, por lo que ha sido necesario incorporar diversas modificaciones que posibiliten la cuantificación del efecto de estos fenómenos sobre la estabilidad y el contenido helicoidal de un péptido determinado (Gans et al., 1991; Scholtz et al., 1993; Shalongo & Stellwagen, 1995; Vázquez et al., 1987; Vázquez & Scheraga, 1988). Asimismo, se han desarrollado formalismos que permiten acoplar la formación de hélices peptídicas con el proceso de dimerización (Qian, 1994), de especial interés para el estudio de sistemas como las denominadas cremalleras de leucina ("leucine zippers") en las que ambos procesos se encuentran estrechamente relacionados (O'Neil & DeGrado, 1990).

### **II.A.3.-Escalas de propensión helicoidal.**

La propensión helicoidal de un aminoácido es una medida de la influencia que su cadena lateral ejerce sobre la conformación del esqueleto peptídico, y tiene su origen en las interacciones de corto alcance establecidas entre éste y las cadenas laterales, así como entre ambos y el disolvente. Son varios los posibles modos de expresar la propensión helicoidal de un aminoácido determinado, pudiendo identificarse con su frecuencia de aparición en hélices de proteínas, con el cambio de estabilidad relativo provocado por su introducción en la secuencia, o bien con los parámetros de propagación de las teorías de transición hélice-ovillo estadístico (valores  $s$  y  $w$  de las teorías de Lifson-Roig (Lifson & Roig, 1961) y Zimm-Bragg (Zimm & Bragg, 1959), respectivamente). En este último caso, un aminoácido puede clasificarse como formador de hélices, indiferente, o destructor, según su propensión helicoidal sea mayor que uno, uno, o menor que uno, respectivamente.

El establecimiento de escalas de propensión helicoidal en las que se encuentre reflejado el potencial formador de hélices de cada cadena lateral es importante para evaluar las contribuciones de los distintos aminoácidos a la estabilidad de proteínas así como para predecir la localización de las hélices en la estructura de una proteína a partir de su secuencia. Numerosos grupos de investigación han abordado la determinación de dichas propensiones helicoidales en sistemas de naturaleza muy diversa (desde copolímeros sintéticos a pequeños péptidos helicoidales o proteínas), empleando asimismo una gran variedad de técnicas, que incluyen desde correlaciones estadísticas hasta dicroísmo circular, resonancia magnética nuclear, o incluso dinámica molecular.



### II.A.3.1.-Correlaciones estadísticas: Frecuencias de aparición en hélices proteicas.

En los años sesenta, tras la resolución de las primeras estructuras tridimensionales de proteínas, se observó cierta correlación entre la frecuencia con que cada aminoácido aparecía en conformación helicoidal en las proteínas de estructura conocida y su tendencia a formar hélices en polipéptidos sintéticos. Como consecuencia de dicha observación, surge la idea de la existencia de una relación entre la frecuencia con que un aminoácido aparece en una posición determinada en la estructura y su efecto estabilizador en dicha posición.

#### II.A.3.1.1.-Análisis estadístico de Chou y Fasman.

Chou y Fasman mejoraron y extendieron correlaciones anteriores, más incompletas, con el objeto de estimar las propensiones helicoidales de las distintas cadenas laterales a partir de su frecuencia de aparición en las secuencias helicoidales de las proteínas de estructura conocida (Chou & Fasman, 1974; Chou & Fasman, 1978; Levitt, 1978), proponiendo la existencia de un paralelismo entre el mecanismo de formación de hélices en proteínas y en pequeños péptidos (Chou & Fasman, 1974). El análisis estadístico de la frecuencia con que los aminoácidos naturales se encuentran formando parte de diferentes estructuras secundarias sugiere que las distintas cadenas laterales poseen diversas preferencias conformacionales (Chou & Fasman, 1978). Este tipo de análisis, sin embargo, no tiene en consideración las restricciones en la secuencia debidas a fenómenos como el elevado empaquetamiento en la estructura nativa, la tendencia de los restos hidrofóbicos a ocupar posiciones internas en la estructura tridimensional, o las preferencias de algunos aminoácidos por encontrarse en las posiciones cercanas a los extremos de la hélice.

#### II.A.3.1.2.-Otros estudios estadísticos: Consideración de las propensiones N-cap y C-cap.

El aumento del número de estructuras tridimensionales en la base de datos ha permitido el desarrollo de estudio estadísticos que describen las preferencias de cada aminoácido en las distintas posiciones de la hélice con respecto a sus extremos (Presta & Rose, 1988; Richardson & Richardson, 1988). La determinación de la posición de éstos es una de las cuestiones de importancia en este tipo de tratamientos, existiendo diversas posibilidades. Así, los extremos de una hélice alfa pueden definirse de acuerdo a varios criterios, como los valores adoptados por los ángulos  $\phi$  y  $\psi$  del esqueleto peptídico (IUPAC-IUB, 1970), el patrón de enlaces de hidrógeno (IUPAC-IUB, 1970, Kabsch & Sander, 1983), o la posición de los carbonos alfa (Levitt & Greer, 1977).

Los restos que forman parte de la hélice y sus posiciones adyacentes se nombran de acuerdo con el esquema:



donde los restos N1, N2, N3, ..., C3, C2, C1 se encuentran en conformación helicoidal y contribuyen al establecimiento de la red de enlaces de hidrógeno de la hélice. Los restos N<sub>cap</sub> y C<sub>cap</sub>, sin embargo, aunque participan en la red de puentes de hidrógeno, presentan ángulos diedros no helicoidales, por lo que sólo pueden considerarse como parcialmente integrados en la estructura helicoidal. Por último, los restos en situados en las posiciones N'', N', C' y C'' se encuentran completamente fuera de la hélice.

Entre las preferencias más notables encontradas mediante el análisis de la base de datos (Presta & Rose, 1988; Richardson & Richardson, 1988) cabe destacar, la elevada frecuencia con que Gly aparece en la posición C<sub>cap</sub>, la ausencia de prolina en el interior de las hélices, así como su papel como iniciador de la hélice (como se desprende de su tendencia a encontrarse en posición N1). La asimetría en cuanto a las preferencias de los restos cargados por ocupar posiciones en los extremos de la hélice (los aminoácidos con carga negativa se encuentran preferentemente en el primer giro de la hélice y los de carga positiva en el último) constituye, a su vez, otra observación interesante.

La posición N-cap se encuentra frecuentemente ocupada por restos como Asn, Ser, Asp o Thr, que poseen la capacidad de establecer un enlace de hidrógeno con alguno de los grupos NH descompensados de la primera vuelta de la hélice. Como es de esperar, estos mismos aminoácidos son relativamente poco frecuentes en el interior de la hélice debido a la posibilidad de desplazar a los grupos del esqueleto peptídico de sus enlaces de hidrógeno. De acuerdo con la hipótesis de Presta y Rose (Presta & Rose, 1988), basada en estas observaciones, la presencia de restos aminoácidos en posiciones adyacentes a los extremos de la hélice con cadenas laterales que puedan proporcionar grupos complementarios a los amida y carbonilos desapareados del esqueleto peptídico, representa una condición necesaria para la formación de la hélice. Por lo tanto, la localización de las hélices en proteínas dependería exclusivamente de la información contenida en la secuencia a nivel local.

### II.A.3.2.-Estudios con copolímeros sintéticos.

Los primeros estudios experimentales llevados a cabo por Scheraga y colaboradores se basaron en la introducción de los distintos aminoácidos en copolímeros aleatorios (fundamentalmente basados en hidroxibutil o hidroxipropil L glutamina, HBLG y HPLG) de distintas longitudes y secuencias (Sueki et al., 1984; Wójcik et al., 1990). El análisis de los datos experimentales se llevó a cabo mediante la aplicación de la teoría de Zimm y Bragg (Zimm & Bragg, 1959), determinándose, por tanto, los parámetros  $s$  y  $\sigma$  para cada aminoácido. Los resultados obtenidos llevaron a la conclusión de que todos los aminoácidos son, con la excepción de prolina (que por sus características especiales desestabiliza notablemente la estructura helicoidal (Altmann et al., 1990)), indiferentes a la formación de hélices.

El que las propensiones helicoidales adopten valores cercanos a la unidad (entre 0.6 y 1.3 a 20 °C) para todos los aminoácidos es indicativo de una estabilidad muy reducida para cualquier hélice aislada en disolución. De hecho, la aplicación de la teoría de Zimm y Bragg (considerando los parámetros determinados experimentalmente para estos sistemas) predice la imposibilidad de observar formación alguna de hélices en disolución acuosa para cadenas polipeptídicas de menos de 20 aminoácidos de longitud (Shoemaker et al., 1985).

Sin embargo, estos resultados se encuentran en contradicción con los obtenidos a partir del estudio de pequeños péptidos ricos en Ala. Para estos sistemas, estables, aunque marginalmente, en disolución acuosa, se han obtenido valores de propensión helicoidal distribuidos en un amplio intervalo energético, mayor de 1 kcal·mol<sup>-1</sup>. Asimismo, tampoco se ha encontrado una buena correlación con los valores obtenidos con estos copolímeros sintéticos y los procedentes del análisis estadístico de la base de datos de proteínas.

### **II.A.3.3.-Estudios con péptidos monoméricos de bajo peso molecular.**

#### **II.A.3.3.1.-Interés del estudio de pequeños péptidos monoméricos.**

Como se ha comentado en el apartado anterior, la aplicación de las teorías de transición hélice-ovillo estadístico, considerando los parámetros obtenidos del estudio de las sustituciones en copolímeros sintéticos (Sueki et al., 1984), predice la inestabilidad de cualquier péptido aislado de longitud menor de veinte aminoácidos, que debería encontrarse completamente desestructurado en disolución acuosa (Bierzynski et al., 1982). Sin embargo, aún a pesar de la imposibilidad de detectar ninguna estructura residual en un gran número de fragmentos correspondientes a regiones helicoidales de varias proteínas, el descubrimiento de un pequeño grupo de estos péptidos que, aislados de la proteína, mantienen cierto porcentaje de estructura helicoidal en disolución (Brown & Klee, 1971; Hughson et al., 1991; Jiménez et al., 1993; Sancho et al., 1992; Waltho et al., 1993) despertó el interés sobre la aplicación de este tipo de sistemas al estudio de las propensiones helicoidales de los distintos aminoácidos.

Son de destacar los resultados obtenidos del estudio de los péptidos C y S, (correspondientes a los fragmentos 1-13 y 1-20, respectivamente, de Ribonucleasa A), que revelaron la existencia de interacciones estabilizadoras de tipo electrostático entre las cadenas laterales y el dipolo de la hélice (Armstrong et al., 1993; Fairman et al., 1990; Shoemaker et al., 1990; Shoemaker et al., 1985; Shoemaker et al., 1987), de gran importancia para la estabilidad estructural del péptido, y que no habían sido consideradas en los tratamientos teóricos. El análisis de los efectos producidos por numerosas sustituciones realizadas, tanto en estos péptidos, como en otros de naturaleza sintética y características similares, permitió demostrar la existencia de diferencias entre las propensiones helicoidales de los distintos aminoácidos de magnitud notablemente mayor que las derivadas inicialmente a partir de los trabajos con copolímeros



sintéticos. Estos resultados revelaron la importancia que dichas propensiones tienen en la determinación de la estabilidad estructural de los fragmentos peptídicos helicoidales (Chakrabarty et al., 1991, Padmanabhan et al., 1990).

El estudio de péptidos de pequeña longitud, que forman hélices monoméricas en disolución, se ha convertido en una práctica común para el análisis de las propensiones helicoidales de los aminoácidos (Chakrabarty et al., 1991; Gans et al., 1991; Kemp et al., 1991; Lyu et al., 1990; Merutka et al., 1990; Padmanabhan et al., 1990; Stellwagen et al., 1992). Dada su simplicidad estructural, este tipo de péptidos proporcionan sistemas modelo útiles, ya que facilitan el examen de las distintas fuerzas que intervienen en la estabilización de los elementos de estructura secundaria en proteínas. De este modo, se evitan las complicaciones asociadas al estudio de sistemas más complejos como proteínas o hélices enrolladas ("coiled-coils"), en los que la presencia de interacciones terciarias y el elevado grado de empaquetamiento pueden enmascarar los efectos que las propensiones helicoidales intrínsecas de los aminoácidos puedan tener en la estabilidad. Como contrapartida, el principal inconveniente de este tipo de sistemas reside en la complejidad de interpretación de los resultados, ya que, como demuestran numerosas evidencias experimentales, el proceso de desplegamiento de una hélice aislada en disolución no sigue un modelo de dos estados, por lo que su análisis requiere la aplicación de modelos teóricos multiestado para la transición hélice-ovillo estadístico.

#### *II.A.3.3.2.-Necesidad de la aplicación de teorías mecanico-estadísticas al análisis de transiciones hélice-ovillo estadístico en pequeños péptidos.*

La aplicación de cualquiera de los modelos teóricos (Zimm-Bragg, Lifson-Roig o alguna de sus modificaciones) a la simulación del proceso de desplegamiento de una hélice peptídica aislada en disolución revela que no se trata de una transición entre dos estados, ya que la helicidad no se encuentra uniformemente distribuida a lo largo del péptido. De este modo, únicamente un pequeño porcentaje de las moléculas en disolución adoptan conformaciones completamente helicoidales. Por el contrario, la mayoría de ellas presentan diversos grados de distorsión, especialmente notable en las regiones terminales, de modo que los segmentos helicoidales se encuentran confinados fundamentalmente a las posiciones centrales de la hélice (Chakrabarty & Baldwin, 1995; Schellman, 1958). La transición, inducida mediante el incremento de la temperatura o de la concentración de agentes desnaturizantes, provoca el ensanchamiento de la distribución de probabilidad con respecto a la longitud de los segmentos helicoidales y la consecuente disminución de su longitud promedio.

Para comprobar experimentalmente estas predicciones sería necesario cuantificar la probabilidad que cada uno de los restos en el péptido tiene de encontrarse en conformación helicoidal; magnitud que puede determinarse mediante técnicas como resonancia magnética nuclear (más concretamente técnicas de intercambio de hidrógeno-deuterio), que permiten

obtener información sobre restos individuales. Las medidas obtenidas mediante dicroísmo circular, por el contrario, son un reflejo de la helicidad total del péptido, por lo que, dado que no se trabaja con un sistema de dos estados, la obtención de la propensión helicoidal de un aminoácido particular a partir de su análisis requiere el empleo de alguno de los modelos teóricos mencionados anteriormente.

Numerosos estudios sobre la formación de hélices alfa mediante métodos de resonancia magnética nuclear (medidas de los desplazamientos químicos de los protones del carbono alfa, constantes de acoplamiento o señales N.O.E.) han demostrado que la distribución de restos helicoidales no es uniforme en toda la cadena (Bradley et al., 1990; Gans et al., 1991; Liff et al., 1991; Osterhout et al., 1989). Aunque estos métodos proporcionan información específica de cada aminoácido y corroboran la idea de que los péptidos helicoidales se encuentran fuertemente distorsionados en sus extremos, no pueden ser fácilmente utilizados para llevar a cabo un análisis de carácter cuantitativo. Por el contrario, las medidas de intercambio de hidrógeno-deuterio de los protones amida permiten observar y cuantificar la distribución de restos helicoidales en la cadena peptídica, de modo que hacen posible establecer una relación entre las probabilidades de formación de enlaces de hidrógeno y los parámetros estadísticos de Zimm y Bragg (Rohl et al., 1992). Los resultados obtenidos, junto con la notable variación con la posición en la hélice de los efectos que las sustituciones Ala->Gly tienen en la estabilidad de péptidos ricos en Ala (Chakrabarty et al., 1991), concuerdan con la hipótesis de la existencia de importantes distorsiones en los extremos de la hélice (Kuwajima & Baldwin, 1983; Rohl et al., 1992; Wand et al., 1986). Como consecuencia, queda comprobada la imposibilidad de aplicar el modelo de dos estados al análisis de estos procesos así como la idoneidad de los tratamientos mecánico-estadísticos para el estudio del desplegamiento de pequeñas hélices peptídicas.

#### *II.A.3.3.3.-Principales sistemas peptídicos utilizados para el estudio de propensiones helicoidales.*

El estudio de las propensiones helicoidales se ha abordado en diversos sistemas peptídicos de bajo peso molecular y carácter monomérico, que se diferencian entre sí en el número de interacciones existentes entre las cadenas laterales de los aminoácidos que constituyen la hélice y que contribuyen a su estabilización. Sin embargo, el procedimiento seguido para estimar las propensiones helicoidales es común a todos ellos, y consiste en la introducción del aminoácido correspondiente en una o varias posiciones de la hélice, y el posterior análisis, haciendo uso de las teorías de transición hélice-ovillo estadístico, del cambio en contenido helicoidal (medido generalmente mediante técnicas de dicroísmo circular) producido por la sustitución.

Uno de los resultados de mayor relevancia desprendido de los primeros estudios con este tipo de sistemas es la excepcionalmente elevada propensión helicoidal de la alanina (Marqusee et al., 1989). Éste es el único aminoácido capaz de formar hélices estables en ausencia de

interacciones estabilizadoras entre las cadenas laterales o bien entre éstas y el dipolo de la hélice. La utilización de péptidos ricos en Ala, solubilizados mediante la inclusión de algunos restos Lys o Gln, ha servido de base para el estudio de las propensiones helicoidales de cada uno de los veinte aminoácidos (Chakrabartty et al., 1994; Chakrabartty et al., 1991; Padmanabhan & Baldwin, 1991; Padmanabhan et al., 1990). De igual modo, ha proporcionado información sobre los efectos asociados a muy diversos tipos de interacciones, como pueden ser las interacciones de "capping", las interacciones electrostáticas con el dipolo de la hélice, o las interacciones específicas entre dos cadenas laterales (Armstrong & Baldwin, 1993; Armstrong et al., 1993; Chakrabartty et al., 1993b; Doig et al., 1994; Huyghues-Despointes et al., 1993; Padmanabhan et al., 1994; Scholtz et al., 1993). El interés especial de este tipo de sistemas reside en la ausencia de cadenas laterales que puedan interactuar entre sí o con el aminoácido introducido, de modo que el análisis e interpretación de los resultados experimentales se simplifican (Chakrabartty et al., 1994).

La estrategia alternativa al alto contenido en Ala para conseguir hélices estables en disolución consiste en introducir, en las posiciones adecuadas, cadenas laterales con capacidad para establecer interacciones estabilizadoras, como puentes salinos, interacciones de "capping", etc. Siguiendo este criterio, se han determinado las propensiones de numerosos aminoácidos en péptidos altamente polares, estabilizados por numerosos puentes salinos solapados entre restos Lys<sup>+</sup> y Glu<sup>-</sup> (Gans et al., 1991; Lyu et al., 1990; Lyu et al., 1989). Otros sistemas similares, de naturaleza intermedia entre los dos primeros, hacen uso tanto del alto contenido en Alanina como de la introducción de interacciones específicas para conseguir la estabilización de las hélices en disolución (Merutka et al., 1990; Park et al., 1993a; Park et al., 1993b; Stellwagen et al., 1992).

Muy recientemente se han evaluado, asimismo, las propensiones helicoidales de los aminoácidos substituidos en un péptido (correspondiente a una de las hélices de Ribonucleasa T1), que muestra cierta estabilidad en disolución (Myers et al., 1997a). Paralelamente, se han determinado los efectos de substituciones equivalentes en la proteína completa, de modo que en este caso ha sido posible realizar una comparación directa de las propensiones helicoidales en sistemas peptídicos y proteínas. Los valores obtenidos son muy similares en ambos sistemas de lo que se desprende que, en ciertas ocasiones y en contextos similares, las propensiones helicoidales de los aminoácidos en péptidos y proteínas son equivalentes.

#### **II.A.3.4.-Estudios con proteínas y hélices enrolladas (coiled-coils).**

##### **II.A.3.4.1.-Consideraciones generales.**

Como se ha mencionado en los apartados anteriores, las hélices peptídicas aisladas son sólo marginalmente estables en disolución, encontrándose fuertemente distorsionadas en los extremos. La situación en proteínas es diferente, dado que en estos sistemas la estabilidad de las hélices está determinada en gran medida por las interacciones terciarias establecidas con el resto

de la molécula. En este caso, la naturaleza cooperativa de su desplegamiento permite el análisis de los datos experimentales haciendo uso de la aproximación de dos estados.

Como contrapartida, la complejidad estructural de las proteínas puede complicar la interpretación de los resultados obtenidos, ya que la sustitución de un aminoácido en una posición determinada produce con frecuencia cambios en las conformaciones de las cadenas laterales vecinas, que pueden conducir a la pérdida de interacciones previamente existentes así como al establecimiento de otras nuevas. El costo energético de este posible reajuste estructural se alteraría la medida de las propensiones helicoidales (Bell et al., 1992), por lo que, aún a pesar de la imposibilidad de corregir cuantitativamente estos efectos, es recomendable el desarrollo de estudios cristalográficos paralelos que permitan la identificación de posibles alteraciones estructurales (Bell et al., 1992; Blaber et al., 1993);

El procedimiento seguido habitualmente consiste en la sustitución de un resto, situado en una posición helicoidal apropiada, por una nueva cadena lateral mediante técnicas de mutagénesis dirigida, y en la posterior evaluación del cambio en estabilidad frente al desplegamiento provocado por la mutación. De la diferencia en energía libre entre ambos mutantes puede obtenerse la relación entre las propensiones helicoidales (entendidas como el parámetro  $s$  ó  $w$  de los modelos teóricos), de modo que:

$$\Delta G^\circ(\text{mutante A}) - \Delta G^\circ(\text{mutante B}) = RT \cdot \ln \frac{S_{\text{aminoácido A}}}{S_{\text{aminoácido B}}} \quad (\text{II.A.3})$$

Como puede observarse, el análisis de los efectos de las distintas sustituciones en hélices de proteínas sólo permite la determinación de las diferencias en propensión helicoidal entre pares de aminoácidos, por lo que, en este caso, no es posible obtener valores absolutos de las tendencias helicoidales de los distintos aminoácidos. Dado que glicina es uno de los aminoácidos, con la excepción de prolina, con menor propensión helicoidal, es común su elección como referencia, por lo que muy frecuentemente las propensiones helicoidales se expresan como el cambio en estabilidad relativo al mutante de Gly, de forma que  $\Delta\Delta G = \Delta G^\circ(\text{mutante X}) - \Delta G^\circ(\text{mutante Gly})$ .

Este procedimiento de análisis supone que las diferencias en propensiones helicoidales son responsables de la totalidad del cambio en estabilidad entre ambos mutantes y que el resto sustituido no participa en interacciones con otras cadenas laterales o con el esqueleto peptídico. Por tanto, para llevar a cabo este tipo de estudios en proteínas es necesario elegir cuidadosamente la posición en la que realizar las mutaciones, de forma que las diferencias en propensión helicoidal entre los aminoácidos sean el único factor determinante de los cambios en estabilidad asociados a la mutación. El sitio de sustitución ideal debe encontrarse en el tramo central de una hélice regular y relativamente larga, para evitar posibles interferencias ocasionadas por el establecimiento de interacciones específicas con los extremos de la hélice ("capping"). Asimismo, debe encontrarse completamente expuesto al disolvente y suficientemente alejado de grupos

cargados. Debe también carecer de posibilidades de contacto con cadenas laterales cercanas y encontrarse alejado de contactos cristalinos que pudieran distorsionar la conformación adoptada por su cadena lateral en el cristal y obstaculizar el estudio estructural.

#### II.A.3.4.2.-Diferentes sistemas proteicos utilizados para el estudio de propensiones helicoidales.

Aún a pesar del elevado porcentaje de restos en conformación helicoidal en proteínas globulares, aquéllos que cumplen los requisitos necesarios para constituir un buen sitio en el que llevar a cabo el estudio mutacional son muy escasos. De entre los trabajos llevados a cabo con proteínas cabe destacar los realizados en las posiciones Ala-32 en Barnasa (Horovitz et al., 1992), Ser-44 y Val-131 en Lisozima T4 (Blaber et al., 1994) y Ala-21 en Ribonucleasa T1 (Myers & Pace, 1996; Myers et al., 1997a).

La Ala 32 se encuentra ubicada en la segunda hélice de Barnasa, formada por los restos 26 a 34 y de secuencia TKSEAQALG. Dado que el aminoácido en esta posición se encuentra completamente expuesto al disolvente y no participa en interacciones con el corazón de la molécula, Fersht y colaboradores (Horovitz et al., 1992) llevaron a cabo la sustitución de Ala-32 por los 19 aminoácidos, estimando sus propensiones helicoidales mediante la determinación de las estabilidades de cada uno de los mutantes a partir de experimentos de desnaturalización en urea a 25° C seguidos mediante fluorescencia.

En el caso de Lisozima T4, las sustituciones se han realizado en la posición 44, situada en el interior de una hélice de secuencia Leu-Asn-Ala-Ala-Lys-Ser-Glu-Leu-Asp-Lys-Ala-Ile y que comprende los restos 39 a 50 en la proteína, así como en la posición 131 correspondiente a la hélice constituida por los restos 126 a 134 y que tiene la secuencia Trp-Asp-Glu-Ala-Ala-Val-Asn-Leu-Ala. Aunque la primera de estas posiciones (Ser 44) cumple con todos los requisitos necesarios para llevar a cabo al análisis mutacional, en el segundo caso existen ciertos inconvenientes (los valores adoptados por los ángulos del esqueleto peptídico, se encuentran próximos al intervalo característico de las hélices  $3_{10}$  y existe la posibilidad de establecimiento de interacciones específicas con otras cadenas laterales), por lo que únicamente se ha llevado a cabo un número muy limitado de sustituciones. Las propensiones helicoidales se determinaron a partir de experimentos de desnaturalización térmica y se expresaron como valores de  $\Delta\Delta G^\circ$  medidos a la temperatura de transición conformacional,  $T_m$ , de la proteína.

Así mismo, se han sustituido los 19 aminoácidos en la posición 21 de Ribonucleasa A, correspondiente a la hélice formada por los restos 13 a 29 de secuencia SSDVSTAQAAGYKLHED. Esta posición posee muy buenas características, dado que se encuentra en el centro exacto de la hélice, en su cara expuesta al disolvente y con pocas probabilidades de establecer contactos con otras cadenas laterales, ya que los restos en posición  $(i+3)$  e  $(i+4)$  se encuentra implicados en otras interacciones. La estabilidad de los distintos



mutantes se determina mediante desnaturalización con urea a 25 °C seguida por fluorescencia. El análisis de las curvas de desnaturalización se llevó a cabo empleando el modelo de desplegamiento de dos estados y el método de extrapolación lineal.

#### II.A.3.4.3.-Hélices enrolladas ("Coiled coils").

Como se ha discutido anteriormente, uno de los problemas fundamentales que presenta el estudio de propensiones helicoidales en proteínas proviene de la complejidad de interpretación de los cambios en estabilidad asociados a las substituciones, debido a posibles interferencias procedentes de cambios conformacionales o a la modificación de las interacciones como consecuencia de las mutaciones, que pudieran enmascarar las contribuciones de las propensiones intrínsecas. El empleo de sistemas de mayor simplicidad, pero suficientemente cooperativos como para permitir la aplicación del modelo de dos estados al análisis de los resultados experimentales, constituye una alternativa interesante para la determinación de las propensiones intrínsecas de los distintos aminoácidos.

Aunque la posibilidad de aplicación del modelo de dos estados a su estudio aún es motivo de discusión, el desplegamiento de pequeños péptidos "coiled-coil" es bastante cooperativo, habiéndose obtenido resultados muy similares tanto aplicando dicho modelo como tratamientos mecánico-estadísticos (Qian, 1994). De este modo, O'Neil y DeGrado (O'Neil & DeGrado, 1990) han medido las propensiones helicoidales de los veinte aminoácidos a partir de los experimentos de substitución llevados a cabo en un péptido con las características correspondientes a las cremalleras de leucinas, diseñado de forma que presente un comportamiento dímérico en disolución y se minimicen las interacciones no estrictamente locales entre los aminoácidos substituidos y el resto de la cadena. Como consecuencia, el sitio de substitución se encuentra expuesto al disolvente en el estado dímérico y rodeado de aminoácidos con cadenas laterales no cargadas y de pequeño tamaño. Las propensiones helicoidales se determinaron mediante el estudio de la dependencia con la concentración de la formación de hélices en 5 M de urea. El análisis de los datos se llevó a cabo considerando un modelo de dos estados para la transición entre un dímero helicoidal y monómeros desplegados, de modo que las propensiones helicoidales se expresan como los  $\Delta\Delta G^\circ$  correspondientes a los cambios en la estabilidad de los dímeros con respecto al mutante de Gly.

#### II.A.3.5.-Estudios teóricos y semiempíricos.

Paralelamente a los estudios de tipo experimental, se han utilizado numerosas aproximaciones teóricas para analizar las propensiones helicoidales y los determinantes de la estabilidad de las hélices alfa en disolución. De este modo, diversos autores han estudiado el proceso de formación y estabilización de hélices modelo o fragmentos helicoidales de proteínas



mediante el empleo de simulaciones de dinámica molecular (Creamer & Rose, 1992; Creamer & Rose, 1994; Hermans et al., 1992; Tirado-Rives & Jorgensen, 1991; Yang & Honig, 1995) o análisis de Monte Carlo (Okamoto, 1994).

De igual modo, se ha recurrido a la utilización de métodos de carácter semiempírico para llevar a cabo el análisis de las propensiones helicoidales con objeto de establecer ciertas relaciones que permitan predecir el grado de helicidad de un péptido determinado a partir de su secuencia. Entre ellos cabe destacar el enfoque de Ooi y colaboradores (Ooi & Oobatake, 1991), que hacen uso de su parametrización, previamente desarrollada para el análisis de la termodinámica de plegamiento de proteínas, de las distintas magnitudes termodinámicas en función de los cambios en área accesible al disolvente. Así, los autores predicen un cambio de entalpía asociado a la formación de una hélice de polialanina muy similar al obtenido experimentalmente (Scholtz et al., 1991), y la posibilidad de obtener hélices de pequeña longitud marginalmente estables en disolución.

De especial interés es el algoritmo AGADIR desarrollado por Muñoz y Serrano (Muñoz et al., 1995a; Muñoz & Serrano, 1994; Muñoz & Serrano, 1995b; Muñoz & Serrano, 1995c; Muñoz & Serrano, 1995d) mediante el análisis estadístico de numerosos resultados experimentales con péptidos helicoidales en disolución. Estos sistemas permiten obtener un juego de parámetros que, empleados dentro del contexto de un modelo termodinámico-estadístico adecuado, posibilitan la predicción con elevada precisión del contenido helicoidal de un péptido cualquiera en disolución.

La similitud entre los resultados obtenidos a partir de estos estudios y las propensiones procedentes de los trabajos experimentales valida su utilización para llevar a cabo un análisis detallado de los diferentes factores responsables de las preferencias conformacionales de cada aminoácido y de las diferencias en estabilidad de las hélices de las que forman parte. De este modo, a pesar del origen fundamentalmente entálpico de la estabilidad de la hélice (Ooi & Oobatake, 1991; Scholtz et al., 1991), se ha identificado la diferencia en libertad conformacional entre las distintas cadenas laterales en conformación helicoidal como uno de los orígenes y principal determinante de las diferentes tendencias helicoidales de los aminoácidos. Asimismo, se ha demostrado que la diferencia en propensión helicoidal entre Ala y Gly puede ser atribuida casi en su totalidad al mayor espacio conformacional accesible para Gly en el estado desplegado, lo cual desfavorece notablemente la formación de la hélice. Aunque en menor grado, otros factores como el establecimiento de interacciones no covalentes, la hidratación diferencial de los distintos restos, así como el potencial para establecer interacciones específicas (puentes iónicos, enlaces de hidrógeno o contactos hidrofóbicos) entre sus cadenas laterales y el resto de la hélice son de importancia en la determinación de la propensión helicoidal de cada aminoácido.

#### **II.A.4.-Principales conclusiones de los estudios experimentales.**

##### **II.A.4.1.-Intervalo de propensiones helicoidales.**

En la tabla II.A.1 se muestran las distintas escalas experimentales obtenidas con pequeños péptidos aislados, "coiled coils" y proteínas. Las cuatro primeras columnas corresponden a péptidos de pequeña longitud, estabilizados mediante su elevado contenido en Ala o bien mediante la inclusión de interacciones específicas entre algunas de sus cadenas laterales, especialmente puentes salinos entre los restos ( $i, i+4$ ) e ( $i, i+3$ ). La quinta y sexta columnas

**Tabla II.A.1: Escalas experimentales de propensión helicoidal**  
 $\Delta\Delta G$  (Kcal/mol)

Aminoácido	AK/AQ <sup>(1)</sup> (0 °C)	KEAKE <sup>(2)</sup> (0 °C)	K <sub>2</sub> AE <sub>2</sub> <sup>(3)</sup> (4 °C)	E <sub>4</sub> K <sub>4</sub> <sup>(4)</sup> (4 °C)	T1-Pep <sup>(5)</sup> (0 °C)	T1-21 <sup>(6)</sup> (25 °C)	Ba-32 <sup>(7)</sup> (25 °C)	T4-44 <sup>(8)</sup> (52 °C)	T4-131 <sup>(9)</sup> (52 °C)	C-Coil <sup>(10)</sup> (20 °C)	Cop. <sup>(11)</sup> (20 °C)
Ala	-1.97	-1.95	-1.05	-0.79	-0.97	-0.90	-0.91	-0.96	-0.94	-0.77	-0.35
Arg <sup>+</sup>	-1.75	-1.96	-0.64	-----	-0.50	-0.49	-0.77	-0.77	-----	-0.68	-0.33
Asn	-1.01	-0.83	-0.20	-0.18	-0.35	-1.24	-0.25	-0.39	-----	-0.07	-0.16
Asp <sup>o</sup>	-1.18	-0.85	-----	-----	-0.31	-1.23	-----	-0.42	-0.77	-----	-----
Asp <sup>-</sup>	-1.16	-0.85	-0.27	-----	-0.29	-0.19	-0.20	-----	-----	-0.15	-0.08
Cys	-1.06	-1.10	0.12	-----	-0.45	-0.16	0.09	-0.42	-----	-0.23	-0.30
Gln	-1.42	-1.33	-0.25	-0.48	-0.67	-0.57	-0.43	-0.80	-----	-0.33	-0.30
Glu <sup>o</sup>	-1.49	-1.63	-----	-----	-0.80	-0.95	-----	-0.53	-0.88	-----	-----
Glu <sup>-</sup>	-1.35	-1.56	-0.84	-----	-0.66	-0.21	-0.36	-----	-----	-0.27	-0.29
Gly	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
His <sup>o</sup>	-1.13	-0.78	-0.51	-----	-0.30	-0.73	-----	-----	-----	-0.06	-----
His <sup>+</sup>	-0.86	-0.78	-----	-----	0.23	-0.34	-0.13	-0.57	-----	-----	-0.09
Ile	-1.26	-1.17	-0.67	-0.39	-0.63	-0.46	-0.10	-0.84	-0.84	-0.23	-0.39
Leu	-1.61	-0.64	-0.69	-0.62	-0.75	-0.77	-0.56	-0.92	-0.77	-0.62	-0.23
Lys <sup>+</sup>	-1.68	-1.75	-0.46	-----	-0.59	-0.39	-0.72	-0.73	-----	-0.65	-0.27
Met	-1.45	-1.50	-0.52	-0.57	-0.82	-0.75	-0.60	-0.86	-0.81	-0.50	-0.42
Phe	-0.97	-1.50	0.06	-----	-0.37	-0.33	-0.22	-0.59	-----	-0.41	-0.36
Ser	-1.18	-0.94	-0.29	-0.28	-0.50	-0.45	-0.50	-0.53	-0.64	-0.35	-0.15
Thr	-0.75	-0.70	-0.16	-0.23	-0.32	-0.33	-0.12	-0.54	-0.56	-0.11	-0.19
Trp	-1.01	-1.33	0.06	-----	-0.89	-0.60	0.07	-0.58	-----	-0.45	-0.37
Tyr	-1.28	-1.17	0.90	-----	-0.59	-0.51	-0.09	-0.72	-----	-0.17	-0.32
Val	-0.93	-0.70	-0.32	-0.34	-0.33	-0.24	-0.03	-0.63	-0.68	-0.14	-0.28

(1) Péptido de secuencia -AAK(ó Q)AXAAK(ó Q)A- : (Rohl et al., 1996).

(2) Péptido de secuencia -AKEAXAKEA- : (Chakrabarty & Baldwin, 1995; Park et al., 1993b).

(3) Péptido de secuencia -AKKAXAEEA- : (Yang et al., 1997).

(4) Péptido de secuencia -KKKKXXXEEE- : (Lyu et al., 1990).

(5) y (6): Hélice de secuencia -STAQXAAYK-, correspondiente a la región 13-29 de Ribonucleasa T1. (Myers et al., 1997b)

(7) Hélice de secuencia -KSAQXLG-, correspondiente a la región 26-34 de Barnasa (los restos  $x+3$  y  $x+4$  no forman parte de la hélice). (Horovitz et al., 1992)

(8) Hélice de secuencia -QAAKXELDK-, correspondiente a la región 39-50 de Lisozima T4. (Blaber et al., 1994)

(9) Hélice de secuencia -DEAAXNLA-, correspondiente a la región 126-134 de Lisozima T4 (el resto  $x+4$  no forma parte de la hélice). (Blaber et al., 1994)

(10) Hélice de secuencia -AALEXKLQA- integrada en la cremallera de leucina diseñada por O'Neill y DeGrado. (O'Neill & DeGrado, 1990)

(11) Copolímeros sintéticos basados en polihidroxibutil glutamina. (Wójcik et al., 1990)

contienen los datos obtenidos para un péptido helicoidal de secuencia correspondiente a una de las hélices de la Ribonucleasa T1, que mantiene cierto contenido helicoidal en disolución acuosa y los resultados de la substituciones en la posición equivalente de la proteína completa, respectivamente. Las restantes columnas muestran los valores de propensión helicoidal correspondientes a mutaciones en diversas proteínas y "coiled coils". Como puede observarse, existe una buena correspondencia entre los valores obtenidos con los distintos sistemas, especialmente si se consideran exclusivamente los aminoácidos no cargados, de modo que, en general, se mantiene el orden de tendencias helicoidales de los distintos aminoácidos en todos ellos.

De acuerdo con los resultados procedentes de los experimentos llevados a cabo con copolímeros de secuencia aleatoria (Wójcik et al., 1990) las diferencias en cuanto a propensiones helicoidales (parámetro  $s$  en la teoría de Zimm y Bragg) entre los distintos aminoácidos son pequeñas. Sin embargo, los resultados obtenidos del estudio de proteínas y pequeños péptidos aislados en disolución revelan diferencias de mayor magnitud (Chakrabarty et al., 1994; Chakrabarty et al., 1993b; Lyu et al., 1990; O'Neil & DeGrado, 1990; Padmanabhan et al., 1990; Park et al., 1993b).

La naturaleza de los copolímeros sintéticos, caracterizados por secuencias aleatorias, longitudes no definidas y grados de polimerización generalmente superiores a 100 restos, parece ser la responsable de su especial comportamiento. Las cadenas laterales de los restos de hidroxialquil-L-glutamina, componente mayoritario de estos sistemas, parecen estar implicadas en numerosas interacciones con otras cadenas laterales, y provocan una disminución de la sensibilidad de los experimentos de substitución a la naturaleza del aminoácido introducido. De este modo, los efectos que las propensiones intrínsecas de los distintos aminoácidos puedan tener en la estabilidad de la conformación helicoidal se encuentran atenuados (Myers et al., 1997b; Pace & Scholtz, 1997; Padmanabhan et al., 1994).

El intervalo energético en que se encuentran distribuidas las propensiones helicoidales en la mayoría de los sistemas, tanto proteínas como péptidos, es muy similar y cercano a 1 kcal·mol<sup>-1</sup>. Sin embargo, el comportamiento de algunos péptidos modelo discrepa de la tendencia general, de modo que los resultados obtenidos con los péptidos basados en alanina duplican las propensiones helicoidales encontradas en proteínas. Por otro lado, los péptidos estabilizados mediante puentes salinos consecutivos dan lugar a diferencias en propensión helicoidal entre los distintos aminoácidos un 30% menores que las correspondientes a proteínas, y, como ya se ha mencionado, los estudios con copolímeros proporcionan valores entre 2 y 3 veces más pequeños.

Como posible origen de la discrepancia entre los valores obtenidos del análisis de péptidos basados en Ala y otros sistemas peptídicos, se ha sugerido la posible existencia de distorsiones en la señal de dicroísmo circular debidas a la presencia de cadenas laterales aromáticas (Chakrabarty et al., 1993b), así como la utilización de formalismos teóricos para el análisis de los datos experimentales que no consideran la posibilidad de establecimiento de interacciones específicas entre cadenas laterales o efectos de capping (Chakrabarty et al., 1994; Doig et al., 1994). Algunos autores han propuesto, asimismo, la existencia de cierta relación entre la naturaleza de las

cadena laterales en las inmediaciones del sitio de sustitución y la magnitud de las propensiones helicoidales obtenidas en cada sistema (Marqusee et al., 1989; Myers et al., 1997a; Pace & Scholtz, 1997; Padmanabhan et al., 1994). De hecho, algunos experimentos realizados con hélices de carácter anfipático demuestran la existencia de fenómenos de este tipo, como puede deducirse a partir de la diferencia observada entre los efectos que una misma sustitución tiene en ambas caras, hidrofílica e hidrofóbica, de la hélice (Zhou et al., 1993).

Se ha sugerido la posibilidad de que las diferencias observadas entre los distintos sistemas tengan su origen en la naturaleza de los aminoácidos que los constituyen, y concretamente en el efecto de las cadenas laterales en la hidratación del esqueleto peptídico (Myers et al., 1997b; Pace & Scholtz, 1997). En efecto, el intervalo energético en el que se distribuyen las propensiones helicoidales en los distintos sistemas parece encontrarse inversamente relacionado con el tamaño promedio de las cadenas laterales cercanas a la posición de sustitución.

Los péptidos basados en alanina están diseñados de modo que las posiciones adyacentes al sitio de sustitución así como las correspondientes a los restos  $(i+3)$  e  $(i+4)$  estén ocupadas por dicho aminoácido, de modo que, a consecuencia de su pequeño tamaño, su cadena lateral no interferiría, ni en el estado desplegado ni en la conformación helicoidal, en la hidratación de los grupos peptídicos. Por el contrario, en los péptidos y proteínas de secuencias naturales las cadenas laterales presentes en posiciones  $(i, i+1)$ ,  $(i, i+3)$  e  $(i, i+4)$  son de naturaleza heterogénea y generalmente de mayor tamaño. Asimismo, los péptidos modelo estabilizados mediante puentes salinos ( $E_4K_4$ ) están constituidos casi exclusivamente por glutámicos y lisinas, que poseen cadenas laterales voluminosas; los copolímeros sintéticos, que contienen cadenas laterales especialmente largas y voluminosas (hidroxipropil o hidroxibutil glutamina) constituyen un caso extremo. En los mutantes de lisozima T4 (Blaber et al., 1994) se observa, a partir de las estructuras cristalinas, una mejor solvatación de los grupos carbonilo y amida del esqueleto peptídico como consecuencia de la introducción de restos de Ala en lugar de otros más voluminosos; por lo tanto, las cadenas laterales adyacentes a la posición de mutación podrían afectar a la solvatación del esqueleto peptídico y modular, de este modo, la propensión helicoidal observada para el aminoácido sustituido.

Sin embargo, recientemente se ha medido el juego completo de propensiones helicoidales en un péptido de características muy similares a los péptidos basados en Ala (mostrado en la tercera columna de la tabla II.A.1) y en el que las cadenas laterales en posiciones de interés presentan un volumen similar, habiéndose encontrado un valor de  $\Delta\Delta G$  para Gly de  $1.05 \text{ kcal}\cdot\text{mol}^{-1}$  (Yang et al., 1997). Estos resultados parecen indicar que la hipótesis expuesta anteriormente no puede aplicarse de modo universal, por lo que, hasta el momento, no se dispone de una explicación razonable que justifique las diferencias en los valores de  $\Delta\Delta G$  para los péptidos AK/AQ y KEAEK y el resto de los sistemas experimentales.

#### II.A.4.2.-Comparación entre los resultados de análisis estadísticos y estudios experimentales.

Es interesante, asimismo, comparar las propensiones helicoidales obtenidas

experimentalmente con la frecuencia con que un determinado resto aparece en las posiciones centrales de hélices de proteínas de estructura conocida. En la tabla II.A.2 se recogen dos de las escalas obtenidas mediante el análisis estadístico más empleadas: la desarrollada por Chou y Fasman (Chou & Fasman, 1978) a partir de la observación de 29 estructuras y la de Williams y colaboradores (Williams et al., 1987), producto del estudio de un grupo de 75 proteínas.

En general sólo se observa una correlación moderada entre las propensiones helicoidales experimentales y ambas escalas, de modo que se pueden observar casos en los que aminoácidos

*Tabla II.A.2: Otras escalas de propensión helicoidal  
 $\Delta\Delta G$  (kcal·mol<sup>-1</sup>)*

Aminoácido	AGADIR <sup>(1)</sup>	Williams <sup>(2)</sup>	Chou-Fasman <sup>(3)</sup>	Pace <sup>(4)</sup>
Ala	-1.10	-0.98	-0.85	-1.00
Arg	-1.04	-0.78	-0.41	-0.79
Asn	-0.50	-0.33	-0.10	-0.35
Asp	-0.51	-0.56	-0.44	-0.31
Cys	-0.50	-0.23	-0.13	-0.32
Gln	-0.78	-0.84	-0.54	-0.61
Glu	-0.76	-1.16	-0.94	-0.60
Gly	0.00	0.00	0.00	0.00
His	-0.48	-0.62	-0.43	-0.39
Ile	-0.75	-0.66	-0.51	-0.59
Leu	-0.91	-0.91	-0.64	-0.79
Lys	-0.95	-0.80	-0.59	-0.74
Met	-0.89	-0.87	-0.88	-0.76
Phe	-0.63	-0.73	-0.56	-0.46
Ser	-0.58	-0.14	-0.20	-0.50
Thr	-0.53	-0.33	-0.26	-0.34
Trp	-0.63	-0.59	-0.51	-0.51
Tyr	-0.63	-0.31	-0.12	-0.47
Val	-0.64	-0.55	-0.49	-0.39

(1) Escala semiempírica basada en el análisis termodinámico estadístico de 423 secuencias peptídicas: (Muñoz & Serrano, 1995b)

(2) y (3) Escalas basadas en el análisis estadístico de las base de datos estructurales de proteínas: (Williams et al., 1987) y (Chou & Fasman, 1978), respectivamente.

(4) Escala promedio basadas en las distintas escalas experimentales de propensiones helicoidales: (Pace & Scholtz, 1997)

de carácter fuertemente desestabilizador aparecen en las posiciones centrales de las hélices de proteínas con una frecuencia sólo cuatro veces menor que alanina (una vez corregidos los datos por la frecuencia relativa de cada aminoácido) sólo cuatro veces menor que alanina (Chakrabarty et al., 1994). Esta observación no es sorprendente si se considera que Chou y Fasman incluyen en su estudio todos los restos en conformación helicoidal, sin evaluar su posición en la hélice o su situación en el interior de la proteína. Estos resultados indican, por lo tanto, que las propensiones helicoidales intrínsecas únicamente representan una fracción de las preferencias conformacionales de cada aminoácido y que existen otros determinantes adicionales como puede ser la presencia de



interacciones terciarias que pueden ser de importancia.

La situación es distinta con respecto a las propensiones N-cap (Doig & Baldwin, 1995) para las que existe una muy buena correlación con la frecuencia con que cada aminoácido aparece en dicha posición (Presta & Rose, 1988; Richardson & Richardson, 1988). Este fenómeno puede ser consecuencia de la disposición de los extremos de las hélices en la estructura tridimensional, frecuentemente situados en zonas cercanas a la superficie y accesibles al disolvente, que reduce significativamente su probabilidad de participación en interacciones terciarias.

#### **II.A.4.3.-Comparación de las propensiones helicoidales en péptidos y proteínas.**

Una cuestión importante para la evaluación del papel que las propensiones helicoidales de los aminoácidos tienen en la determinación de la estabilidad y mecanismo de plegamiento de una proteína es su aplicabilidad tanto a péptidos como a proteínas. De cualquier modo, dado que las medidas de energía libre en los distintos sistemas se han realizado bajo diferentes condiciones de temperatura, fuerza iónica y presencia de agentes desnaturalizantes, no es posible esperar una correspondencia exacta entre ellos, aunque es interesante investigar la posibilidad de establecer una escala de propensiones helicoidales de aplicabilidad general para cualquier hélice peptídica.

La existencia de correlaciones positivas entre los valores obtenidos con sistemas peptídicos, proteínas y coiled-coils indica que las propensiones helicoidales contribuyen de modo similar a la estabilidad conformacional, tanto en péptidos como en proteínas. La correlación entre los datos correspondientes a la proteínas entre sí es menor que la encontrada entre los sistemas peptídicos, lo que refleja la mayor influencia de factores externos, difícilmente cuantificables, en la medida de las propensiones helicoidales en éstas últimas. Es posible, por tanto, la existencia de cierta dependencia de las medidas de propensión helicoidal del contexto en que se determinan, que resulta de muy difícil corrección (Lin et al., 1994; Pinker et al., 1993).

Como se ha mencionado, es interesante determinar si la propensión helicoidal de un aminoácido proporciona una contribución energética equivalente en un péptido aislado y en una proteína. Con este propósito, Myers y colaboradores han medido las propensiones helicoidales de los aminoácidos naturales en una de las hélices de Ribonucleasa T1 y en un péptido modelo de secuencia idéntica (los resultados se muestran en la tabla II.A.1). Los autores observan que, en este caso, las propensiones helicoidales de la mayoría de los aminoácidos son muy similares en ambos sistemas, siendo la correspondencia especialmente notable entre las cadenas laterales no polares. Es conveniente hacer notar que el análisis de los datos se ha llevado a cabo utilizando modelos termodinámico-estadísticos multiestado para los datos correspondientes al sistema peptídico y la aproximación de dos estados para la proteína, por lo que la buena correspondencia observada es indicativa asimismo de la bondad de los modelos de transición hélice-ovillo estadístico para la descripción de la formación de hélices peptídicas. De igual modo, los resultados obtenidos constituyen una evidencia clara de que, en ciertas ocasiones, las propensiones helicoidales determinadas en sistemas peptídicos son aplicables directamente a proteínas.

#### **II.A.4.4.-Escala de propensiones helicoidales consenso.**



Recientemente se ha propuesto una escala de propensión helicoidal consenso (mostrada en la última columna de la tabla II.A.2) que incluye los once juegos de datos experimentales recogidos en la tabla II.A.1 (Blaber et al., 1994; Horovitz et al., 1992; Muñoz & Serrano, 1995b; Myers et al., 1997b; O'Neil & DeGrado, 1990; Park et al., 1993b; Rohl et al., 1996; Yang et al., 1997), de modo que considera datos procedentes tanto de sistemas peptídicos como de proteínas. De acuerdo con los autores (Pace & Scholtz, 1997), esta escala general presenta una excelente correspondencia con los resultados obtenidos por Muñoz y Serrano mediante su algoritmo AGADIR (Muñoz & Serrano, 1995b) así como con la desarrollada en este trabajo y presentada en la sección IV.A (Luque et al., 1996). La correlación encontrada con las escalas resultantes de los análisis estadísticos de la base de datos estructurales (Chou & Fasman, 1978; Richardson & Richardson, 1988; Williams et al., 1987) es sólo moderada. La tabla II.A.3 se recogen los resultados de las distintas correlaciones.

*Tabla II.A.3: Correlación entre diversas escalas de propensión helicoidal y la escala experimental promedio desarrollada por Pace y colaboradores (Pace & Scholtz, 1997)*

Para aminoácidos no polares (Gly, Ala, Val, Ile, Leu, Met, Phe):

	AK/AQ <sup>1</sup>	AGADIR <sup>2</sup>	WILLIAMS <sup>3</sup>	C. y F. <sup>4</sup>	Luque <sup>5</sup>
<b>Coefficiente de Correlación:</b>	0.99	0.98	0.96	0.93	0.94
<b>Pendiente:</b>	1.91	1.05	0.97	0.84	0.74
<b>Ordenada en el origen:</b>	-0.02	-0.06	-0.11	-0.07	0.01

Para todos los aminoácidos:

	AK/AQ	AGADIR	WILLIAMS	C. y F.	Luque
<b>Coefficiente de Correlación:</b>	0.94	0.97	0.78	0.74	0.93
<b>Pendiente:</b>	1.75	1.06	1.05	0.86	0.81
<b>Ordenada en el origen:</b>	-0.07	-0.08	-0.12	-0.01	0.00

<sup>1</sup>(Rohl et al., 1996); <sup>2</sup>(Muñoz & Serrano, 1995b); <sup>3</sup>(Williams et al., 1987); <sup>4</sup>(Chou & Fasman, 1978); <sup>5</sup>(Luque et al., 1996)

De acuerdo con esta escala general, Alanina es el aminoácido con mayor tendencia a encontrarse en conformación helicoidal, de modo que, en términos absolutos, es el único que contribuye favorablemente a la estabilidad de la hélice (Rohl et al., 1996). A continuación se encuentran aquellos aminoácidos que poseen cadenas laterales largas y de carácter hidrofóbico como Met o Leu, junto con los que presentan grupos polares separados del esqueleto por varios átomos de carbono como Arg, Lys<sup>+</sup>, Gln o Glu<sup>-</sup>. En este último caso, es interesante hacer notar la desestabilización introducida por el grupo cargado de los restos Glu<sup>-</sup> y Asp<sup>-</sup>, que probablemente refleja la disminución de su propensión helicoidal debido a la capacidad de sus cadenas laterales para establecer enlaces de hidrógeno con los grupos amida del esqueleto peptídico. Tanto Gln como Glu<sup>-</sup> son mejores formadores de hélices que Asn y Asp<sup>-</sup>, debido posiblemente a que, a

consecuencia de la menor longitud de estos últimos, el costo entrópico que supone el establecimiento de interacciones específicas entre sus cadenas laterales y el esqueleto polipeptídico, que desestabilizarían la hélice, es menor.

La diferencia en entropía conformacional entre las cadenas laterales de los aminoácidos alifáticos en la hélice es la responsable de la mayor tendencia helicoidal de Leu frente Ile o Val, aunque no existen motivos evidentes que justifiquen la mayor propensión de Ile frente a Val. Los tres aminoácidos aromáticos presentan tendencias helicoidales muy similares, lo que sugiere que la presencia de grupos con capacidad para establecer enlaces de hidrógeno en las cadenas laterales de Trp y Tyr no tiene un efecto notable en sus correspondientes propensiones helicoidales. Todos ellos ocupan posiciones intermedias en la escala, equidistantes entre Ala y Gly.

Los peores formadores de hélices son aquellos restos que poseen cadenas laterales ramificadas en el carbono beta o bien presentan un grupo polar separado del esqueleto únicamente por un grupo  $-\text{CH}_2-$ . Prolina es claramente el resto de mayor potencial desestabilizador cuando se encuentra en posiciones centrales de la hélice debido a la geometría característica de su anillo que le impide adoptar la conformación helicoidal y a la carencia de un hidrógeno amida que permita el establecimiento de enlaces de hidrógeno con el esqueleto peptídico.

#### II.A.4.5.- Interacciones adicionales de carácter estabilizador.

##### II.A.4.5.1.-Interacciones *N-cap* y *C-cap*.

La longitud promedio de las hélices alfa encontradas en proteínas globulares es de unos doce restos, de modo que, en este caso, se establecen ocho enlaces de hidrógeno entre grupos amina y carbonilo de las posiciones internas de la hélice mientras que los grupos amina donores de los primeros cuatro restos, así como los carbonilos aceptores de los cuatro últimos se encuentran descompensados. Por tanto, de acuerdo con la hipótesis propuesta por Presta y Rose (Presta & Rose, 1988), la presencia en las posiciones adyacentes a los extremos de la hélice de aminoácidos, cuyas cadenas laterales posean la capacidad satisfacer los grupos donores y aceptores descompensados, es de gran importancia para la determinación de la estabilidad de la estructura helicoidal.

Los restos *N-cap* y *C-cap* pueden definirse como aquellos que ocupan posiciones adyacentes a la hélice y que, aunque participan en su red de enlaces de hidrógeno, no se encuentran en conformación helicoidal (Presta & Rose, 1988); o bien como el primer y último resto en la secuencia cuyo carbono alfa se encuentra incluido en el cilindro definido por la hélice (Richardson & Richardson, 1988). En proteínas, la determinación de los extremos de la hélice no presenta demasiados problemas y ambas definiciones conducen a resultados muy similares. Por el contrario, cuando se trabaja con péptidos pequeños, caracterizados por la existencia de numerosas conformaciones parcialmente helicoidales que se interconvierten entre sí, la demarcación de los límites de las hélices no es tan sencilla, ya que está definida por la distribución conformacional, a su vez determinada por la secuencia aminoácídica y las propensiones helicoidales de cada resto.

El fenómeno del “capping” de la hélice consiste en el establecimiento de interacciones

de corto alcance entre las cadenas laterales de los aminoácidos en posición N-cap o C-cap y el esqueleto peptídico. Asimismo, son importantes las interacciones que tanto estos grupos como el esqueleto peptídico de los restos vecinos establecen con el disolvente. La tendencia de un aminoácido cualquiera a ocupar estas posiciones en la hélice es una propiedad intrínseca de su cadena lateral, y es independiente de su propensión helicoidal. El estudio de este fenómeno en péptidos modelo requiere la modificación de las teorías originales de transición helice-ovillo estadístico con objeto de incluir los restos N-cap y C-cap, inicialmente no considerados, en la función en partición mediante la asignación de sus correspondientes pesos estadísticos (ver figura II.A.2).

Diversos estudios experimentales demuestran que las interacciones de "capping" estabilizan tanto las hélices peptídicas (Bruch et al., 1991; Forood et al., 1993; Lyu et al., 1993) como las encontradas en proteínas (Bell et al., 1992; Serrano & Fersht, 1989; Serrano et al., 1992a; Serrano et al., 1992b). La tendencia de los distintos aminoácidos a encontrarse en la posición N-cap en los péptidos ricos en alanina de la serie AK/AQ se ha determinado mediante la aplicación de la teoría de Lifson y Roig modificada (Doig et al., 1994) al análisis de los resultados experimentales (Chakrabartty et al., 1994; Doig et al., 1994). Asimismo se han determinado la propensiones relativas a alanina de varios aminoácidos en proteínas como Barnasa (Serrano et al., 1992b) y Lisozima T4 (Bell et al., 1992), mostrados en la tabla II.A.4.

**Tabla II.A.4:** Cambios en energía libre asociados a la formación de interacciones N-cap  
 $\Delta\Delta G(\text{kcal}\cdot\text{mol}^{-1})$

Aminoácido	AK/AQ <sup>(1)</sup>	Lisozima T4 <sup>(2)</sup>	Barnasa <sup>(3)</sup>
Ala	0.00	0.00	0.00
Asn	-1.41	-2.20	-0.86
Asp <sup>o</sup>	-----	-----	-2.02
Asp <sup>-</sup>	-----	-1.90	-----
Gln	0.93	-----	-0.42
Glu <sup>-</sup>	-----	-----	-0.25
Gly	-1.08	-0.60	-0.69
Ile	-0.58	-----	-0.16
Leu	-0.71	-----	-----
Lys <sup>+</sup>	0.50	-----	-----
Met	-0.39	-----	-----
Pro	-0.33	-----	0.87
Ser	-1.12	-1.90	-1.64
Thr	-0.64	-2.80	-2.05
Val	-0.10	0.00	0.15

(1) Valores obtenidos mediante la aplicación de la teoría de Lifson y Roig modificada (Doig et al., 1994) al análisis de los péptidos ricos en Alanina de la serie AK/AQ (Chakrabartty et al., 1994).

(2) Valores obtenidos de la substitución de varios aminoácidos en posición 59 de Lisozima T4 a pH 2.0 (Bell et al., 1992).

(3) Valores obtenidos de la substitución de varios aminoácidos en las posiciones 6 y 26 de Barnasa (Serrano et al., 1992a; Serrano et al., 1992b).

De los resultados de estos estudios se desprende la existencia de una amplia dispersión en la tendencia de los aminoácidos no cargados a ocupar la posición N-cap, de modo que, en general, estas tendencias son diferentes de sus propensiones helicoidales (Chakrabarty et al., 1993a; Lyu et al., 1993; Serrano et al., 1992b). La magnitud de los cambios en energía libre asociados a estas interacciones son similares a los correspondientes a la formación de la hélice (Chakrabarty et al., 1994), de tal forma que la sola substitución de un aminoácido en el extremo amino-terminal es suficiente para producir un profundo efecto en la estructura secundaria de péptidos aislados (Forood et al., 1993).

Por el contrario, los efectos de las substituciones en la posición C-cap de restos neutros y no aromáticos son muy pequeños. De este modo, con la excepción de Gly, muy frecuentemente encontrada en esta posición, las diferencias entre las propensiones C-cap de los distintos aminoácidos en hélices peptídicas son muy pequeñas. Es posible, por tanto, que el establecimiento diferencial de estas interacciones sea el responsable de la ligera asimetría en contenido helicoidal observada en pequeños péptidos de secuencias simétricas, justificando la mayor estabilidad observada en el extremo amino-terminal en comparación con el carboxilo-terminal (Chakrabarty et al., 1993a).

Al igual que ocurre con las propensiones helicoidales, la tendencia de cada aminoácido a ocupar estas posiciones puede racionalizarse de acuerdo con ciertos determinantes físicos, entre los que cabe destacar el establecimiento de enlaces de hidrógeno entre las cadenas laterales y el esqueleto peptídico, la solvatación de los grupos amida descompensados, o la protección de superficies apolares del contacto con el disolvente.

De este modo, los restos que poseen cadenas laterales polares y de pequeño tamaño (Asn, Asp, Thr ó Ser) presentan las mayores propensiones a ocupar estas posiciones debido a la posibilidad de establecer un enlace de hidrógeno entre sus cadenas laterales y el esqueleto peptídico (Presta & Rose, 1988; Richardson & Richardson, 1988). Por otro lado, la elevada propensión N-cap de Gly se debe posiblemente a la ausencia de cadena lateral, lo que posibilita una mejor solvatación de los grupos amida descompensados de la primera vuelta de la hélice (Serrano et al., 1992a). Los aminoácidos no polares poseen bajas o moderadas propensiones N-cap, que se correlacionan con tamaño de sus cadenas laterales, posiblemente a consecuencia de la deshidratación parcial, como consecuencia de la formación de la hélice, de los grupos hidrofóbicos que las constituyen. La extremadamente baja propensión N-cap de Gln se ha atribuido a su capacidad de formación de enlaces de hidrógeno con el esqueleto peptídico en el estado desplegado (Chakrabarty et al., 1993a).

La mayor tendencia de Gly en el extremo carboxilo terminal refleja su capacidad única para adoptar conformaciones caracterizadas por ángulos diedros positivos, de modo que favorece la solvatación de los grupos carbonilo del esqueleto y permite el establecimiento simultáneo de dos puentes de hidrógeno, imposibilitando, de este modo, la propagación de la hélice.

#### II.A.4.5.2.-Interacciones entre grupos cargados y el dipolo de la hélice.

Como se ha descrito anteriormente, cualquier hélice alfa se asemeja a un macrodipolo con un polo positivo (de carga aproximadamente igual a +0.5 en el extremo N-terminal) originado por las cargas parciales de los cuatro grupos NH no implicados en la formación de enlace de hidrógeno, y otro negativo (de carga -0.5) asociado a los grupos CO de los últimos cuatro restos del extremo C-terminal (Hol et al., 1978; Sheridan et al., 1982; Wada, 1976). Como consecuencia, la presencia en la hélice de cadenas laterales con grupos cargados da lugar a la aparición de interacciones de tipo electrostático entre éstos y las cargas parciales de los grupos NH y CO del esqueleto peptídico (Åquist et al., 1991; Tidor & Karplus, 1991), denominadas interacciones "carga-dipolo de la hélice" o "carga-esqueleto".

La naturaleza exclusivamente electrostática, sin participación de enlaces de hidrógeno, de este tipo de interacciones se ha demostrado a través de numerosos estudios experimentales. Así, los resultados obtenidos de los experimentos llevados a cabo con copolímeros constituidos por dos bloques, uno de aminoácidos cargados y otro de restos de alanina [(Glu)<sub>20</sub>(Ala)<sub>20</sub> y (Lys)<sub>20</sub>(Ala)<sub>20</sub>] no pueden interpretarse mediante la formación de enlaces de hidrógeno. Por el contrario, es necesario recurrir al modelo de interacción carga-dipolo para justificar los resultados. Asimismo, el análisis de los efectos que diversos factores, como la situación de los bloques entre sí o la variación de la temperatura de transición (T<sub>m</sub>) con la concentración de sales, tienen en la estabilidad demuestran la naturaleza electrostática de las interacciones (Ihara et al., 1982; Takahashi et al., 1989). La dependencia con el pH de la T<sub>m</sub>, observada para diversos mutantes de Lisozima T4 en los que se ha introducido un grupo cargado en las proximidades del extremo N-terminal, constituye una prueba adicional de la existencia de dichas interacciones electrostáticas.

De igual modo, los efectos desestabilizadores observados como consecuencia de la introducción de restos cargados en el extremo de la hélice con carga de igual signo (Ihara et al., 1982), así como los efectos de apantallamiento a consecuencia de la concentración de sales en disolución (Huyghues-Despointes et al., 1993; Ihara et al., 1982; Lockhart & Kim, 1993; Shoemaker et al., 1987), únicamente pueden ser interpretados en términos electrostáticos.

Entre otros estudios dirigidos a elucidar la naturaleza y características de estas interacciones cabe destacar los llevados a cabo con el péptido S de Ribonucleasa con objeto de observar la variación de su carácter helicoidal con la carga en el extremo N-terminal (Mitchinson & Baldwin, 1986), o el análisis de los efectos que la introducción de restos cargados en diversas posiciones de una hélice peptídica de carácter neutro tiene en su estabilidad (Armstrong & Baldwin, 1993; Huyghues-Despointes et al., 1993; Scholtz et al., 1993). El ajuste de los datos experimentales en este último caso se ha llevado a cabo utilizando una modificación de la teoría de Lifson y Roig, que incorpora los efectos de la interacción entre un grupo cargado en cualquier posición de la hélice y su dipolo.



### II.A.4.5.3.-Interacciones específicas entre cadenas laterales.

La comprensión de la formación y estabilidad de las hélices alfa requiere la cuantificación de todas las interacciones existentes en estos elementos de estructura secundaria. Junto a interacciones entre los grupos cargados de las cadenas laterales y el dipolo de la hélice, el establecimiento de interacciones específicas entre cadenas laterales constituye uno de los principales factores estabilizadores de las hélices peptídicas en disolución (Bierzynski et al., 1982; Shoemaker et al., 1985; Shoemaker et al., 1987).

En este contexto, diversos estudios experimentales llevados a cabo con pequeños péptidos helicodales aislados de proteínas (como el péptido C de Ribonucleasa A) demostraron la existencia de interacciones específicas entre diversas cadenas laterales. Entre ellas destacan los puentes salinos establecidos entre grupos cargados (Fairman et al., 1990; Osterhout et al., 1989) y las interacciones entre grupos amino y cadenas laterales aromáticas (Shoemaker et al., 1990), esenciales para la estabilidad estructural de las hélices aisladas en disolución.

Posteriormente diversos estudios estructurales han permitido la identificación en proteínas completas de interacciones similares a las encontradas en los péptidos aislados, lo cual corrobora el papel de estos elementos de estructura secundaria como unidades cooperativas independientes, de gran importancia en la determinación del mecanismo de plegamiento.

#### a.-Motivos estructurales característicos de los extremos de la hélice.

Las interacciones entre las cadenas laterales en posición  $(i, i+3)$  y  $(i, i+4)$ , se habían identificado inicialmente, junto con las propensiones helicoidales intrínsecas de los distintos aminoácidos, como los principales determinantes de la estabilidad de las hélices alfa. Sin embargo, más recientemente, el análisis estadístico de la base de datos de proteínas ha conducido a la localización, en los extremos amino y carboxilo terminal de las hélices, de motivos estructurales específicos de carácter local importantes para la estabilización y localización de las hélices en la secuencia (Aurora et al., 1994; Harper & Rose, 1993; Muñoz et al., 1995b; Seale et al., 1994). Entre los motivos estructurales de mayor interés cabe destacar los denominados "capping box" e "hydrophobic staple" en el extremo amino-terminal así como el motivo de Schellman en el carboxilo-terminal.

##### a.1.-"The capping box".

Este motivo específico, característico del extremo amino-terminal, consiste en el establecimiento de un enlace de hidrógeno recíproco entre los restos en posición N-cap y N3 (Harper & Rose, 1993) (de acuerdo con la notación descrita en la sección II.A.3.1.2). De este modo, la cadena lateral del resto N-cap se encuentra enlazada con el grupo amida en posición N3 y, recíprocamente, la cadena lateral de este último participa en un enlace de hidrógeno con el



correspondiente nitrógeno amida del resto en posición N-cap. Por lo tanto, este motivo estructural permite la compensación de dos de los cuatro grupos donores libres de la primera vuelta de la hélice, y podría actuar como la señal que indique el final de la hélice tanto en péptidos como en proteínas (Zhou et al., 1993).

#### a.2.-“The hydrophobic staple”.

El segundo motivo estructural de importancia, localizado asimismo en el extremo amino-terminal, es el resultado de la interacción hidrofóbica establecida entre las cadenas laterales de los restos en posición N' y N4, asociada con la presencia de un resto con buenas propiedades N-capping como Ser, Thr, Asp o Asn (Muñoz et al., 1995b).

El análisis de la base de datos estructurales de proteínas (Muñoz & Serrano, 1995b) muestra que este motivo estructural se encuentra con frecuencia acompañado de estructuras de “capping box”, de modo que la presencia de ambos motivos ejerce un efecto sinérgico en la estabilidad de la hélice. El efecto estabilizador de este motivo estructural radica fundamentalmente en el establecimiento de interacciones de tipo hidrofóbico entre dos restos voluminosos situados en posiciones ( $i, i+5$ ), aunque, cuando aparece asociado al motivo de “capping box”, la geometría establecida permite el establecimiento adicional de interacciones ( $i, i+8$ ) en la hélice (Muñoz & Serrano, 1995a).

Diversos estudios experimentales llevados a cabo con péptidos monoméricos, diseñados específicamente para el análisis de este tipo de interacciones, han demostrado la contribución significativa de ambos motivos estructurales a la estabilidad de la hélice en disolución (Jiménez et al., 1994; Lyu et al., 1993; Muñoz & Serrano, 1995b).

#### a.3.-El motivo estructural de Schellmann.

Este motivo, característico del extremo carboxilo terminal de la hélice, consiste en la presencia de un resto en posición C' (generalmente Gly) con ángulos diedros positivos acompañado por el establecimiento de dos enlaces de hidrógeno entre los grupos del esqueleto peptídico: uno entre los restos en posición C3 y C'' y el otro entre los situados en C2 y C' (Aurora et al., 1994).

El análisis estadístico de la base de datos de proteínas ha revelado la existencia de un patrón característico (Hidrofóbico-X-Polar-X-Glicina-Hidrofóbico') en la secuencia asociado a la formación de este motivo estructural (Aurora et al., 1994; Viguera & Serrano, 1995b), consistente en la presencia de Gly en posición  $i$ , junto con restos hidrofóbicos en las posiciones ( $i-4$ ) e ( $i+1$ ) y polares (o bien Ala) en posición ( $i-2$ ). De este modo, la interacción hidrofóbica entre las cadenas laterales, junto con el establecimiento de los dos enlaces de hidrógeno, facilitan la adopción del motivo estructural de Schellmann en este tipo de secuencias.

Al contrario que las interacciones descritas anteriormente, este motivo estructural

contribuye sólo ligeramente a la estabilidad helicoidal. El estudio experimental del papel que estas secuencias juegan en la formación del motivo de Schellman y, por tanto, en la estabilización de la hélice, llevado a cabo con pequeños péptidos monoméricos, ha demostrado la inestabilidad de este tipo de disposiciones estructurales en disolución acuosa. Esto se debe probablemente al alto costo entrópico asociado con la inmovilización del resto de Gly, que no se encuentra compensado por las interacciones hidrofóbicas entre las cadenas laterales (Viguera & Serrano, 1995b). Los autores apuntan hacia importancia de los efectos de contexto en los análisis estadísticos llevados a cabo con proteínas, y que están ausentes en el caso de péptidos monoméricos en disolución acuosa. En proteínas, la formación del motivo estructural de Schellman puede resultar un modo cómodo de proteger amplias superficies hidrofóbicas del contacto con el disolvente y satisfacer, simultáneamente, el potencial formador de enlaces de hidrógeno del extremo carboxilo terminal de la hélice, que, de otro modo, necesitaría del establecimiento de interacciones de más largo alcance.

*b.-Otras interacciones estabilizadoras entre cadenas laterales.*

La existencia de interacciones entre las cadenas laterales de los distintos restos que componen una hélice alfa ha demostrado ser una importante fuente de estabilización, de modo que, en ocasiones, el cambio en energía libre que conlleva su establecimiento es de magnitud similar al asociado con la formación interacciones fuertemente estabilizadoras como puentes salinos. Las energías de estabilización asociadas a algunas de estas interacciones se han determinado cuantitativamente mediante el estudio de sistemas simplificados como pequeños péptidos ricos en Ala (Chakrabarty & Baldwin, 1995).

Como se ha comentado anteriormente, la elevada propensión helicoidal de este aminoácido permite obtener hélices peptídicas de 15 o 20 restos estables en disolución, que permiten controlar, en cierto modo, la naturaleza de las interacciones establecidas entre sus cadenas laterales. De este modo, mediante la inserción de cadenas laterales de naturaleza variada en distintas posiciones en la hélice, es posible cuantificar la estabilización adicional aportada por cada tipo de interacción, así como su dependencia con las posiciones relativas ocupadas por las distintas cadenas laterales. En condiciones óptimas, cuando el péptido de referencia posee aproximadamente un 50% de contenido helicoidal, situación en la que una pequeña variación en energía libre origina un cambio significativo en la helicidad, es posible determinar con suficiente precisión interacciones del orden de  $-0.5$  kcal/mol (Scholtz et al., 1993). Al igual que en los casos anteriores, el análisis cuantitativo de los datos experimentales requiere el desarrollo de un modelo termodinámico estadístico extendido que contemple la posibilidad de interacciones entre cadenas laterales, no consideradas en los tratamientos originales (Scholtz et al., 1993; Shalongo & Stellwagen, 1995).

Entre las interacciones más frecuentemente observadas en péptidos y proteínas es

necesario mencionar el establecimiento de puentes salinos entre las cadenas laterales de aminoácidos cargados (Fairman et al., 1990; Gans et al., 1991; Huyghues-Despointes et al., 1993; Lyu et al., 1990; Marqusee & Baldwin, 1987; Merutka et al., 1990; Scholtz et al., 1993), las interacciones hidrofóbicas entre cadenas laterales apolares (Creamer & Rose, 1995; Padmanabhan & Baldwin, 1994), el establecimiento de enlaces de hidrógeno entre dos cadenas laterales (Scholtz et al., 1993), las interacciones entre cadenas aromáticas y aminas cargadas (Armstrong & Baldwin, 1993; Shoemaker et al., 1990), o las interacciones entre restos azufrados y aromáticos (Reid et al., 1985; Viguera & Serrano, 1995a).

## **II.B.-PROTEASAS ASPARTICAS.**

### **II.B.1.-Enzimas proteolíticas.**

Las enzimas proteolíticas desempeñan una función de gran importancia en el ciclo de vida de las proteínas, dado que la ruptura de enlaces peptídicos constituye una de sus modificaciones más frecuentes. Tradicionalmente, la proteólisis enzimática se ha asociado con procesos digestivos, habiendo sido las proteasas presentes en el tracto digestivo de los mamíferos el principal foco de atención. Sin embargo, recientemente el interés se ha dirigido hacia proteasas que desempeñan funciones regulatorias en gran variedad de procesos fisiológicos de enorme relevancia. Así, las enzimas proteolíticas intervienen en la activación de zimógenos, hormonas peptídicas y receptores, en el procesamiento de cadenas polipeptídicas en el momento de su síntesis, en la activación de proteasas reguladoras implicadas en procesos de coagulación sanguínea, así como en diversos procesos celulares desarrollados en gran variedad de tejidos.

En términos generales, el mecanismo mediante el cual se lleva a cabo la ruptura de un enlace peptídico (mostrado en la figura II.B.2 para el caso específico de las proteasas aspárticas) es común a todas las proteasas. De este modo, durante la etapa inicial del proceso de hidrólisis, el carbono carbonílico del enlace peptídico, cuyo carácter electrofílico se encuentra exaltado debido a la acción de un grupo electrofílico sobre el oxígeno carbonílico que incrementa la polarización del enlace, sufre el ataque de un nucleófilo adecuado. Dicho ataque nucleofílico se encuentra adicionalmente facilitado por la eliminación de un protón del nucleófilo, proceso que generalmente se encuentra asistido por la intervención de algún grupo de la enzima. La posterior ruptura del intermedio tetraédrico generado como consecuencia de la adición del nucleófilo requiere, asimismo, la asistencia de algún tipo de catálisis ácida general con objeto de facilitar la salida del grupo amino terminal. La naturaleza de los grupos específicos que realizan dichas funciones (ataque nucleofílico, catálisis ácido-base y asistencia electrofílica) es, sin embargo, característica de cada tipo de proteasa.

Las proteasas han sido clasificadas según diversos criterios como tamaño molecular, carga eléctrica o especificidad por diversos sustratos. Actualmente, de acuerdo con un sistema de clasificación más racional basado en la comparación de sus sitios activos, mecanismos de acción y estructuras tridimensionales, la Unión Internacional de Bioquímica reconoce cuatro clases de proteasas, agrupadas en seis familias (proteasas de serina I y II, tioproteasas, proteasas aspárticas y metaloproteasas I y II), cada una de las cuales presenta un juego característico de aminoácidos funcionales que, dispuestos en una configuración específica, conforman su sitio activo.

### **II.B.2.-Proteasas Aspárticas.**

#### **3.B.2.1.-Características Generales.**

Las proteasas aspárticas se encuentran ampliamente representadas en la naturaleza, de modo que están presentes en muy diversos tipos de sistemas, desde vertebrados, plantas y hongos hasta retrovirus. Al igual que ocurre con otros grupos de enzimas, los miembros de esta familia desempeñan funciones muy distintas, dadas sus diferencias en actividad, especificidad y estabilidad. Así, las enzimas gástricas como Pepsina, Gastricsina o Quimosina, participan en los procesos digestivos. Cathepsina E y D, localizadas en los lisosomas de varios tipos celulares, intervienen en la degradación intracelular de proteínas y se encuentran posiblemente implicadas en el desarrollo de procesos cancerígenos. Otras enzimas, como Renina, participan indirectamente en el control de la presión sanguínea mediante la hidrolización altamente específica de angiotensinógeno, dando origen al decapeptido precursor de angiotensina II. Varias de las proteínas del grupo participan en los procesos de esporulación de hongos, como es el caso de Endothiapepsina, Penicillopepsina o Mucorpepsina. Finalmente, las proteasas retrovíricas constituyen una pieza clave en el procesado de las distintas proteínas víricas, sintetizadas originalmente como precursores poliprotéicos inactivos, por lo que su intervención es esencial para la maduración del virus.

Existen ciertas características universales que identifican una enzima como proteasa aspártica (Davies, 1990; Tang & Wong, 1987). En general, todos los miembros del grupo presentan un pH de máxima actividad en el intervalo ácido (con la excepción de renina y las proteasas retrovíricas que operan a un pH cercano a la neutralidad), masas moleculares comprendidas entre 35 y 40 kDalton, así como especificidad por sustratos peptídicos en conformación extendida. Sin embargo, el rasgo más característico de esta familia de proteasas es la intervención de dos grupos aspárticos en el mecanismo catalítico, por lo que estas enzimas son susceptibles de ser inhibidas irreversiblemente por reactivos específicos del sitio de unión como el éster metílico de diazoacetil-norleucina (Rajagopalan et al., 1966) y el epóxido de p-nitro-fenoxi-propano (Tang, 1971), cada uno de los cuales reacciona selectivamente con el grupo carboxilo de uno de los aspárticos catalíticos (Asp<sup>215</sup> y Asp<sup>32</sup> respectivamente, en notación correspondiente a pepsina), conduciendo a la inactivación de la enzima. Asimismo, Pepstatina, pentapéptido de secuencia Iva-Val-Val-Sta-Ala-Sta, producido de forma natural por varias especies de actinomicetos (Umezawa, 1970), constituye un inhibidor de muy alta afinidad, universal a todas las proteasas aspárticas (Bott et al., 1982; Cooper et al., 1989; Rao & Singh, 1991; Rich et al., 1985), de forma que uno de los restos de Estatina, (ácido 4(S)-amino-3(S)-hidroxi-6-metil heptanoico), actúa como análogo del estado tetraédrico de transición (Bailey et al., 1993; Blundell et al., 1987; Goldblum, 1988; James et al., 1992; Rich et al., 1980). La estructura química de este inhibidor se muestra en la figura II.B.3.

Las proteasas celulares, encontradas en mamíferos, plantas y hongos, están constituidas por una única cadena polipeptídica de unos 327 aminoácidos de longitud, habiéndose encontrado un 5% de homología entre las secuencias de las diferentes enzimas. La mayoría de estas proteínas se sintetizan como zimógenos inactivos, los cuales contienen unos 45 aminoácidos adicionales en el

extremo amino-terminal que deben ser eliminados para dar origen a la forma activa de la enzima (Davies, 1990; Tang & Wong, 1987). Por el contrario, las proteasas retrovíricas son enzimas de menor tamaño, constituidas por unos 130 aminoácidos en total, y de naturaleza homodimérica, de modo que cada uno de los monómeros aporta la mitad de los elementos necesarios para conformar el sitio activo.

### II.B.2.2.-Estructura Tridimensional.

En la actualidad se dispone de las estructuras tridimensionales de un grupo significativo de estas proteínas, tanto de las proteasas del grupo de la Pepsina (pepsina porcina (Abad-Zapatero et al., 1990; Cooper et al., 1990), Pepsinógeno (James & Sielecki, 1986), Quimosina bovina (Newman et al., 1991), Renina humana (Sielecki et al., 1989), y las proteasas fúngicas Penicillopepsina (James & Sielecki, 1983), Rhizopuspepsina (Suguna et al., 1987a), Endothiapepsina (Blundell et al., 1990) y Mucorpepsina (Watson et al., 1988)) como de diversas enzimas retrovíricas (Rous sarcoma virus (Miller et al., 1989b) o V.I.H.-1 (Wlodawer et al., 1989)). En términos generales, las estructuras tridimensionales de todas las proteasas aspárticas son muy similares, lo que, en combinación con la homología encontrada entre sus secuencias, sugiere un origen evolutivo común para todas ellas.

La estructura secundaria, mostrada en la figura II.B.4 para el caso concreto de Endothiapepsina, está constituida fundamentalmente por láminas beta, con un muy pequeño porcentaje de conformación helicoidal. Estas proteínas presentan estructuras bilobulares, y están constituidas por dos dominios muy similares (o bien dos monómeros idénticos en el caso de las proteasas retrovíricas) que se encuentran relacionados espacialmente por un eje binario de rotación. Ambos dominios estructurales se hallan separados por una profunda hendidura de unos 40 Å de longitud, donde se encuentra localizado el sitio de unión de sustratos e inhibidores competitivos, constituyendo, por tanto, el centro catalítico de la molécula.

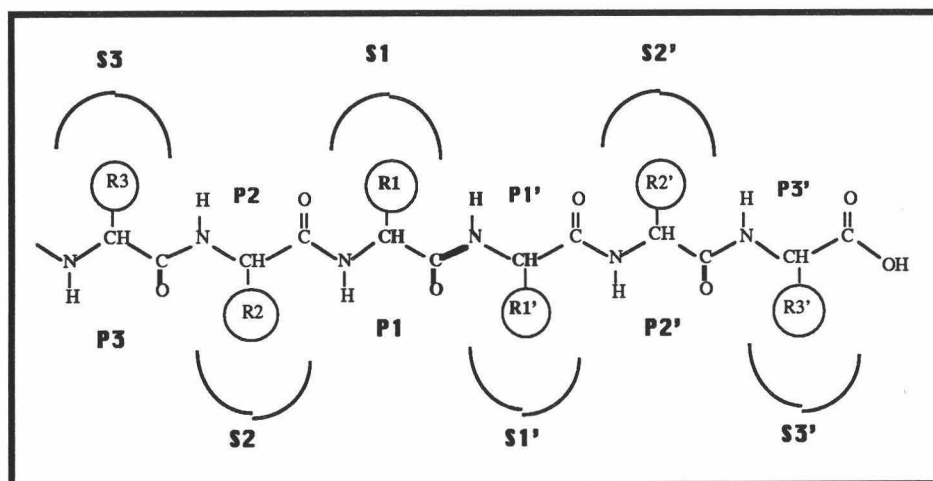
Todas las proteasas aspárticas, tanto las enzimas monoméricas del grupo de Pepsina como las proteasas retrovíricas, de naturaleza dimérica, presentan secuencias características, altamente conservadas, en las inmediaciones de los aspárticos catalíticos. El motivo *X-Asp-Thr-Gly-Ser*, donde X es un resto hidrofóbico (generalmente Phe), aparece tanto en el dominio amino como carboxilo terminal. Estos aminoácidos desempeñan un papel esencial no sólo en la definición y estabilización de la estructura del sitio activo a través de una intrincada y extremadamente simétrica red de puentes de hidrógeno, sino en el establecimiento de interacciones con inhibidores y sustratos.

El sitio activo puede acomodar hasta siete aminoácidos del ligando en los subsitios S4-S3' (según la notación establecida por Schechter y Berger (Schechter & Berger, 1967) y representada en la figura II.B.1) con preferencia para la hidrolización del enlace peptídico establecido entre los resíduos en posición S1 y S1'. Las variaciones en especificidad y actividad entre las diversas



proteasas aspárticas pueden racionalizarse como el resultado de alteraciones discretas en alguno de los subsitios, dado que son las interacciones establecidas en cada uno de ellos las responsables de situar el sustrato, de modo que el enlace peptídico hidrolizable se encuentre orientado adecuadamente entre los dos grupos aspárticos catalíticos, esenciales para el proceso de hidrólisis.

Un rasgo peculiar de este tipo de proteasas es la existencia de una región excepcionalmente flexible o "flap" que descansa sobre el sitio activo en la proteína libre y que debe abrirse para permitir el acceso del ligando. Una vez formado el complejo, esta aleta se cierra sobre el ligando constituyendo la parte superior del sitio de unión y forzando la casi completa deshidratación tanto del sustrato o inhibidor como del centro activo de la enzima. Las numerosas interacciones establecidas con el ligando estabilizan notablemente esta región con la consecuente pérdida de flexibilidad, que se refleja en la disminución de los factores de temperatura en el complejo con respecto a la proteína libre. Las proteasas retrovíricas presentan dos de estos lazos flexibles, cada uno aportado por uno de los monómeros, de tamaño algo mayor que los encontrados en muchas enzimas eucarióticas, lo cual se traduce en el incremento del grado de desolvatación del ligando como consecuencia de la unión a la proteína con respecto a las éstas.



*Figura II.B.1: Nomenclatura de las distintas posiciones del sustrato y de los subsitios de unión en la enzima con respecto a la posición del sitio de hidrólisis, situado entre las posiciones P1 y P1'.*

### II.B.2.3.-Estado de ionización del sitio activo.

El centro activo de estas proteasas está constituido por los grupos carboxílicos de los dos aspárticos catalíticos, cada uno de ellos aportado por uno de los dominios (o monómeros en el caso de las proteasas retrovíricas) de la proteína. Una complicada red de enlaces de hidrógeno en

la que participan restos de ambos dominios confiere rigidez y estabilidad al sitio activo, de modo que los dos carboxilos se encuentran forzados en sus posiciones coplanares, manteniendo una molécula de agua coordinada entre ellos. Dada la pequeña distancia observada entre sus oxígenos en la estructura tridimensional, se ha postulado la existencia de un átomo de hidrógeno compartido por ambos grupos carboxílicos, así como el establecimiento de diversos enlaces de hidrógeno con la molécula de agua.

La disposición simétrica de ambos grupos aspárticos, la existencia de la molécula de agua entre ellos, así como la conservación de los aminoácidos en sus inmediaciones sugieren una estructura promedio resonante entre dos posibles conformaciones en las que el átomo de hidrógeno ocupa posiciones cercanas al OD2 de los restos Asp<sup>32</sup> y Asp<sup>215</sup> (numeración correspondiente a pepsina porcina) con igual probabilidad. Sin embargo, diversos estudios teóricos apuntan hacia la existencia de cierta asimetría en el enlace de hidrógeno, lo que se refleja en una ligera preferencia del grupo Asp<sup>32</sup> a encontrarse desprotonado en la forma libre de la enzima (Beveridge & Heywood, 1993; Goldblum, 1988; Goldblum et al., 1993; Iliadis et al., 1996; Iliadis et al., 1994).

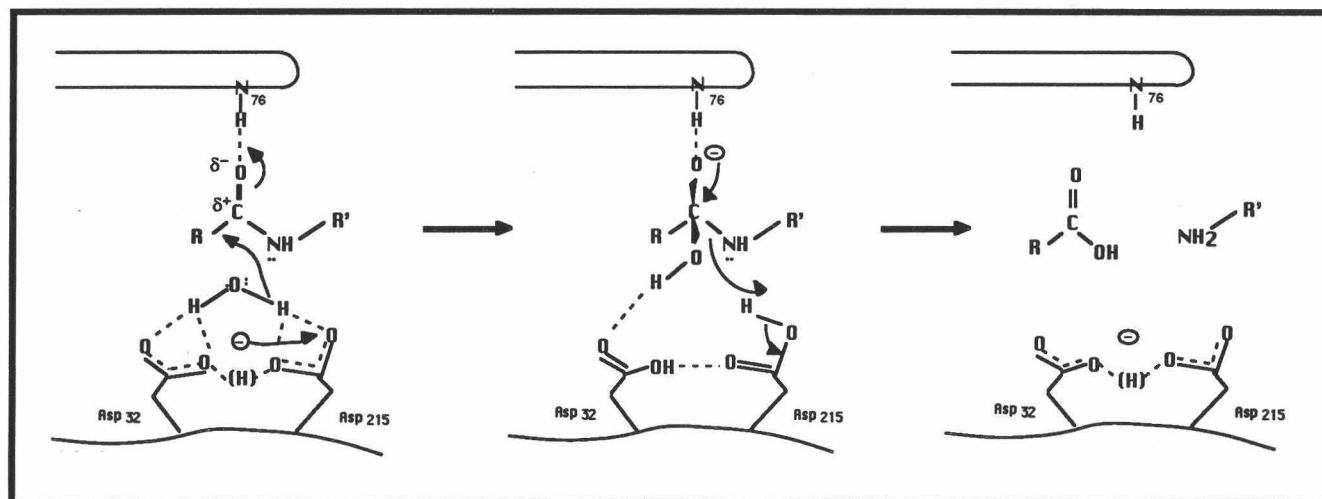
En general, las proteasas aspárticas presentan curvas de actividad en función del pH de forma acampanada (Lin et al., 1992), de tal modo que la mayoría de ellas presentan máxima actividad catalítica en el intervalo de pH entre 2 y 4, con la excepción de renina y las proteasas retrovíricas que actúan a pH cercanos a la neutralidad, entre 5.5 y 7.5. Este comportamiento frente a los cambios de pH ha sido atribuido a la existencia de anomalías en los  $pK_a$  de los grupos Asp catalíticos con respecto a los valores promedio observados en compuestos modelo con un único grupo carboxílico, de forma que, mientras que uno de los aspárticos presenta un  $pK_a$  excepcionalmente elevado, alrededor de 5, el segundo está caracterizado por un  $pK_a$  anormalmente bajo, cercano a 2.

Este comportamiento parece tener su origen en la proximidad espacial de ambos grupos carboxílicos, que deben estar en contacto y mantener sus orientaciones relativas dada la rigidez del centro activo, originándose, por tanto, importantes interacciones repulsivas. Como consecuencia, el estado completamente desprotonado se encuentra muy desfavorecido energéticamente, lo que conlleva el incremento del  $pK_a$  correspondiente a la pérdida del último protón. Por el contrario, una vez que existe un protón compartido por ambos aspárticos, la adición de un segundo es menos probable por lo cual el valor de  $pK_a$  correspondiente debe ser menor de lo habitual. Esta situación se ha asimilado a la existente en ácidos dicarboxílicos con restricciones conformacionales, habiéndose demostrado que la simple existencia de un grupo carboxilato y un carboxilo orientados convenientemente en presencia de una molécula de agua es suficiente para desarrollar la actividad catalítica característica de las proteasas aspárticas (Iliadis et al., 1996).

#### II.B.2.4.-Mecanismo catalítico.

En contraposición al mecanismo propuesto inicialmente, que suponía la existencia de intermedios de tipo covalente en el proceso de hidrólisis (Fruton, 1976; Newmark & Knowles, 1975; Takahashi et al., 1974), diversas evidencias experimentales, obtenidas mediante estudios cinéticos en estado estacionario y pre-estacionario, experimentos de intercambio isotópico (Antonov, 1985; Antonov et al., 1978; Antonov et al., 1981), estudios de resonancia magnética nuclear (Schmidt et al., 1985), así como el análisis de la variación de los parámetros cinéticos con el pH (Hofmann et al., 1984), han conducido al establecimiento de un mecanismo de acción no covalente que comprende una etapa de catálisis ácido-base general (Hsu et al., 1977; James & Sielecki, 1985; James et al., 1985; Pearl, 1987; Pearl & Blundell, 1984; Polgar, 1987; Suguna et al., 1987a). Esta hipótesis ha sido validada, tanto mediante estudios cristalográficos que demuestran que la geometría del sitio activo no permite la formación de intermedios covalentes durante la catálisis (Hsu et al., 1977), como por numerosos estudios estructurales de los complejos de varias enzimas fúngicas con inhibidores oligopeptídicos (Bott et al., 1982; Hallet et al., 1985; James et al., 1982).

Cualquier modelo mecanístico aceptable debe ajustarse a diversas observaciones experimentales, características de la catálisis de estas enzimas, como la amplia variación de la velocidad catalítica con la longitud y composición de los sustratos, la eficiencia similar en la catálisis de la hidrólisis tanto de ésteres como de enlaces peptídicos, o la ausencia de efectos isotópicos.



*Figura II.B.2: Mecanismo de hidrólisis del enlace peptídico catalizado por proteasas aspárticas, ilustrado para el caso concreto de Endothiapepsina.*

El mecanismo aceptado universalmente (Iliadis et al., 1996; James et al., 1992; Pearl,

1987; Polgar, 1987; Silva et al., 1996; Suguna et al., 1987b) y mostrado en la figura II.B.2, consiste en el ataque nucleofílico al carbono del enlace peptídico, llevado a cabo por la molécula de agua presente en el centro activo, coordinada entre los dos aspárticos catalíticos, y que, encontrándose distorsionada por el efecto del fuerte campo ejercido por éstos, se convierte en un nucleófilo muy efectivo. Por otro lado, existen diversas evidencias de una ligera distorsión en el enlace peptídico que sufre el proceso de hidrólisis a consecuencia de las interacciones establecidas entre el sustrato y los distintos restos de la enzima, lo que se traduce en la disminución de su carácter de doble enlace y en una mayor polarización del grupo carbonilo, de forma que el ataque nucleofílico se encuentra adicionalmente facilitado.

Como resultado del ataque nucleofílico inicial se produce la ruptura de la molécula de agua con la consecuente adición de un grupo hidroxilo al carbono del enlace peptídico, originando un intermedio tetraédrico de transición, proceso que se encuentra asociado con la protonación simultánea de uno de los carboxilos. Posteriormente el nitrógeno amida del enlace peptídico acepta un protón procedente del grupo carboxílico conduciendo a la ruptura del intermedio. Se ha propuesto la posibilidad de que los enlaces de hidrógeno establecidos entre el "flap" y el átomo de oxígeno puedan asistir en la orientación de los orbitales, de modo que durante la división del intermedio tetraédrico el grupo saliente sea la amina y no el nucleófilo original (Deslongchamps, 1975).

#### II.B.2.5.-Especificidad de sustratos.

En contraste con la situación encontrada con otras proteasas que se caracterizan por el establecimiento de una única interacción covalente con el sustrato, dado que las proteasas aspárticas llevan a cabo la catálisis mediante el ataque nucleofílico de una molécula de agua, el sustrato debe encontrar varios sitios de anclaje a ambos lados del enlace peptídico hidrolizable que posibiliten su orientación adecuada con respecto a los grupos carboxilo de los aspárticos catalíticos. Así, la unión de sustratos e inhibidores al centro activo de las proteasas aspárticas se lleva a cabo gracias al establecimiento de un gran número de interacciones de diversa naturaleza a lo largo del sitio de unión que, como se ha mencionado anteriormente, constituye un orificio cilíndrico que se extiende a través de toda la molécula y en el cual se pueden acomodar hasta ocho aminoácidos en diferentes subsitios de unión.

La selectividad intrínseca de la proteasa, que se encuentra determinada por las preferencias por los diferentes aminoácidos en cada uno de los subsitios de unión, constituye uno de los factores fundamentales que determinan la afinidad de unión de los sustratos así como la extensión con que la enzima lleva a cabo la hidrólisis de un enlace peptídico determinado. Como consecuencia, la selectividad en la catálisis de estas enzimas depende, en cierto modo, de la severidad de las restricciones impuestas en cada uno de los subsitios de unión. En general, con la excepción de Renina, que se cuenta entre una de las enzimas más específicas, las proteasas

aspárticas son poco selectivas, de forma que no muestran claras preferencias por aminoácidos determinados en ninguna de las posiciones. No obstante, existen limitaciones adicionales asociadas a la capacidad del sustrato para adoptar la conformación adecuada, así como a la accesibilidad al centro catalítico del enlace peptídico hidrolizable, que pueden convertirse en factores determinantes en el establecimiento de la selectividad de estas proteínas.

En general, el empleo de pequeños sustratos oligopeptídicos constituye una técnica ampliamente utilizada para el establecimiento de la selectividad de estas proteínas que permite examinar el efecto que alteraciones sistemáticas en la secuencia de los sustratos tienen en los parámetros cinéticos de hidrólisis, lo cual, de llevarse a cabo empleando sustratos proteicos, podría conducir a situaciones de muy difícil interpretación, dada la posibilidad de cambios conformacionales en la proteína sustrato como consecuencia de la mutación. Así, aunque el estudio de este tipo de sustratos permite el establecimiento de las preferencias en los distintos subsitios cercanos al centro catalítico, debe tenerse presente que su utilización elimina la posibilidad de evaluar los efectos de la conformación adoptada por el sustrato en la especificidad de estas proteasas.

Diversos estudios llevados a cabo con varias proteasas aspárticas han determinado la influencia tanto de la secuencia polipeptídica como de la longitud del sustrato en los parámetros de hidrólisis (Balbaa et al., 1993; Dunn et al., 1986). Así, se ha podido observar que pequeñas variaciones en la composición de un sustrato determinado tienen efectos muy variados en las distintas proteasas, lo cual demuestra la gran influencia en la especificidad de pequeñas variaciones en las secuencias de proteínas homólogas.

#### **II.B.2.6.-Estudios de inhibición.**

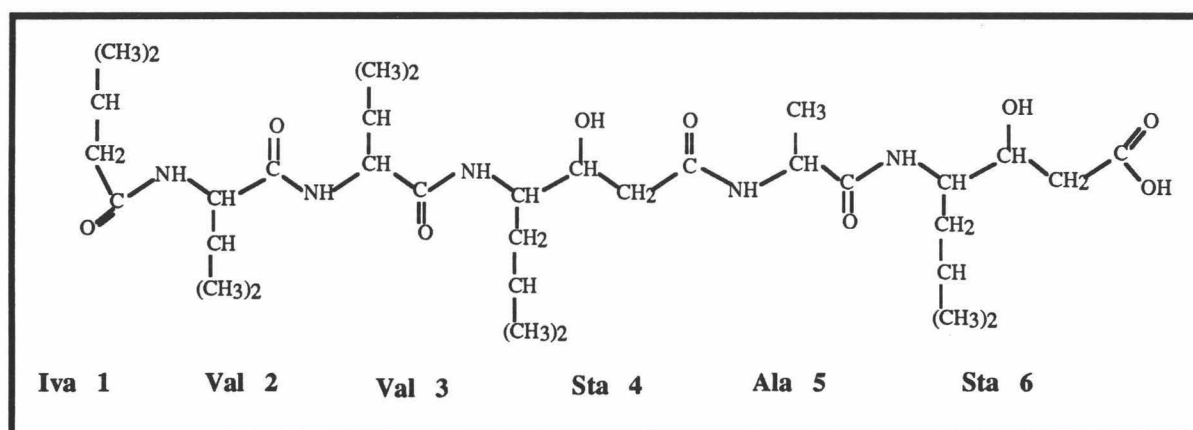
Dado que las proteasas aspárticas llevan a cabo la catálisis sin la intervención directa de ninguno de sus grupos funcionales, sino que es la molécula de agua coordinada entre los aspárticos del sitio activo la responsable del ataque inicial al enlace peptídico, su centro activo carece de la agresividad nucleofílica característica de las proteasas de serina o cisteína. El diseño de inhibidores para estas enzimas, por tanto, no puede estar orientado a la modificación química de un grupo específico del centro activo, sino que debe implicar estrategias más sutiles como la optimización de las interacciones secundarias a lo largo del sitio de unión o la substitución del enlace peptídico del sustrato por grupos no hidrolizables semejantes al estado de transición.

Diversos estudios sobre la especificidad de estas proteasas han permitido la identificación de aquellas secuencias aminoacídicas que optimizan las interacciones del sustrato en los distintos subsitios a lo largo del sitio de unión, de forma que, una vez obtenido el sustrato óptimo para la enzima, la estrategia más fructífera y más comúnmente empleada en el diseño de inhibidores potentes y específicos de las proteasas aspárticas se reduce conceptualmente a la substitución del enlace peptídico entre los aminoácidos en posiciones P1 y P1' por un grupo no hidrolizable,

análogo al estado tetraédrico de transición.

### II.B.2.6.1.-Inhibidores que contienen Estatina.

Pepstatina A, producto natural de fermentaciones bacterianas y cuya estructura química se muestra en la figura II.B.3, es un pentapéptido de secuencia Iva-Val-Val-Sta-Ala-Sta (donde Iva representa al ácido Isovalérico) que contiene dos residuos del aminoácido poco común Estatina(ácido 4(S)-amino-3(S)-hidroxi-6-metil heptanoico). Este compuesto es un potente inhibidor universal de las proteasas aspárticas, habiéndose determinado mediante diversos estudios cinéticos la importancia del resto central de Estatina en la afinidad de unión a estas enzimas (Rich et al., 1977), de tal forma que el grupo hidroxilo 3(S) de la estatina central se encuentra relacionado estructuralmente con uno de los grupos hidroxilo del estado tetraédrico de transición originado durante el proceso de hidrólisis.



*Figura II.B.3: Estructura química del inhibidor universal de las proteasas aspárticas Pepstatina-A*

Las constantes de inhibición son muy variables entre las diversas proteasas aspárticas de modo que, aunque en la mayoría de los casos pepstatina es un inhibidor de gran potencia ( $K_i$  para Pepsina es del orden de  $5 \times 10^{-11}$  M), la potencia inhibitoria es mucho menor en el caso de Renina y las proteasas retrovíricas, siendo en este último caso, una variante de pepstatina A, acetil-pepstatina, un inhibidor más eficiente.

Los derivados de pepstatina que contienen restos modificados de estatina en posición central que carecen del hidroxilo en configuración 3(S) constituyen, en general, inhibidores de menor potencia. De este modo, mientras que la eliminación del grupo hidroxilo conduce a la reducción de las constantes de afinidad a pepsina unas 4000 veces, la inversión de su configuración, de S a R, origina compuestos con capacidades inhibitorias unas 1000 veces



inferiores (Rich et al., 1977; Rich et al., 1980).

Datos estructurales de diversos complejos de estas enzimas con pepstatina (Bott et al., 1982; Cooper et al., 1989; Fitzgerald et al., 1990; James et al., 1982; Rich et al., 1985) y otros inhibidores basados en sustratos que poseen un resto de estatina en posición central muestran que el oxígeno de su grupo hidroxilo se encuentra situado entre los grupos carboxilo de los aspárticos del centro catalítico, como sustituto de la molécula de agua presente en la enzima libre, estableciendo diversos puentes de hidrógeno con los oxígenos de los carboxilos catalíticos. Los inhibidores que contienen estatinas con el hidroxilo en configuración 3(R) no presentan la disposición geométrica óptima para la formación de estas interacciones, lo que justifica la disminución observada en su afinidad (Rao & Singh, 1991). Asimismo, el desplazamiento de la molécula de agua coordinada entre ambos aspárticos puede aportar una contribución entrópica significativa, que justificaría las 5 kcal·mol<sup>-1</sup> de diferencia encontradas entre las afinidades de pepstatina y dehidroxipepstatina.

#### **II.B.2.6.2.-Reducción del enlace peptídico hidrolizable.**

A pesar de que la reducción del carbono carbonílico perteneciente al enlace peptídico hidrolizable constituye una posibilidad atractiva dada la similitud del grupo resultante [-CHOH-NH-] con el estado de transición [-CH(OH)<sub>2</sub>-NH-], la inestabilidad de la amina resultante limita sobremanera la utilidad de este tipo de compuestos, de forma que ha sido necesario la utilización de grupos funcionales estables de características similares.

Así, el reemplazamiento del hidroxilo por un grupo metileno [-CH<sub>2</sub>-NH-], empleado por primera vez en el desarrollo de inhibidores de Renina (Szelke et al., 1982), origina compuestos de gran potencia inhibitoria. La elevada afinidad de estos compuestos parece residir en la naturaleza tetraédrica del átomo de carbono que reemplaza al carbonilo así como en la presencia de la amina que presumiblemente se asemeja al nitrógeno de carácter básico generado como producto de la hidrólisis del enlace peptídico. El enlace reducido, que carece de las restricciones conformacionales del grupo peptídico de naturaleza planar, permite un mayor grado de flexibilidad en el enlace P1-P1', de modo similar a lo que cabría esperar en el estado de transición (Cooper et al., 1987).

Los análogos hidroxietilénicos [-CHOH-CH<sub>2</sub>-], cuya estructura retiene el grupo hidroxilo característico del enlace reducido, y en los cuales el nitrógeno se encuentra substituido por un grupo metileno, constituyen asimismo inhibidores de gran potencia, en el intervalo nanomolar; lo que sugiere la importancia de la función desempeñada por el hidroxilo (Szelke, 1985). Se ha observado que su presencia es crucial en la unión de varios inhibidores de Renina y así como la proteasa del virus V.I.H.-1 (Dreyer et al., 1989; Erickson et al., 1990; Rich et al., 1990; Roberts et al., 1990; Tomasselli et al., 1990b). Al igual que lo encontrado con el hidroxilo de los restos de estatina, parecen existir evidencias de la unión preferencial de uno de los posibles conformeros.

Un procedimiento alternativo para conseguir la estabilización de la amina originada a consecuencia de la reducción del enlace peptídico consiste en la interposición de un grupo metileno adicional entre el hidroxilo y el grupo amina, de modo que este tipo de inhibidores [-CHOH-CH<sub>2</sub>-NH-], a pesar de la presencia de un átomo adicional en el enlace P1-P1' en comparación con un enlace peptídico estándar, mantienen las interacciones favorables características de ambos grupos funcionales (Dann et al., 1986).

#### II.B.2.6.3.-Otros inhibidores.

Diversos tipos de inhibidores han sido generados mediante la introducción de distintas funciones como substitutos del enlace peptídico, como es el caso de glicoles (Lunney et al., 1993), norstatinas (Dhanaraj et al., 1992), fosfostatinas (Lunney et al., 1993), etc. Es de destacar el empleo de derivados cetónicos [-CO-CH<sub>2</sub>-] (Szelke, 1985) que, aunque más parecidos al sustrato que al estado de transición, se unen a la enzima en su forma gem-diol [-C(OH)<sub>2</sub>-CH<sub>2</sub>-], favorecida por la elevada afinidad de la enzima por el intermedio tetraédrico. La hidratación de las cetonas puede facilitarse incrementando el carácter electrofílico del carbono carbonílico, haciéndolo más susceptible al ataque nucleofílico, de forma que los derivados difluorocetónicos [-CO-CF<sub>2</sub>-] así como las difluorostatonas (Badasso, 1994; Veerapandian et al., 1992) producen inhibidores de mayor potencia (Gelb et al., 1985).

#### II.B.2.6.4.-Estructuras tridimensionales de los complejos.

Actualmente se dispone de un número considerable de estructuras tridimensionales de los complejos de las distintas proteasas aspárticas con sus inhibidores (Bailey & Cooper, 1994; Bailey et al., 1993; Blundell et al., 1987; Cooper et al., 1989; Cooper et al., 1987; Cooper et al., 1992; James et al., 1992; James et al., 1982; Lunney et al., 1993; Miller et al., 1989a; Rao & Singh, 1991; Rich et al., 1985; Sali et al., 1989; Suguna et al., 1992; Suguna et al., 1987b; Swain et al., 1990; Tong et al., 1993; Veerapandian et al., 1990; Workman & Burkitt, 1979), observándose en términos generales una gran similitud en el modo de unión así como en las interacciones establecidas entre la proteína y el inhibidor en todos ellos.

Todos los inhibidores se unen en conformación extendida, adoptando la posición correcta en el centro activo mediante el establecimiento de un elevado número de enlaces de hidrógeno entre el esqueleto de la cadena peptídica del inhibidor y diversos grupos de la enzima. Estos se encuentran situados tanto en la base del sitio de unión, caracterizada por su elevada estabilidad y rigidez, como en las zonas más flexibles de la enzima, que se estabilizan notablemente como consecuencia de dichas interacciones. Las cadenas laterales se encuentran dispuestas alternativamente a ambos lados del inhibidor, contribuyendo adicionalmente, mediante el establecimiento de extensivos contactos de van der Waals con los distintos subsitios de unión, a la

similitud conformacional entre los distintos inhibidores.

Como consecuencia de la formación del complejo, la región flexible o “flap”, que en la proteína libre descansa sobre el sitio activo, se cierra sobre el inhibidor, conduciendo a su casi total deshidratación. Por tanto, aunque el complejo mantiene algunas de las moléculas de agua, la unión de los inhibidores provoca la casi completa deshidratación del centro de unión, siendo especialmente relevante el desplazamiento de la molécula de agua que se encuentra coordinada entre los dos grupos carboxílicos de los aspárticos en la proteína libre y cuya posición pasa a ser ocupada por el grupo hidroxilo de ciertos análogos del estado de transición como estatinas, hidroxietilenos o difluorocetonas.

### II.B.3.-Endothiapepsina

#### II.B.3.1.-Origen y secuencia.

Endothiapepsina (EC 3.4.23.6) es una proteasa aspártica segregada de forma natural por el hongo filamentoso *Endothia parasitica*.. Actualmente se produce a escala comercial dada su elevada actividad como coagulador de leche, empleándose en la fabricación de quesos como sustituto de Quimosina aunque, en contraste con la elevada especificidad de ésta, Endothiapepsina presenta una actividad proteolítica más general.

La proteína está constituida por una única cadena polipeptídica de 330 resíduos, cuya secuencia presenta una homología elevada con otras proteasas aspárticas. En general, la identidad en las secuencias esta comprendida entre un 25 y un 30% para las enzimas procedentes de vertebrados como Pepsina, Quimosina, Gastricsina, Renina, o Cathepsina D, y no se encuentra especialmente relacionada con ninguna de ellas. En cuanto a su similitud con las otras proteasas fúngicas, Penicillopepsina, proteína producida por el hongo *Penicillium Jantinellum* y perteneciente a su misma familia presenta un 52% de homología, mientras la proteasa de *Mucor Pusillus*, muestra la mínima correspondencia en la secuencia, alrededor del 21%, inferior incluso que la encontrada con organismos superiores (Barkholt, 1987).

En cuanto al contenido en aminoácidos, la polaridad global de la proteína, un 45%, expresada como suma de los porcentajes de grupos polares (Asp, Asn, Glu, Gln, Thr, Ser, His, Lys y Arg) en la enzima, se encuentra dentro de los valores promedio para otras proteínas. Sin embargo, si se analizan los dominios amino y carboxi terminal independientemente, las polaridades son 55% y 37% respectivamente, valores realmente en los extremos de la distribución para las proteínas de la base de datos. Aunque no se conoce la relevancia fisiológica de este fenómeno, es sorprendente la similitud estructural existente con otras proteasas del grupo, aún a pesar de la ausencia de diferencias tan significativas entre sus dominios (Capaldi & Vandekooi, 1972).

### II.B.3.2.-Estructura tridimensional.

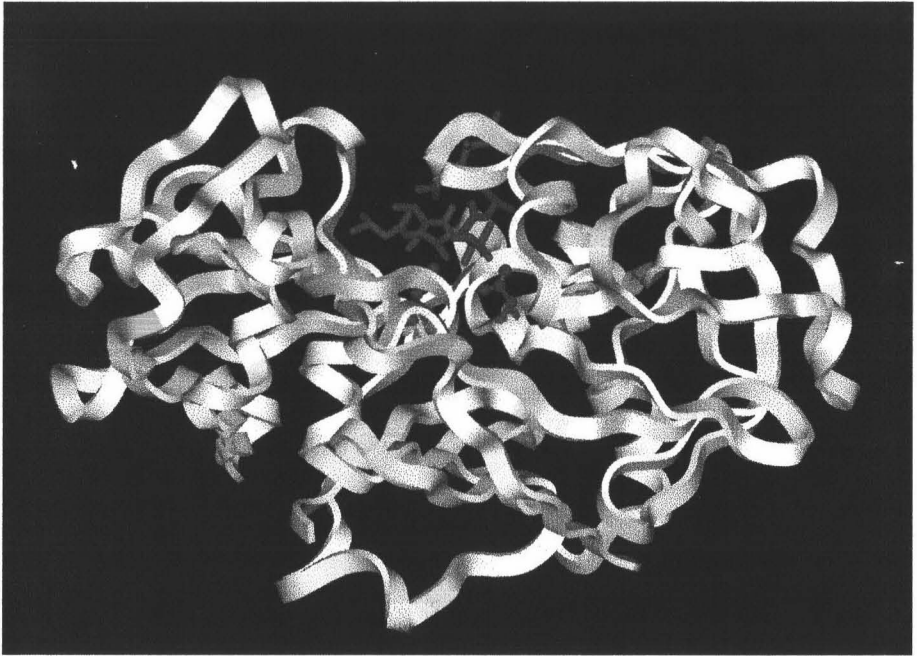
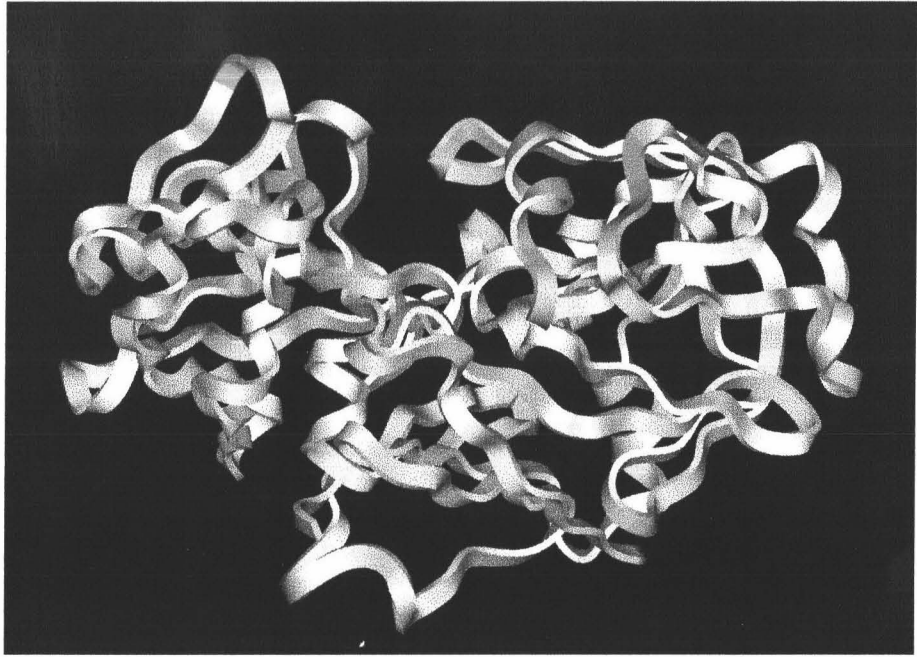
En los primeros estadios, el análisis cristalográfico de los datos de difracción de rayos X se vió obstaculizado por la indefinición de la secuencia aminoacídica, a pesar de lo cual existen en la literatura diversos análisis preliminares a alta resolución (Blundell et al., 1985; Jenkins et al., 1975; Pearl & Blundell, 1984; Subramanian et al., 1977). La determinación completa de la secuencia ha posibilitado finalmente el refinamiento de la estructura cristalográfica a 2.1 Å de resolución, y, por tanto, una descripción detallada de la estructura tridimensional de la proteína (Blundell et al., 1990). En el panel superior de la figura II.B.4 se muestra la estructura tridimensional de esta enzima libre en disolución, coloreada de acuerdo con los factores de temperatura de modo que las regiones en blanco representan las zonas con menor flexibilidad y las azules las más móviles.

En rasgos generales, Endothiapepsina presenta el plegamiento bilobular característico de las proteasas aspárticas, consistente en dos dominios de estructura secundaria fundamentalmente beta, que se encuentran separados por una hendidura que atraviesa la molécula y que conforma el centro activo. La estructura, en su mayoría constituida por láminas beta ligeramente distorsionadas de modo que cada una de sus hebras interacciona mediante dos o más enlaces de hidrógeno con las demás, consiste fundamentalmente en una gran lámina beta central de seis hebras antiparalelas, compartida por ambos lóbulos, y que se encuentra flanqueada por otras dos láminas mixtas, una por cada dominio estructural, de ocho hebras cada una y equivalentes entre sí que se encuentran dispuestas en forma de sandwich. Cada una de estas láminas laterales comparte la mitad de una de sus hebras con la lámina central, presentando una ligera torsión a la derecha de modo que ambas convergen por encima de ella para constituir la base del centro activo. En cada uno de los dominios existen adicionalmente dos pequeñas hélices alfa, ambas en la superficie de la proteína y que, por tanto, presentan una de las caras al disolvente.

Cada uno de los dominios, tanto el amino (restos -2 a 173 en notación de pepsina) como el carboxilo-terminal (restos 174 a 326), está constituido por 65 residuos topológicamente equivalentes que pueden superponerse, con una desviación promedio de 1.4 Å, mediante la rotación alrededor de un eje binario que atraviesa el sitio activo (Tang & Wong, 1987). La existencia de dicha simetría intramolecular conduce a la hipótesis de una posible evolución de la enzima a partir de un ancestro de naturaleza dimérica constituido por dos subunidades idénticas que se fusionaron con posterioridad para originar una sola cadena. Dicha hipótesis ha sido corroborada por la posterior identificación de las enzimas retrovíricas como proteasas aspárticas de carácter dimérico, que pueden ser consideradas como "fósiles" de las enzimas de organismos superiores.

Las secuencias alrededor del eje binario interlobular se encuentran muy conservadas en todas las proteasas aspárticas, de modo que la simetría es especialmente llamativa en las inmediaciones de los dos aspárticos catalíticos, Asp<sup>32</sup> y Asp<sup>215</sup> pertenecientes al dominio amino y carboxi-terminal respectivamente (James & Sielecki, 1983; Pearl & Blundell, 1984; Suguna et al.,

*Figura II.B.4: Estructura tridimensional de Endothiapepsina libre (panel superior) y de su complejo con Pepstatina A (panel inferior). Los átomos del inhibidor se muestran en color rojo. La proteasa se ha coloreado de acuerdo con los factores de temperatura de los distintos aminoácidos, de modo que las regiones en blanco representan las zonas de la roteína más rígidas y las azules las de mayor flexibilidad estructural.*





1987b). Asimismo, cada uno de los lóbulos presenta cierta simetría interna existiendo dos ejes binarios secundarios que atraviesan ambos dominios (Andreeva et al., 1984; Blundell et al., 1979). De este modo, la lámina beta interna de cada uno de los lóbulos presenta una arquitectura simétrica de enlaces de hidrógeno, a ambos lados del eje de rotación.

### II.B.3.3.-Estructura del centro activo.

A pH 4.5, el centro activo de la molécula de Endothiapepsina se caracteriza por una intrincada red, aparentemente simétrica, de enlaces de hidrógeno en la que se encuentran implicadas las secuencias homólogas Asp<sup>32</sup>-Thr<sup>33</sup>-Gly<sup>34</sup>-Ser<sup>35</sup> en el dominio amino-terminal y Asp<sup>215</sup>-Thr<sup>216</sup>-Gly<sup>217</sup>-Thr<sup>218</sup> en el carboxilo terminal.

Una molécula de agua, coplanar con ambos carboxilos, se encuentra situada a distancia adecuada para la formación de puentes de hidrógeno con los cuatro oxígenos de dichos grupos. La distancia entre los dos oxígenos carboxílicos (OD2), aproximadamente 3.0 Å, sugiere la existencia de un enlace de hidrógeno adicional entre ellos. Los dos oxígenos más internos interactúan con los nitrógenos de los grupos amida de los residuos Gly<sup>34</sup> y Gly<sup>217</sup> y, a pesar de que las distancias y orientaciones no son las adecuadas para el establecimiento de una interacción especialmente intensa, la proximidad de ambos restos al centro catalítico sugiere la participación de estos grupos en la estabilización del estado de transición. De la misma manera, los oxígenos externos establecen puentes de hidrógeno con los grupos alcohólicos de las cadenas laterales de los residuos Ser<sup>35</sup> y Thr<sup>218</sup>.

Las cadenas laterales de los aminoácidos Thr<sup>33</sup> y Thr<sup>216</sup>, situados simétricamente a ambos lados de los aspárticos catalíticos, están orientadas hacia el corazón hidrofóbico de la molécula e intervienen en una complicada red de enlaces de hidrógeno de tal forma que cada resto de Thr establece contactos con diversos átomos de la cadena principal del dominio opuesto. Este esquema, denominado "the fireman's grip", confiere gran estabilidad y rigidez el centro activo de la molécula (Blundell et al., 1990; Pearl & Blundell, 1984).

Al igual que ocurre en el resto de las proteasas aspárticas, en contraste con la rigidez y elevada estabilidad tanto del sitio activo como del núcleo hidrofóbico de la molécula, existe una región en el dominio amino-terminal, que en Endothiapepsina corresponde a los residuos 74 a 83, caracterizada por una elevada flexibilidad de modo que presenta los valores más elevados para los factores de temperatura en la proteína libre. Este loop, denominado "flap", que en la forma libre de la enzima descansa sobre el sitio de unión, debe abrirse para permitir el acceso de los sustratos e inhibidores. La unión del ligando a la enzima se produce gracias al establecimiento de fuertes interacciones favorables con la base del sitio activo de la proteína, mientras que, por su parte, el "flap" interactúa con los restos centrales de la cara expuesta del ligando de forma que éste queda casi totalmente imbuido en el centro activo de la enzima. La restricción conformacional a la que se ven sometidos los residuos involucrados en dichas interacciones se manifiesta en la

disminución de los factores de temperatura correspondientes a esta región en el complejo, como puede observarse en el panel inferior de la figura II.B.4 donde se muestra la estructura tridimensional del complejo de Endothiapepsina con Pepstatina A.

#### II.B.3.4.-Actividad hidrolítica.

El centro activo de la enzima, situado en la extensa abertura que atraviesa toda la molécula (Tang et al., 1978), puede albergar al menos siete residuos del ligando acomodados en los subsitios S6-S3', (Schechter & Berger, 1967), de modo que la ruptura del enlace peptídico tiene lugar entre los aminoácidos, generalmente de carácter hidrofóbico, que ocupan las posiciones S1 y S1' (Davies, 1990). La variación en la actividad catalítica de Endothiapepsina con los cambios en pH muestra la forma acampanada característica de todas las proteasas aspárticas, presentando máxima actividad hidrolítica a valores de pH comprendidos entre 2.0 y 2.5. Los valores de la constante de Michaelis para hidrólisis de diversos sustratos a pH 7 se encuentran en el intervalo milimolar, mientras que las constantes catalíticas toman valores comprendidos entre 15.2 y 0.004 s<sup>-1</sup> (Balbaa et al., 1993) según la naturaleza del sustrato.

El análisis de los parámetros cinéticos característicos de los procesos de hidrólisis de pequeños sustratos sintéticos demuestra que las constantes de velocidad no sólo dependen de la naturaleza de los restos adyacentes al enlace peptídico hidrolizable, sino que también se ven afectadas por la presencia y carácter de los aminoácidos en posiciones más alejadas de la unidad dipeptídica básica (Fruton, 1970; Sampath-Kumar & Fruton, 1974). Este hecho indica que la especificidad de Endothiapepsina se encuentra en gran parte determinada, no sólo por la naturaleza de los aminoácidos situados a ambos lados del enlace peptídico hidrolizable, sino que también está afectada por las interacciones secundarias del ligando con los distintos subsitios en el centro activo.

En general, los valores de la constante de Michaelis,  $K_m$ , parecen ser independientes de la longitud de la cadena, aunque existen pequeñas variaciones no sistemáticas, mientras que la constante catalítica puede verse modificada en varios órdenes de magnitud (Balbaa et al., 1993), lo que es indicativo del papel de las interacciones establecidas con los distintos subsitios en la disminución de la barrera de activación del proceso de hidrólisis mediante la estabilización del estado de transición (Fruton, 1970; Sampath-Kumar & Fruton, 1974). Los mayores incrementos en  $k_{cat}$  se producen al ocuparse las posiciones P3 y P2', lo cual es atribuible al establecimiento de enlaces de hidrógeno específicos entre la proteína y los aminoácidos en dichas posiciones.

Diversos estudios estructurales revelan la existencia de cierta flexibilidad entre los dominios, de forma que la proteína estaría constituida por dos "cuerpos rígidos" que pueden rotar uno con respecto al otro, y originar cambios en la geometría del sitio de unión (Sali, 1992; Abad-Zapatero, 1990). El papel asignado al resto en posición S3 como desencadenante de dichos cambios conformacionales justificaría las variaciones observadas en las constantes

catalíticas como resultado de sustituciones en dicha posición. Esta interpretación se ve limitada por la posible existencia de contactos cristalinos que puedan estar afectando las conformaciones de las regiones presuntamente implicadas en dichos cambios conformacionales (Sali, 1992). Desde un punto de vista mecanicista, algunos autores (Hofmann et al., 1988) sugieren que la unión del ligando en posiciones S3 y S2' facilita la conversión de un primer complejo enzima-sustrato en un complejo productivo mediante la distorsión del enlace peptídico hidrolizable, preparándolo para el ataque nucleofílico de la molécula de agua presente en el sitio activo (Pearl, 1987; Suguna et al., 1987b).

#### II.B.3.5.-Estudios de inhibición.

Como se ha mencionado anteriormente, el diseño de inhibidores para este tipo de proteasas se basa fundamentalmente en la generación de análogos del estado tetraédrico de transición, así como en la optimización de las interacciones secundarias a lo largo del sitio de unión. Por tanto, en un alto porcentaje, los inhibidores han sido desarrollados, bien mediante la sustitución del enlace peptídico hidrolizable por un análogo del estado de transición en sustratos naturales o sintéticos previamente optimizados, o bien mediante modificaciones de inhibidores naturales como pepstatina.

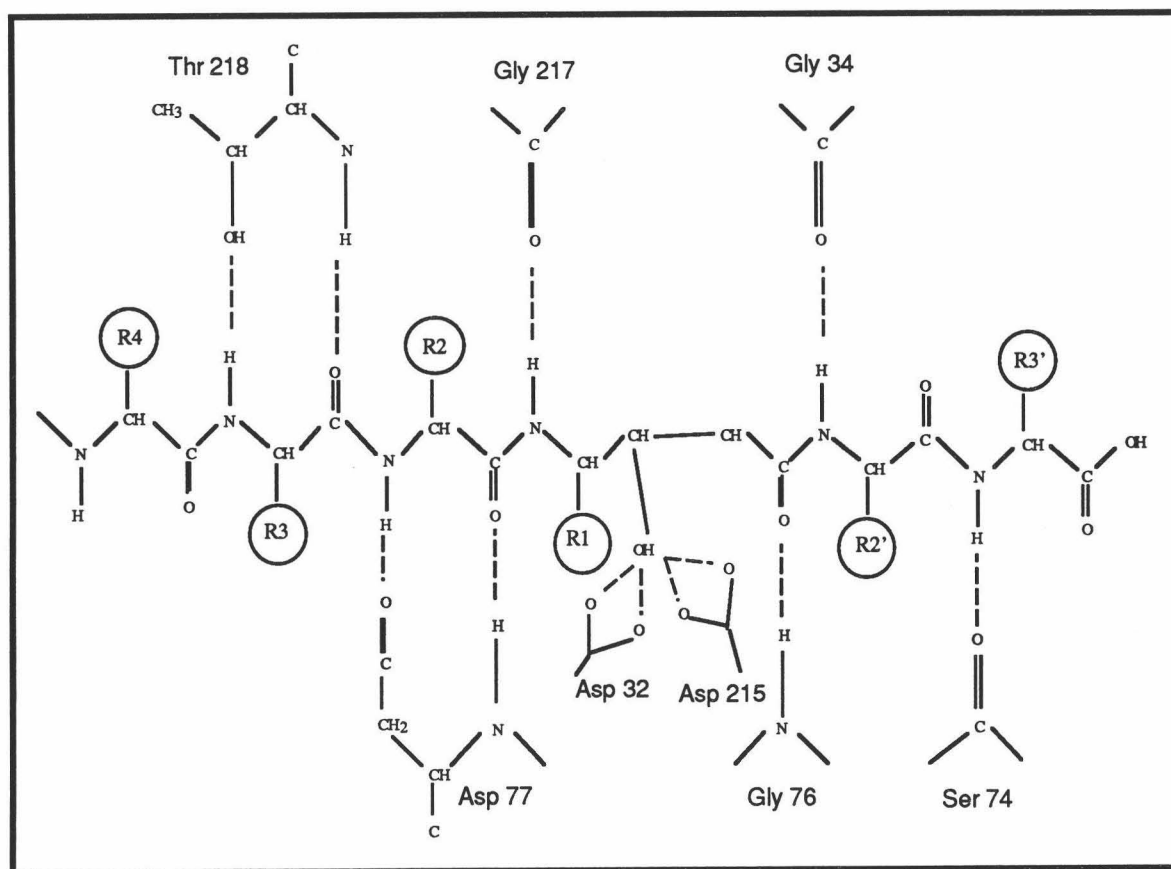
##### II.B.3.5.1.-Estructura Tridimensional.

En cuanto al análisis de las estructuras tridimensionales correspondientes a los distintos complejos con Endothiapepsina, se observa que todos los inhibidores adoptan conformaciones muy similares, de modo que los valores de los ángulos de torsión de sus cadenas principales en posición P4, P3, P1' y P2' se encuentran muy cercanos a los mínimos de energía correspondientes a una conformación extendida en lámina beta. La situación es ligeramente distinta para los residuos P1 y P2, cuyos ángulos adoptan valores correspondientes a una región del mapa de Ramachandran poco común en proteínas globulares, lo que refleja, bien la presencia de grupos de naturaleza no peptídica, o bien la existencia de deformaciones en la geometría del enlace peptídico como consecuencia del establecimiento de interacciones extensivas en los demás subsitios. Se ha postulado que la deformación del enlace peptídico entre los grupos P1 y P1' del sustrato es imprescindible para la adopción de conformaciones similares a las presentadas por los inhibidores, por lo que dicha deformación podría desempeñar un papel decisivo en el proceso de hidrólisis (Cooper et al.; 1987, Pearl, 1987).

La comparación de los distintos complejos de Endothiapepsina con sus inhibidores (Badasso, 1994; Bailey et al., 1993; Cooper et al., 1988; Cooper et al., 1989; Cooper et al., 1987; Cooper et al., 1992; Foundling et al., 1987; Lunney et al., 1993; Sali et al., 1989; Veerapandian et al., 1990; Veerapandian et al., 1992), revela una casi absoluta conservación de los enlaces

establecidos entre los átomos de la cadena principal del inhibidor y la enzima, mostrados esquemáticamente en la figura II.B.5. De especial relevancia son los enlaces formados entre el grupo hidroxilo central, presente en la gran mayoría de los análogos del estado de transición, y los dos Asp catalíticos, cuyas distancias y ángulos de enlace se encuentran cercanos a los valores óptimos. De entre los enlaces formados con la base del sitio activo cabe destacar los establecidos entre los nitrógenos amida de los grupos en posición P3, P2 y P2' con el grupo hidroxilo de la Thr<sup>219</sup>, y los oxígenos carbonílicos de los restos Gly<sup>217</sup> y Gly<sup>34</sup>, respectivamente, así como puente de hidrógeno entre el oxígeno carbonílico del grupo P3 y el nitrógeno amida del resto Thr<sup>219</sup>, como los conservados con mayor fidelidad entre todos los complejos.

En general, los enlaces de hidrógeno que mantienen la estructura del centro catalítico en Endothiapepsina no se ven alterados a causa de la unión del inhibidor ya que no se observa



*Figura II.B.5: Esquema de los enlaces de hidrógeno más relevantes establecidos entre Endothiapepsina y sus inhibidores.*

ningún cambio conformacional relevante, aunque se puede apreciar cierta disminución en la coplanaridad de los grupos carboxilos de los asparticos catalíticos en el complejo con respecto a la proteína libre. De entre las interacciones establecidas con la parte superior del centro activo cabe

destacar los enlaces entre los nitrógenos amida de las posiciones P2 y P2' con uno de los oxígenos carboxílicos del Asp<sup>77</sup> y el oxígeno del enlace peptídico de la Ser<sup>74</sup> respectivamente, así como el enlace bifurcado entre el nitrógeno amida de la Gly<sup>76</sup> y sendos oxígenos carbonílicos del inhibidor, en posiciones P1' y P2, estando este último implicado en un enlace adicional con el nitrógeno amida del Asp<sup>77</sup>.

Como consecuencia de la formación de estos enlaces, la flexibilidad característica de la región del flap (resíduos 71-82) en la proteína libre, debida fundamentalmente al alto contenido en Gly y Ser, se ve restringida ostensiblemente en el complejo, como se refleja en la disminución de sus factores de temperatura cristalográficos, y como puede apreciarse en la figura II.B.4, donde se representan las estructuras tridimensionales de la proteína libre y de su complejo con pepstatina A, coloreadas de acuerdo con los factores de temperatura. Es interesante destacar el hecho de que en los enlaces de hidrógeno establecidos con el "flap", los grupos donores se encuentran en el ligando y los aceptores en la enzima, mientras que la tendencia opuesta se manifiesta entre los enlaces formados con la base del centro activo, con la excepción de la interacción entre el nitrógeno amida de la Thr<sup>219</sup> con el carboxilo en posición P3.

Las cadenas laterales se encuentran dispuestas de forma alternante a ambos lados de la cadena principal e intervienen en la formación tanto de puentes de hidrógeno como de contactos de van der Waals con los grupos de la enzima que definen los subsitios de unión S6-S3'. En términos de interacciones de van der Waals, son los residuos en posiciones P1 y P3 los de mayor relevancia, debido a la naturaleza altamente hidrofóbica tanto de los grupos del inhibidor como de los subsitios de la enzima. En contraste con el comportamiento de la cadena principal, los enlaces de hidrógeno establecidos con las distintas moléculas de agua aún presentes en el complejo, así como entre las cadenas laterales del inhibidor y los grupos de la proteína, no se encuentran conservados entre los distintos inhibidores. La falta de correlación entre las diferencias observadas y la potencia inhibitoria de los distintos compuestos es consistente con la hipótesis de que únicamente los puentes de hidrógeno establecidos con la cadena principal son relevantes para la inhibición, determinando el alineamiento preciso del ligando en el sitio de unión de la enzima. De esta forma, la afinidad y, por tanto, la potencia de un determinado compuesto como inhibidor, debe estar determinada por otro tipo de interacciones, posiblemente contactos de van der Waals, entre las cadenas laterales de los ligandos y los distintos subsitios de unión de la proteína (Foundling et al., 1987; Sali et al., 1989).

#### II.B.3.5.2.-Tipos de inhibidores.

##### a.- Pepstatina A.

Como se ha mencionado anteriormente, Pepstatina A, pentapéptido de secuencia Iva-Val-Val-Sta-Ala-Sta, donde las abreviaturas Iva y Sta corresponden respectivamente al ácido



isovalérico y al aminoácido poco común estatina (ácido 4(S)-amino-3(S)-hidroxi-6-metil heptanoico), constituye un potente inhibidor para todas las proteasas aspárticas. El resto de estatina en posición central actúa como un análogo del estado tetraédrico de transición, ocupando las posiciones correspondientes a los dos aminoácidos entre los que se encuentra el enlace peptídico que sufre el proceso de hidrolización (Rich et al., 1977), de modo que el grupo hidroxilo en posición C3 sustituye a la molécula de agua unida a los grupos aspárticos en la proteína libre. Los inhibidores en los que el grupo estatina presenta configuración 3S presentan una mayor potencia que sus enantiómeros 3R, dada su mayor capacidad para el establecimiento de enlaces de hidrógeno con la enzima, como se desprende de diversos estudios estructurales (Bailey et al., 1993; Bott et al., 1982; James & Sielecki, 1985).

Pepstatina A se une a Endothiapepsina con alta afinidad, presentando constantes de asociación que dependen fuertemente del pH, de modo que, a valores cercanos al máximo de actividad catalítica, la constante de asociación tiene el valor  $2 \times 10^9$  M (pH 3.1 y 37 °C) [Boger, 1985 #81] mientras que a pH cercanos a la neutralidad la asociación disminuye bruscamente hasta  $8 \times 10^6$  M<sup>-1</sup> (pH 7.0 y 16 °C) (Gómez & Freire, 1995). Este comportamiento, junto con la fuerte dependencia de la entalpía aparente de unión con las entalpías de protonación de diversos tampones a pH 7, indica que la unión de pepstatina A a Endothiapepsina bajo estas condiciones se encuentra acoplada al equilibrio de protonación de uno de los grupos de la proteína, presumiblemente uno de los aspárticos catalíticos, de pK 4.3, de acuerdo con la dependencia de la constante de inhibición con el pH (Gómez & Freire, 1995). El análisis termodinámico del proceso indica que la unión de pepstatina A a Endothiapepsina se encuentra favorecida tanto entálpica como entrópicamente, de forma que la pérdida de entropía conformacional asociada a la restricción de los grados de libertad translacionales y conformacionales está compensada por la alta entropía de hidratación debida a la desolvatación de superficies hidrofóbicas en la formación del complejo.

El análisis termodinámico-estructural del proceso (Gómez & Freire, 1995) revela la existencia de varias regiones en la proteína que contribuyen significativamente a la energía libre de unión como es el caso de los aspárticos catalíticos y los aminoácidos en sus cercanías, así como la región correspondiente al "flap" flexible. Asimismo, la contribución de diversos restos lejanos al sitio de unión que refleja la existencia del cambio conformacional en el "flap", así como pequeños reajustes en la estructura.

#### *b.-Otros inhibidores.*

Entre los distintos inhibidores de Endothiapepsina se cuentan diversos derivados de Pepstatina A (Bailey et al., 1993) así como otros inhibidores que contienen estatina o sus derivados en posición P<sub>1</sub> (Cooper et al., 1989; Cooper et al., 1992; Lunney et al., 1993; Sali et al., 1989) y compuestos en los que el enlace peptídico hidrolizable ha sido reducido (Blundell et al.,



1987; Cooper et al., 1987; Foundling et al., 1987) o bien sustituido por grupos como fosfostatinas (Lunney et al., 1993), glicoles (Lunney et al., 1993), norstatinas (Dhanaraj et al., 1992), y aminoalcoholes (Cooper et al., 1988). Asimismo han sido cristalizados los complejos con Endothiapepsina de inhibidores conteniendo grupos hidroxietilénicos (Cooper et al., 1992; Veerapandian et al., 1990) o difluorostatonas (Badasso, 1994; Veerapandian et al., 1992).

Los mejores inhibidores de Endothiapepsina son generalmente aquéllos que contienen análogos hidroxietilénicos, estatinas, glicoles o fluorocetonas en posición  $P_1$  como H-261 (Veerapandian et al., 1990), H-189 (Bailey et al., 1993), pepstatina (Bailey et al., 1993), PD-133,450 (Lunney et al., 1993) o CP-81,282 (Veerapandian et al., 1992) cuyas constantes de inhibición se encuentran en el intervalo de 1 a 10 nM. Los análogos de fosfostatina y aminoalcoholes, PD-130,328 (Lunney et al., 1993) y BW624 (Cooper et al., 1988), respectivamente, presentan constantes de unión más débiles probablemente debido a su menor similitud con el estado de transición.

#### **II.B.4.-Proteasa del virus del S.I.D.A. tipo I.**

##### **II.B.4.1.-Consideraciones generales.**

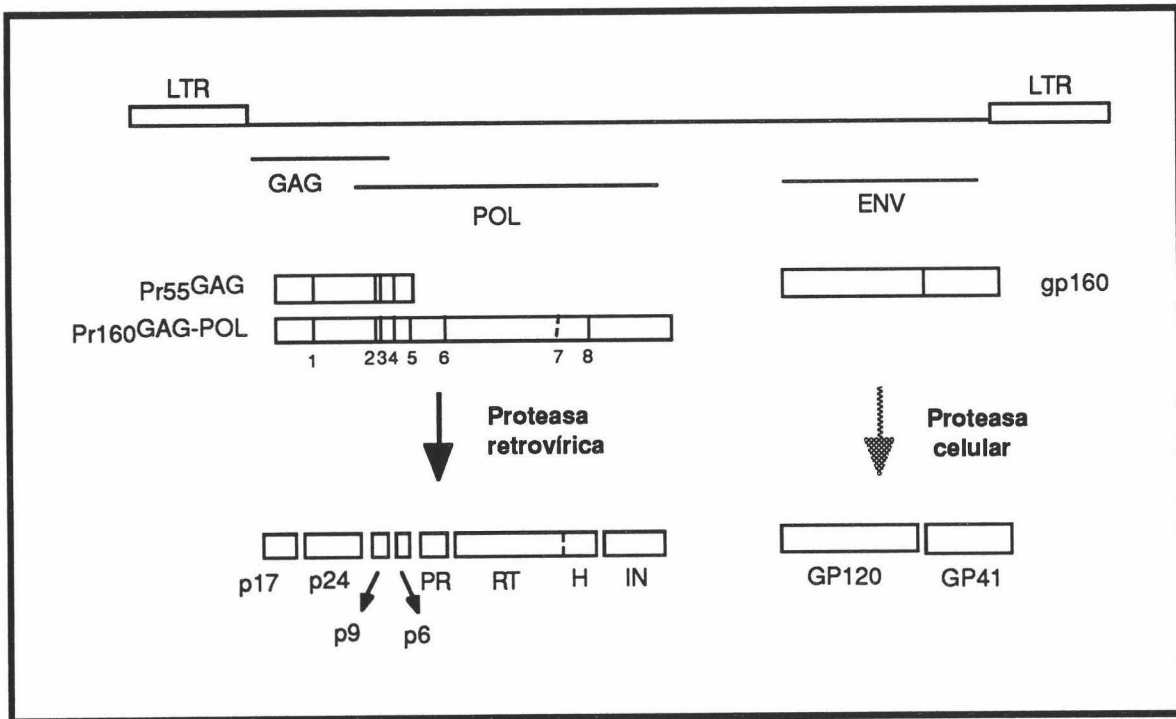
###### **II.B.4.1.1.-Proteasas retrovíricas.**

El virus del S.I.D.A. o V.I.H. pertenece al grupo de los lentivirus, que se integra en la ampliamente extendida familia de los retrovirus (Retroviridae) y como tal su genoma se encuentra constituido por tres regiones características, dispuestas en el orden 5'-LTR-gag-pol-env-LTR-3'. De este modo, el resultado del proceso de transcripción es la producción de ciertos precursores poliproteicos cuyo posterior procesamiento, a manos de una proteasa específica (Debouck et al., 1987; Giam & Boros, 1988; Strickler et al., 1989), origina las diversas proteínas estructurales así como las distintas enzimas necesarias para la generación de partículas infecciosas. En la figura II.B.6 se esquematiza el proceso de transcripción y procesado de los precursores poliproteicos en el virus V.I.H.-1.

Durante el proceso de transcripción, el virus V.I.H. produce tres precursores poliproteicos: Gp160, Pr55gag y Pr160gag-pol que contienen las distintas proteínas funcionales y estructurales necesarias para su desarrollo. En el caso del precursor Gp160, que contiene dos proteínas estructurales (gp120, que participa en la formación del envoltorio externo del virus y la proteína transmembrana gp41), es una proteína celular de la familia de la tripsina la encargada del proceso de hidrólisis. Por el contrario, el procesamiento de los precursores Pr55gag y Pr160gag-pol requiere la intervención de la propia proteasa del virus.

La proteólisis de estas dos regiones, Pr55gag (que contiene diversas proteínas estructurales organizadas en la secuencia  $\text{NH}_2\text{-p17-p24(25)-p7-p6-COOH}$ ) y Pr160gag-pol

(originado como fusión de las regiones gag y pol y que, junto con las proteínas estructurales mencionadas anteriormente, contiene diversas enzimas como la proteasa del virus (PR), la transcriptasa reversa (RT), la ribonucleasa H (H) y la integrasa (IN)), es esencial para la culminación del ciclo de vida del virus y la producción de partículas infecciosas.



**Figura II.B.6:** Organización genética y precusores poliproteicos del virus V.I.H.-1. Se muestran las secuencias terminales repetitivas (Long Terminal Repeat) en ambos extremos, las tres zonas de lectura principales (GAG, POL y ENV), los precusores poliproteicos resultantes de su transcripción ( $Pr55^{GAG}$ ,  $Pr160^{GAG-POL}$ , y  $gp160$ ), así como sus correspondientes proteínas maduras.

Dado que la supervivencia del virus depende de su capacidad para separar los distintos componentes de los precusores polipeptídicos de forma efectiva, así como de la limitación de dicha hidrólisis a ciertos puntos muy específicos, su evolución se ha dirigido hacia el desarrollo de proteasas propias, evitando de este modo la intervención en el proceso de enzimas celulares, que carecen de la especificidad necesaria y que generalmente se encuentran reguladas por un gran número de inhibidores celulares específicos. Por tanto, la posibilidad de inhibición de la actividad de estas proteasas retrovíticas de forma selectiva, evitando interferencias con otras enzimas celulares, constituye una atractiva posibilidad terapéutica.

Las proteasas retrovíticas, al contrario que las encontradas en otros tipos de virus,

pertenecen al grupo de las proteasas aspárticas, y presentan, por tanto, las características esenciales de este tipo de enzimas. La intervención de dos grupos aspárticos en su mecanismo catalítico, así como la conservación del motivo Asp-Thr-Gly en el centro activo, fué demostrada mediante análisis mutacionales, observándose que la substitución de estos grupos conduce a la inactivación de la enzima (Darke et al., 1988; Kohl et al., 1988; Loeb et al., 1989a; Loeb et al., 1989b). Por lo tanto, estas proteasas son susceptibles de inhibición por los reactivos característicos de las proteasas aspárticas, entre los que cabe destacar el inhibidor universal del grupo, pepstatina (Darke et al., 1989), siendo las proteasas de los virus V.I.H-1 y V.I.H.-2 especialmente sensibles a uno de sus derivados: acetil-pepstatina (Richards et al., 1989a; Richards et al., 1989b). Estas enzimas presentan asimismo curvas características acampanadas que describen la variación de  $V_{max}$  así como  $V_{max}/K_m$  en función del pH (Meek et al., 1989), lo que implica la intervención de un grupo carboxílico desprotonado en el mecanismo catalítico.

Por otra parte, la elucidación de la estructura tridimensional de diversas proteasas retrovíricas (Miller et al., 1989b; Navia et al., 1989; Wlodawer et al., 1989) ha revelado la presencia de los rasgos estructurales característicos de las proteasas aspárticas, aunque estas enzimas, al contrario que las eucarióticas, constituidas por una única cadena polipeptídica, son proteínas diméricas constituidas por dos monómeros idénticos. Cada uno de los monómeros presenta un único dominio, adquiriéndose la estructura bilobular característica de esta familia de proteasas tras el proceso de dimerización (Kato et al., 1989; Meek et al., 1989; Miller et al., 1989b). Cada monómero aporta la mitad de los componentes esenciales del único centro activo de la molécula, incluyendo los aspárticos catalíticos, los residuos Thr y Gly (conservados en todas las proteasas aspárticas y de gran importancia para la determinación de la geometría del sitio de unión), así como los lazos flexibles o "flaps" que constituyen la parte superior del sitio de unión. Dado que un monómero aislado, aún conteniendo la secuencia Asp-Thr-Gly, carece de un sitio catalítico completamente desarrollado que permita la interacción adecuada con el sustrato y el desarrollo de la catálisis, la formación del dímero activo debe tener lugar cuando la proteasa aún se encuentra formando parte del precursor poliproteico gag-pol, lo que, según algunos autores podría contribuir a la delicada regulación de su actividad enzimática *in vivo*.

#### II.B.4.1.2.-Proteasa del virus del S.I.D.A. como objetivo terapéutico.

Los objetivos tradicionales en el desarrollo de fármacos de acción antivírica han sido las proteínas encargadas de los procesos de ensamblaje y maduración del virus. La comprensión a nivel molecular de las estrategias desarrolladas por los diversos virus para llevar a cabo la infección celular, ha revelado el mecanismo de acción de muchos medicamentos utilizados habitualmente y, por tanto, ha conducido al descubrimiento y desarrollo de nuevos fármacos, más potentes y específicos.

En el caso del S.I.D.A., en un principio la lucha contra la infección se centró

fundamentalmente en el desarrollo de inhibidores específicos de la retrotranscriptasa, habiéndose obtenido resultados esperanzadores con diversos fármacos como azytodimidina (AZT) o aciclovir. Sin embargo, la aparición de cepas víricas resistentes a estos compuestos junto con los avances en los estudios sobre la proteasa, que han permitido la mejor comprensión de su función en los procesos de maduración del virus y de su mecanismo de acción, han convertido a esta última en uno de los objetivos fundamentales en el desarrollo de agentes terapéuticos.

Entre las estrategias más importantes orientadas hacia la inactivación de la proteasa se cuentan desde la búsqueda de productos naturales con capacidad de inhibición (Gnabre et al., 1995) o el desarrollo de análogos del estado de transición basados en sustratos peptídicos, hasta la aplicación, posibilitada gracias a la elucidación de un número creciente de estructuras tridimensionales de complejos entre la enzima y diversos inhibidores, de técnicas de modelado molecular al diseño de nuevos fármacos de estructuras complementarias a la cavidad delimitada por el sitio activo (Martin, 1992; Vacca, 1994; Wlodawer & Erickson, 1993).

La necesidad del proceso de dimerización para la adquisición de la actividad catalítica de la proteasa proporciona un mecanismo alternativo para la inhibición de la enzima, complementario a los métodos de inactivación dirigidos al centro activo, dado que la obstaculización del ensamblaje de los homodímeros, al igual que la destrucción de la interfase de dimerización, deben ser métodos efectivos para bloquear la actividad catalítica. En este contexto, se ha observado la pérdida de actividad enzimática en presencia del tetrapéptido correspondiente al extremo amino-terminal de la enzima (Zhang et al., 1991), así como la inactivación de la proteasa a consecuencia de la formación de complejos heterodiméricos entre monómeros de la proteína nativa y monómeros mutantes que presentan, bien sustituciones en el sitio activo que imposibilitan la catálisis (Babé et al., 1995), o bien modificaciones del extremo carboxi-terminal que impiden el establecimiento de las interacciones necesarias para la dimerización (Babé et al., 1991; McPhee et al., 1996).

#### **II.B.4.2.-Estructura Tridimensional.**

El avance de los estudios estructurales sobre la proteasa del virus del S.I.D.A. se vió inicialmente obstaculizado por la dificultad que presentaba su obtención dada su presencia minoritaria en el virus, por lo que fué necesario el empleo de técnicas de síntesis, así como la expresión de la proteína en diversos sistemas bacterianos, para la consecución de las cantidades requeridas para la aplicación de las técnicas estructurales. Estas dificultades iniciales condujeron al desarrollo de varios modelos estructurales, basados inicialmente en las estructuras tridimensionales de las proteasas aspárticas eucarióticas (Pearl & Taylor, 1987; Pechik et al., 1989) y, más recientemente, en la estructura cristalográfica de la proteasa del Rous Sarcoma Virus (Weber et al., 1989).

Una vez clonada en varios sistemas (Cheng et al., 1990; Darke et al., 1988; Debouck et

al., 1987), así como sintetizada químicamente en su totalidad (Schneider & Kent, 1988), la estructura tridimensional, tanto de la enzima recombinante (Lapatto et al., 1989; Navia et al., 1989) como de la sintética, (Wlodawer et al., 1989) han sido finalmente elucidadas. La comparación entre ambas ha permitido demostrar por vez primera la equivalencia estructural entre una enzima recombinante y la generada mediante síntesis química. La estructura tridimensional de esta enzima se muestra en la figura II.B.7, donde el panel superior presenta la enzima libre y el inferior el complejo entre ésta y el inhibidor acetil-pepstatina. En ambos casos, las estructuras se han coloreado de acuerdo con los factores de temperatura de modo que las regiones azules representan las zonas de mayor flexibilidad en la enzima.

La topología general de los monómeros de la proteasas del virus del V.I.H. tipo I es muy similar a la de uno de los dominios de las proteasas aspárticas del grupo de Pepsina (Blundell et al., 1985), encontrándose las mayores discrepancias en la región correspondiente a la interfase de dimerización, constituida por los extremos amino y carboxi-terminal de cada monómero. En estas regiones los monómeros de la proteasa retrovírica presentan hebras significativamente más cortas que las topológicamente equivalentes en las proteasas eucarióticas. Una vez constituido el dímero, la estructura general es muy similar a la presentada por las proteasas de organismos superiores, constituidas, asimismo, por dos dominios estructurales.

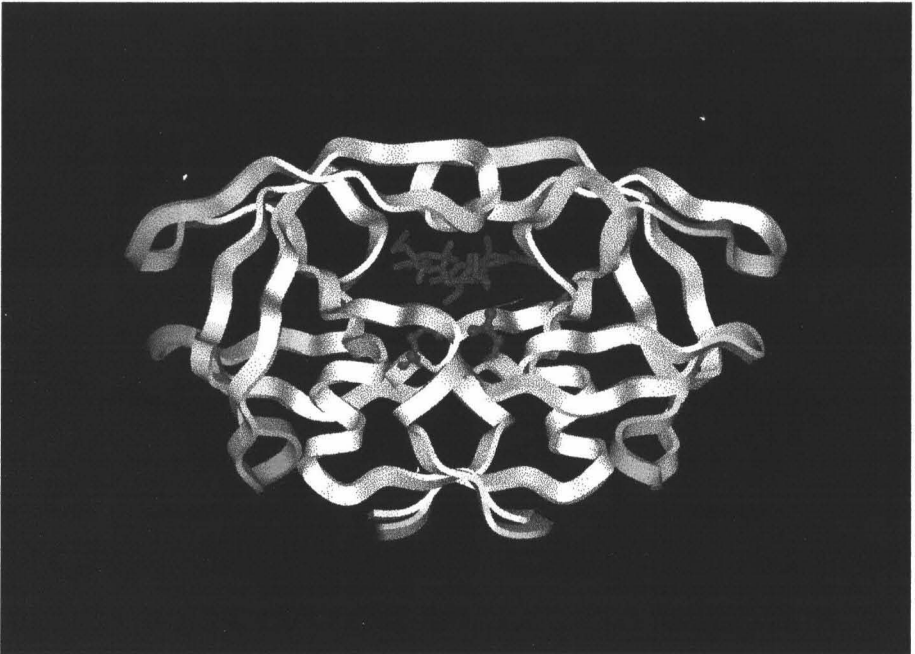
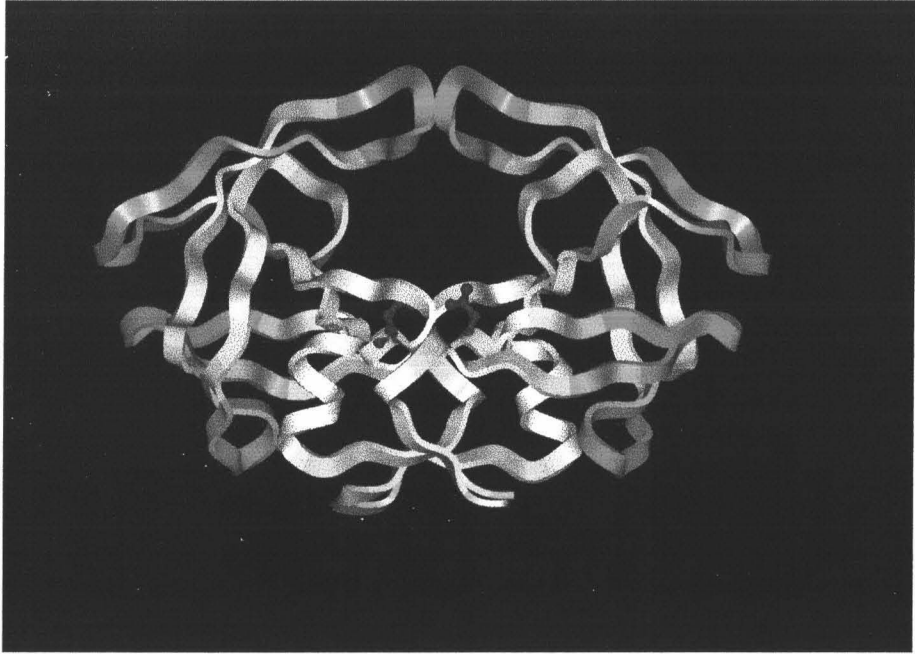
Al igual que en las demás proteasas del grupo, con la excepción de una pequeña hélice, la estructura está constituida fundamentalmente por láminas beta cuyas hebras se encuentran dispuestas de forma aproximadamente simétrica con respecto a un eje de simetría binario, equivalente al encontrado en cada uno de los dominios de las proteasas monoméricas.

La triada característica del centro activo (Asp<sup>25</sup>-Thr<sup>26</sup>-Gly<sup>27</sup>) se encuentra situada en un lazo cuya estructura se encuentra estabilizada por una red de enlaces de hidrógeno muy similar a la encontrada en las proteasas aspárticas eucarióticas (Davies, 1990; Wlodawer & Erickson, 1993; Wlodawer et al., 1989). Los grupos carboxilato de los Asp<sup>25</sup> de ambas cadenas se encuentran dispuestos casi coplanarmente, de modo que sus oxígenos OD2 participan en varios contactos con otros grupos de la enzima. La estructura de la base del centro activo se caracteriza por su elevada rigidez, consecuencia del establecimiento de la interacción conocida como “the fireman grip” entre las cadenas laterales de la Thr<sup>26</sup> de cada monómero con el nitrógeno amida de la Thr correspondiente en la cadena opuesta, encontrándose asimismo cada una de ellas formando parte de un enlace de hidrógeno adicional con el oxígeno carbonílico del residuo 24 en el monómero complementario. Mientras que las características fundamentales del centro catalítico son muy similares en las proteasas retrovíricas y en las celulares, los residuos adyacentes a la triada catalítica son bastante diferentes, encontrándose, por ejemplo, en las proteasas víricas un resto Ala en una posición en la cual Ser o Thr es casi un invariante en pepsinas.

Como consecuencia de su naturaleza dimérica, las proteasas retrovíricas presentan dos “flaps” simétricos, uno en cada cadena, mientras que las proteasas celulares poseen sólo uno, generalmente de menor tamaño. Estos “flaps” que cubren el sitio activo y participan en la unión

*Figura II.B.7: Estructura tridimensional de la proteasa del virus V.I.H.-1 libre (panel superior) y de su complejo con acetil-pepstatina (panel inferior). El inhibidor se muestra en color rojo. La enzima se ha coloreado de acuerdo con los factores de temperatura de los distintos aminoácidos, de modo que las regiones en blanco representan las zonas de la proteína de mayor rigidez y las azules las más flexibles. Las cadenas laterales mostradas corresponden a los dos restos aspárticos catalíticos.*





de inhibidores y sustratos, se encuentran en la proteína libre a unos 7 Å del centro activo, presentando, por tanto, una conformación muy abierta. De cualquier modo, el hecho de que en la proteasa del virus V.I.H.-2 estos lazos flexibles se encuentren en posiciones mucho más cerradas parece indicar que la adopción de esta conformación se debe posiblemente a artefactos de la cristalización y no a la existencia de un mínimo real en la energía. De hecho, las diferentes conformaciones encontradas en las estructuras cristalográficas, además de los valores especialmente elevados de sus factores de temperatura, constituyen evidencias de que, en ausencia de ligandos, estas regiones presentan una elevada flexibilidad en disolución (Wlodawer & Erickson, 1993). Dado que el tamaño de estos "flaps" en la proteasa del virus V.I.H.-1 es mayor que el habitual en sus homólogos eucarióticos, la accesibilidad al disolvente de los sustratos e inhibidores en el complejo es muy pequeña, de modo que el ligando queda prácticamente engullido por la proteína, por lo que no es sorprendente la importancia de estas regiones en la determinación tanto de la especificidad por los distintos sustratos como de la energética de unión.

#### **II.B.4.3.-Actividad hidrolítica.**

##### **II.B.4.3.1.-Especificidad de sustratos.**

La comprensión de los factores determinantes de la especificidad de la proteasa del virus V.I.H.-1 representaría un adelanto importante en el desarrollo de inhibidores selectivos y potentes de la enzima. Sin embargo, al igual que ocurre con el resto de las proteasas aspárticas, la selectividad de esta enzima es el resultado de la compleja combinación de numerosas contribuciones de muy distinta naturaleza, por lo que, a pesar de la abundancia de datos experimentales, la elucidación de los factores que determinan especificidad de la proteasa por sus sustratos presenta grandes dificultades.

El establecimiento de múltiples interacciones entre el sustrato polipeptídico y diversos grupos de la enzima, favorecido por la plasticidad, observada en diversas estructuras tridimensionales (Erickson et al., 1990; Miller et al., 1989b; Swain et al., 1990), que ambos, proteína y sustrato, presentan, conduce a la unión efectiva de éste mediante la formación de una complicada red de puentes de hidrógeno que sitúan el enlace peptídico hidrolizable en posición adecuada para la catálisis, de modo que las cadenas laterales de los distintos residuos deben adaptarse a los subsitios de unión dispuestos alternativamente a ambos lados del centro activo.

Como se ha comentado anteriormente, existen dos alternativas diferentes y en cierto modo complementarias para abordar el problema: por un lado, el estudio de la variación de los parámetros cinéticos con la modificación sistemática de la secuencia de pequeños péptidos y, por otro, el estudio de la hidrólisis de sustratos proteicos.

### a.-Sustratos oligopeptídicos.

El análisis de los péptidos representativos de los puntos de corte en los sustratos naturales de la enzima constituye lógicamente la etapa inicial en el proceso de esclarecimiento de su selectividad (Darke et al., 1988; Henderson et al., 1988; Hui et al., 1990; Konvalinka et al., 1990; Moore et al., 1989; Pettit et al., 1991; Phylip et al., 1990; Poorman et al., 1991; Tomasselli et al., 1991; Tomasselli et al., 1990a; Tozser et al., 1991a). La proteasa del virus V.I.H.-1 es responsable de la hidrólisis de ocho enlaces peptídicos en los precursores gag y gag-pol, que se encuentran flanqueados por secuencias bastante heterogéneas entre sí.

De este modo, los sustratos naturales se han clasificado atendiendo a la naturaleza de los aminoácidos adyacentes al enlace peptídico hidrolizable, habiéndose llegado a establecer inicialmente hasta cuatro clases (Henderson et al., 1988). Sin embargo, actualmente sólo se distingue entre aquéllos sustratos que presentan un resto de prolina en posición P1' y los demás, que generalmente presentan dos residuos de carácter hidrofóbico a ambos lados del sitio de corte (Griffiths et al., 1992a). En general, son muy pocas las enzimas, con excepción de algunos casos específicamente desarrollados con este fin, que pueden hidrolizar enlaces en los que interviene un resto de Prolina, por lo que la relativa frecuencia de este aminoácido en los sustratos de esta enzima, puede ser un mecanismo de protección contra la acción hidrolítica de las proteasas celulares. En la tabla II.B.1 se muestran las secuencias correspondientes a los principales puntos de corte de la proteasa en los distintos precursores poliproteicos.

**Tabla II.B.1:** Parámetros cinéticos característicos de la hidrólisis de péptidos sintéticos de secuencias correspondientes a los sitios de corte en los precursores gag y gag/pol del virus V.I.H.-1.\*

Unión	V.I.H.-1	Km (mM)	kcat (s <sup>-1</sup> )	kcat/Km (s <sup>-1</sup> ·mM <sup>-1</sup> )
p17/p24	Val-Ser-Gln-Asn-Tyr-/Pro-Ile-Val-Gln	0.15	6.8	45
p24/X	Lys-Ala-Arg-Val-Leu-/Ala-Glu-Ala-Met-Ser	0.01	<0.1	90
X/p9	Thr-Ala-Thr-Ile-Met-/Met-Gln-Arg-Gly -Asn	0.05	3.7	74
p9/p6	Arg-Pro-Gln-Asn-Phe-/Leu-Gln-Ser-Arg-Pro	0.53	0.3	1
p15/PR	Val-Ser-Phe-Asn-Phe-/Pro-Gln-Ile-Thr-Leu	<0.01	<0.1	7
PR/RT	Cys-Thr-Leu-Asn-Phe-/Pro-Ile-Ser-Pro	0.07	1.5	24
Rt51-RNasa	Ala-Glu-Thr-Phe-/Tyr-Val-Asp-Gly-Ala-Ala	0.04	0.4	10
RT-IN	Ile-Arg-Lys-Ile-Leu-/Phe-Leu-Asp-Gly	0.01	1.2	200

\* Reproducidos a partir de Tozser et al. (Tozser et al., 1991b). Los ensayos se realizaron a pH 5.6, 0.25 M de fosfato potásico, 7.5% de glicerol, 5mM de ditiotreitól, 1mM de EDTA, 0.2% de Nonidet P-40 y 2M de NaCl.

La especificidad en cada uno de los subsitios ha sido investigada por varios grupos mediante el análisis de la variación de los parámetros cinéticos del proceso de hidrólisis como

consecuencia de la sustitución sistemática de los aminoácidos del sustrato en diversas posiciones (Dunn et al., 1994) tanto para los sustratos peptídicos que contienen prolina (sustituciones en P1 (Billich & Winkler, 1991), P2 (Billich & Winkler, 1991; Tozser et al., 1991b; Tozser et al., 1992) y P2' (Margolin et al., 1990), P3 (Billich & Winkler, 1991), P4 (Griffiths et al., 1992b; Tozser et al., 1991b; Tozser et al., 1992) y P5 (Griffiths et al., 1992b)) como para los pertenecientes a la segunda clase (sustituciones en P1 (Grinde et al., 1992b; Richards et al., 1990), P2 y P2' (Grinde et al., 1992b; Konvalinka et al., 1990; Phylip et al., 1990), P3 y P3' (Grinde et al., 1992b; Konvalinka et al., 1990) y P4 (Grinde et al., 1992b)).

En general, la hidrólisis de estos sustratos está caracterizada por valores relativamente elevados de Km, que en algunos casos alcanzan el intervalo milimolar, lo que, en cierto modo, dificulta los estudios cinéticos. Los resultados de este tipo de estudios sugieren que la especificidad de esta proteína es un proceso complejo, en el que se combina la tendencia a la formación del máximo número de enlaces de hidrógeno posible entre el esqueleto peptídico del sustrato y la proteína, con la optimización de las interacciones entre las cadenas laterales con los distintos subsitios de unión. Mientras que una de las caras del sustrato interviene en el establecimiento de enlaces de hidrógeno con la base del sitio de unión, caracterizada por su elevada rigidez, la otra interacciona con las regiones más flexibles de la proteína, siendo estas últimas generalmente las de mayor relevancia para la adaptación del sustrato al sitio de unión así como para la determinación de la velocidad de hidrólisis. Por lo tanto, las diferencias observadas en los valores de los cocientes kcat/Km para los distintos sustratos reflejan la habilidad de cada uno de ellos para situar sus cadenas laterales en posiciones lo más cercanas posible a las ideales sin interferir en el establecimiento de la red de puentes de hidrógeno decisiva para la catálisis.

Como ilustración, en la tabla II.B.1 se recogen los valores de los parámetros catalíticos obtenidos experimentalmente para algunos de estos sustratos. En líneas generales, cabe destacar la incapacidad de la proteasa para hidrolizar sustratos que presenten aminoácidos ramificados en el carbono beta en posición P1; la preferencia por restos de marcado carácter hidrofóbico y preferentemente ramificados en posición P2; el carácter polar del subsitio S2' (habiéndose encontrado que Glu es una de las mejores opciones en dicha posición); así como el papel desempeñado por los residuos periféricos en posiciones P4 o incluso P5, que, aunque relativamente lejanos al sitio catalítico, pueden afectar significativamente el valor de los parámetros cinéticos.

#### *b.-Sustratos proteicos.*

Como se ha mencionado anteriormente, el empleo de diversas proteínas no víricas como sustratos de la enzima puede proporcionar información complementaria a la obtenida a partir de los estudios con pequeños péptidos concerniente a la influencia de la estructura del sustrato en la especificidad de la proteína. En general, las proteínas hidrolizadas por la proteasa del virus

V.I.H.-1 (Hui et al., 1990; Shoeman et al., 1990; Tomasselli et al., 1991; Tomasselli et al., 1990a) están en su mayoría constituidas por varios dominios, de modo que la hidrólisis se lleva a cabo fundamentalmente en los lazos que los conectan, por lo que se ha especulado acerca de la posible semejanza estructural entre estas regiones y los puntos de corte en los precursores poliprotéicos.

En principio, estos estudios han revelado que la proteasa del virus V.I.H.-1 es capaz de hidrolizar enlaces peptídicos entre cualquier tipo de aminoácidos (Poorman et al., 1991), siendo difícil establecer una clasificación dada la ausencia de preferencias notables, con la única excepción de la elevada frecuencia con que se observa un resto Glu en posición P2'. Asimismo, se ha establecido cierto paralelismo (tanto en cuanto los parámetros cinéticos como en cuanto a la especificidad) entre la hidrólisis de péptidos y proteínas, lo que sugiere que, aunque las constantes de velocidad pueden verse influidas por el contexto estructural, es posible que una región suficientemente flexible de la proteína pueda ser acomodada en el sitio de unión de la enzima en conformación extendida.

#### **II.B.4.3.2.-Influencia del pH y fuerza iónica.**

##### **a.-Parámetros cinéticos del proceso de hidrólisis.**

Como es común a todas las proteasas aspárticas (Pearl & Blundell, 1984), el centro catalítico de la proteasa del virus V.I.H.-1 está constituido por los dos grupos carboxílicos pertenecientes a los Asp<sup>25</sup> de cada uno de los monómeros que, dispuestos coplanarmente entre sí, mantienen una molécula de agua coordinada entre ellos (Wlodawer & Erickson, 1993). Como se ha discutido anteriormente, la proximidad espacial entre los grupos carboxilo, forzada por la geometría y rigidez del sitio activo, provoca anomalías en sus pKas, de forma que el correspondiente a la primera protonación presenta un valor menor que el habitual para compuestos monocarboxílicos completamente solvatados, mientras que el segundo es notablemente más elevado (Polgar, 1987).

Conviene hacer notar que, dada la simetría de la enzima, ambos carboxilos son completamente indistinguibles en la forma libre, por lo que no cabe atribuir ninguno de los pKas a un grupo en particular sino que es preferible considerar el sistema como un ácido dicarboxílico simétrico, situación en cierto modo diferente a la encontrada en las enzimas eucarióticas, en las que, a pesar de la similitud estructural entre ambos aspárticos, pequeñas diferencias en los ambientes e interacciones establecidas por cada uno de ellos, permitía observar la protonación preferencial del Asp<sup>25</sup> (numeración correspondiente a Pepsina).

A pesar de seguir el comportamiento general característico de todas las proteasas aspárticas (Lin et al., 1992), reflejado en la forma acampanada de las curvas de actividad específica (kcat/Km) en función del pH (Hyland et al., 1991a; Ido et al., 1991; Polgar et al., 1994), las enzimas retrovíricas, al contrario que sus homólogas eucarióticas (que presentan máxima actividad a pH comprendidos entre 2 y 4) poseen un pH óptimo en el rango de pH entre 4 y 6.



Este fenómeno ha sido atribuido por algunos autores a la presencia de los restos de Ala<sup>28</sup>, en posiciones topológicamente equivalentes a las ocupadas casi invariablemente por Ser y Thr en las proteasas eucarióticas, de forma que los enlaces de hidrógeno establecidos entre sus grupos hidroxilo y ambos aspárticos catalíticos (Abad-Zapatero et al., 1990; Cooper et al., 1990) podrían ser los responsables de estas diferencias (Ido et al., 1991).

Del ajuste de las curvas de actividad frente al pH puede obtenerse una estimación de los pKas de los grupos ionizables que intervienen en la catálisis, generalmente asignados a los aspárticos del centro activo. Aunque distintos estudios proporcionan valores de pKa ligeramente diferentes (Hyland et al., 1991a; Ido et al., 1991; Polgar et al., 1994), todos coinciden en la existencia de una carga negativa en el sitio activo, de modo que el sustrato se une a la enzima con sólo uno de los aspárticos protonados. Asimismo, la presencia de grupos ionizables en el sustrato, como es el caso del Glu frecuentemente encontrado en posición P2' (Poorman et al., 1991), puede afectar los parámetros catalíticos aunque no modifica sensiblemente la forma general de la curva (Polgar et al., 1994).

En cuanto a la variación de los valores individuales de  $k_{cat}$  y  $K_m$  con el pH, aunque con ciertos sustratos el valor de  $K_m$  parece seguir la ionización de los aspárticos catalíticos (presentando curvas de formas similares a las obtenidas para la velocidad específica así como valores de pKas congruentes para los grupos ionizables (Polgar et al., 1994)), los resultados obtenidos parecen depender de la naturaleza del sustrato y de las condiciones empleadas en el ensayo.

La actividad específica de la enzima se ve incrementada asimismo a alta fuerza iónica (Cheng et al., 1990; Hyland et al., 1991a; Polgar et al., 1994; Szeltner & Polgar, 1996) habiéndose encontrado (tras analizar la dependencia con el pH a diversas concentraciones salinas y haber determinado la escasa variación de los pKas de los grupos ionizables con la fuerza iónica) una variación lineal de los valores, previamente corregidos por efectos de pH, del cociente  $k_{cat}/K_m$  con la concentración de sal (Polgar et al., 1994). Aunque algunos autores sugieren la disminución de la constante de Michaelis a altas concentraciones salinas como el origen del aumento en la constante de velocidad específica,  $k_{cat}/K_m$  (Meek et al., 1994; Wondrak et al., 1991), otros proponen la contribución de factores adicionales como la estabilización conformacional de la enzima a alta fuerza iónica (Szeltner & Polgar, 1996).

#### *b.-pKa de los aspárticos catalíticos en los complejos con distintos inhibidores.*

Diversos estudios de RMN llevados a cabo con varios complejos entre la proteína marcada con <sup>13</sup>C y distintos inhibidores han permitido la estimación de los pKas de los grupos aspárticos catalíticos en la forma complejada de la enzima, ya que el desplazamiento químico correspondiente al carbono cuaternario de sus carboxilos es sensible a su estado de ionización, existiendo una diferencia de unas 4 ppm entre sus formas completamente protonadas y



desprotonadas en agua (Smith et al., 1996).

Así, el espectro correspondiente al complejo de la enzima con Pepstatina A muestra dos señales a 178.8 ppm (correspondiente a un grupo carboxilo protonado) y 172.40 ppm (originada por un grupo carboxilato), que permanecen inalteradas en el rango de pH entre 2.5 y 6.5 durante varios días. Dado que la constante de velocidad para el proceso de disociación es del orden de  $1 \text{ s}^{-1}$ , la ausencia de titulación no puede atribuirse a la inaccesibilidad de los grupos carboxilo al disolvente. Por tanto, estos datos revelan la existencia de un grupo carboxílico con un pKa superior a 6.5 y mientras que el segundo presenta un pKa en el complejo por debajo de 2.5 (Smith et al., 1996).

Idénticos resultados se han obtenido a partir de estudios similares con el inhibidor KNI-272 (Wang et al., 1996), compuesto peptidomimético que constituye un análogo del estado de transición, de forma que la asimetría introducida en la estructura por la presencia del inhibidor permite distinguir las señales correspondientes a los dos monómeros, habiéndose establecido que es el Asp<sup>25</sup> el que se encuentra protonado mientras que el Asp<sup>125</sup> está ionizado en el complejo, de acuerdo con las predicciones realizadas a partir de cálculos mecanocuánticos aplicados a la estructura tridimensional del complejo (Baldwin et al., 1995a).

Diversos estudios teóricos sugieren que los estados de ionización de los grupos aspárticos catalíticos a un pH dado deben encontrarse modulados por las interacciones específicas establecidas entre las cadenas laterales de los grupos aspárticos así como por el ambiente local en que éstos se encuentren en el complejo (Harte & Beveridge, 1993). Esta situación se refleja claramente en los valores de los pKas obtenidos para el complejo con el inhibidor simétrico, de estructura basada en ureas cíclicas y por tanto de naturaleza no peptídica, DMP323 (Lam et al., 1994). La naturaleza simétrica del inhibidor hace que ambas cadenas sean indistinguibles y que, por lo tanto, se obtengan idénticos desplazamientos químicos para los residuos equivalentes en los dos monómeros. De este modo, se observa que ambos presentan un pKa alrededor de 8, por lo que se encuentran protonados en el rango de pH entre 2 y 7 e interaccionando de forma idéntica con los dos grupos hidroxilo del inhibidor dispuestos simétricamente (Yamakazi et al., 1994).

#### *II.B.4.3.3.-Inhibición por los productos de la hidrólisis*

El estudio de los parámetros cinéticos del proceso de hidrólisis de diversos sustratos peptídicos correspondientes a algunas de las secuencias de corte de los precursores poliproteicos gag y gag-pol ha revelado la existencia de procesos de inhibición a cargo de los productos de la hidrólisis, con constantes del orden milimolar y mecanismos que varían según la naturaleza del sustrato, de modo que mientras en algunos casos los dos productos poseen capacidad inhibitoria en otros únicamente uno de ellos se une a la enzima (Hyland et al., 1991b).

Actualmente se dispone de la estructura tridimensional de los complejos de la protefina con los productos de la hidrólisis del sustrato Ac-S-L-N-F=P-I-V-NH<sub>2</sub> (Rose et al., 1996a), donde =

representa el enlace peptídico que sufre el proceso de hidrólisis. El proceso de cristalización llevado a cabo en presencia de los dos productos peptídicos resultado de la hidrólisis (Ac-S-L-N-F y P-I-V-NH<sub>2</sub>) ha permitido obtener la estructura del complejo de la enzima con el péptido correspondiente al extremo amino-terminal, que adopta una conformación muy similar a la presentada por la región correspondiente en los distintos inhibidores peptidomiméticos. Estos resultados están en concordancia con los obtenidos a partir de estudios cinéticos llevados a cabo con un sustrato de secuencia similar, Ac-S-Q-L-Y=P-V-V-NH<sub>2</sub>, en los que únicamente se observa inhibición por parte del tetrapéptido amino-terminal Ac-S-Q-L-Y, caracterizado por una constante de inhibición de 0.23 mM (Hyland et al., 1991b).

La cristalización en presencia exclusivamente del péptido Ac-S-L-N-F conduce a la obtención de una estructura similar a la obtenida en presencia de ambos productos, mientras que al llevarse a cabo el proceso de cristalización en presencia del sustrato completo, es el tripéptido correspondiente al extremo carboxilo terminal (P-I-V-NH<sub>2</sub>) el encontrado en el sitio de unión. En este caso, la orientación del péptido es opuesta a la que presenta en el sustrato, a pesar de lo cual, su cadena principal establece tres de las cuatro interacciones descritas para la estructura del extremo amino-terminal de los inhibidores peptídicos. En ninguno de los casos se han encontrado ambos productos unidos simultáneamente a la enzima.

#### **II.B.4.4.-Estudios de inhibición.**

Dada la conservación del mecanismo catalítico entre todas las enzimas pertenecientes a la familia de las proteasas aspárticas, la clasificación de la proteasa del virus V.I.H.-1 como miembro de este grupo ha permitido la aplicación de estrategias similares a las utilizadas con las otras enzimas de la familia (especialmente elaboradas en el caso de Renina) para el desarrollo de inhibidores efectivos de esta proteasa, posibilitando, por tanto, un avance significativo en el tratamiento terapéutico de la enfermedad.

La similitud en el mecanismo de acción constituye a su vez un claro obstáculo en el desarrollo de drogas efectivas contra el virus, dado que la ausencia de interferencias con la actividad de las proteasas aspárticas humanas como Pepsina, Renina, Cathepsina D o Cathepsina E, constituye un requisito esencial para la posible aplicación terapéutica de cualquier inhibidor (Jupp et al., 1989). A este respecto, la capacidad de llevar a cabo la hidrólisis de enlaces peptídicos entre restos Tyr-Pro, constituye una característica diferencial de la proteasa del virus V.I.H con respecto a las demás proteínas del grupo que, en un principio, proporciona un punto de partida para el desarrollo de inhibidores específicos para esta enzima.

##### **II.B.4.4.1.-Inhibidores peptídicos.**

Como se ha mencionado anteriormente, una de las estrategias de inhibición más

comúnmente adoptadas, y que ha resultado notablemente fructífera en el desarrollo de inhibidores de otras proteasas aspárticas, consiste en la identificación de un sustrato oligopeptídico adecuado en el que incorporar las funcionalidades apropiadas que simulen el estado de transición. De este modo, la exploración de la secuencia y longitud óptimas del sustrato se convierte en una tarea preliminar esencial (Debouck & Metcalf, 1990).

La efectividad obtenida con los distintos análogos del estado de transición es variable (Billich et al., 1988; Tomasselli et al., 1990b), siendo, en general los derivados hidroxietilénicos [CH(OH)-CH<sub>2</sub>] los inhibidores de mayor potencia (Dreyer et al., 1992; Jaskolksi et al., 1991, Thanki et al., 1992; Thompson et al., 1992), seguidos por hidroxietilaminas [CH(OH)-CH<sub>2</sub>-N-] (Rich et al., 1990; Roberts et al., 1990); difluorostatonas [CO-CF<sub>2</sub>-CO] (Schirling et al., 1993); los inhibidores que contienen restos de estatina [CH(OH)-CH<sub>2</sub>-CO]; fosfinatos [P(O<sub>2</sub>)H-CH<sub>2</sub>]; y finalmente las amidas reducidas [CH<sub>2</sub>-NH] (Miller et al., 1989a).

#### *a.-Inhibidores basados en sustratos peptídicos.*

Como se ha mencionado anteriormente, la proteasa del virus V.I.H.-1 es responsable de la hidrolización específica de ocho enlaces diferentes en los precursores gag y gag-pol (Billich et al., 1988; Darke et al., 1988; Krausslich et al., 1989) entre los que cabe destacar los sitios de corte 1,5, y 6 en los que la hidrólisis se produce entre los residuos Phe o Tyr y Pro (ver tabla II.B.1). El hecho de que estos enlaces no sean susceptibles de ser hidrolizados por las otras proteasas humanas, proporcionaría a cualquier inhibidor basado en este tipo de sustratos una elevada especificidad así como mayor resistencia a la degradación.

De este modo, se han desarrollado diversos inhibidores mediante la incorporación de grupos análogos del estado de transición a este tipo de sustratos (Martin, 1992; Rich et al., 1990; Roberts et al., 1990). Diversos estudios comparativos indican que, en general, son los derivados hidroxietilénicos los que originan inhibidores de mayor potencia, posiblemente debido a su mayor semejanza con el enlace peptídico encontrado en los sustratos (Dreyer et al., 1989; Moore et al., 1989; Vacca, 1994). En cuanto a la configuración adoptada por el grupo hidroxilo no existe un comportamiento homogéneo encontrándose tanto compuestos con configuración R o S. Se ha observado mediante estudios cristalográficos y de modelado molecular, que el par electrónico libre del nitrógeno del resto de prolina puede encontrarse orientado tanto hacia la base del centro activo, como hacia el "flap", de modo que, en este caso, es posible conseguir el alineamiento correcto de las cadenas laterales en el sitio de unión con independencia de la estereoquímica del hidroxilo (Krohn et al., 1991; Rich et al., 1991).

#### *b.-Inhibidores de otras proteasas aspárticas.*

Asimismo, la identificación de la proteasa del virus V.I.H.-1 como proteasa aspártica ha permitido la utilización de diversos inhibidores desarrollados en un principio para otras enzimas

del grupo como Pepsina o Renina (Kato et al., 1987; Richards et al., 1989b). De este modo, al igual que las demás enzimas del grupo, el pentapéptido acetil-pepstatina constituye un potente inhibidor de la proteasa del virus de tipo 1 (Richards et al., 1989b), e incluso presenta mayor potencia frente a la proteasa de tipo 2, siendo las constantes de inhibición 20 nM y 5nM respectivamente. La conformación adoptada por el pentapéptido en el complejo (Fitzgerald et al., 1990) es muy similar a la encontrada para otros inhibidores, aún a pesar de la ausencia de cadena lateral en P1'.

Asimismo, la semejanza encontrada entre la secuencia de corte en el pepsinógeno y uno de los sitios de hidrólisis de los precursores poliproteicos, ha permitido el desarrollo de inhibidores de gran potencia (Martin, 1992; McQuade et al., 1990; Vacca et al., 1991) basados en compuestos desarrollados específicamente como inhibidores de Renina.

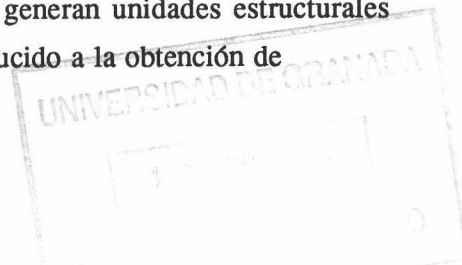
#### II.B.4.4.2.-Inhibidores no peptídicos.

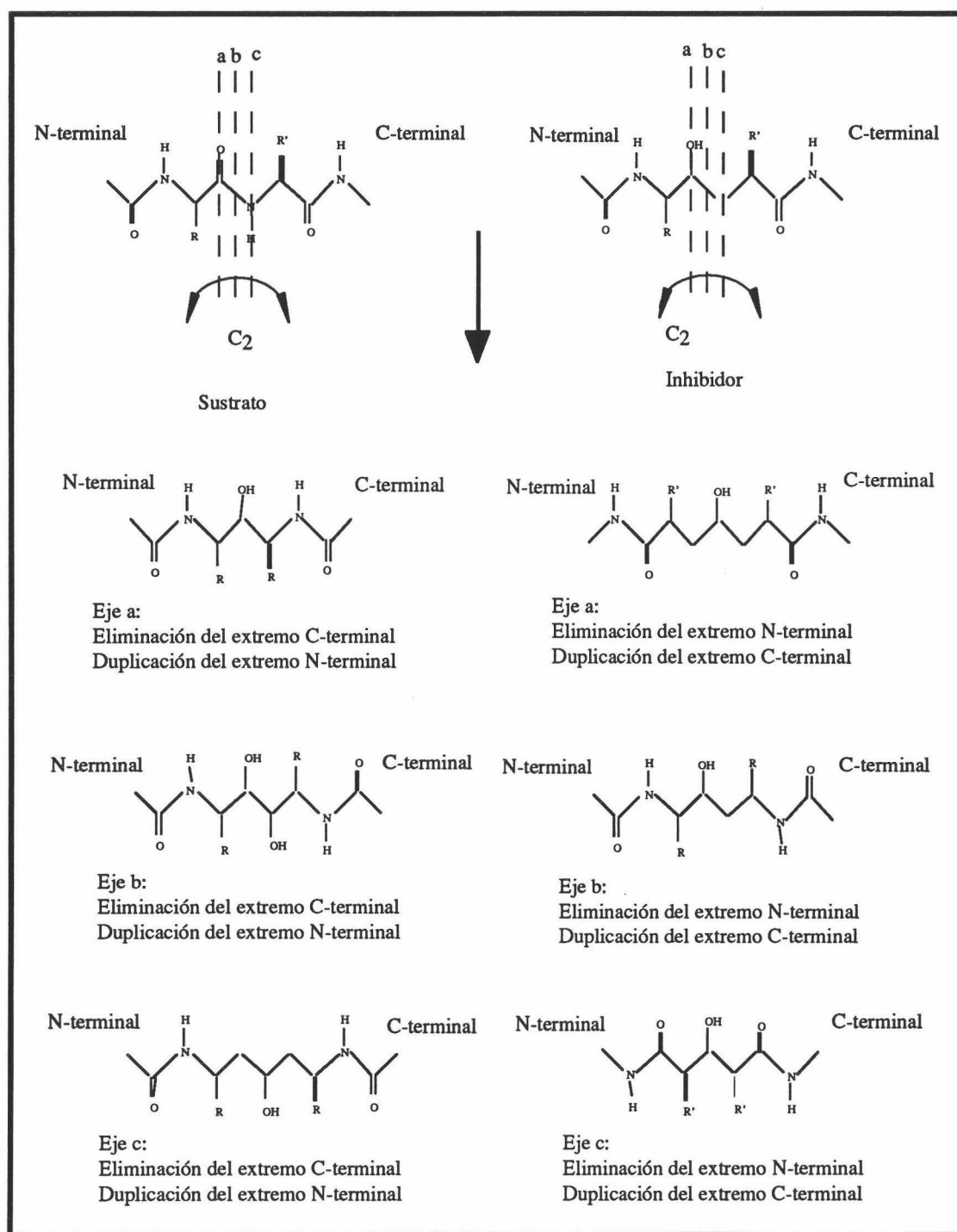
##### a.-Inhibidores simétricos.

La confirmación cristalográfica de la naturaleza homodimérica de las proteasas retrovíricas que, como se ha descrito anteriormente, poseen un único sitio activo de simetría binaria, ha proporcionado un nuevo impulso al desarrollo de inhibidores de naturaleza simétrica. Estos compuestos, de menor carácter peptídico, aunque todavía basados en el concepto clásico del estado de transición, constituyen potentes inhibidores de la enzima que, además de presentar una mayor estabilidad *in vivo*, se caracterizan por una mayor especificidad por las proteasas retrovíricas dada la falta de simetría en el centro activo de las enzimas eucarióticas.

El proceso de diseño de un inhibidor simétrico a partir de un sustrato o inhibidor asimétrico conlleva la elección de una de las porciones (amino o carboxilo terminal) de los compuestos originales y la posterior duplicación de su estructura mediante la aplicación de una operación de simetría  $C_2$ , de modo que se origina una estructura simétrica complementaria a las regiones hidrofóbicas e hidrofílicas de ambas mitades del sitio activo. Así, tras la definición en la estructura original de un eje de simetría imaginario coincidente con el de la enzima y la eliminación de uno de los extremos, la duplicación, mediante el giro alrededor de dicho eje, de la región conservada genera un nuevo inhibidor simétrico (Erickson, 1993, Kempf, 1994, Kempf et al., 1990). En la figura II.B.8 se ilustran algunos de los procedimientos seguidos para el desarrollo de estos inhibidores.

Dependiendo de la elección del eje de simetría (que puede pasar por el centro del enlace peptídico hidrolizable o coincidir con el carbono carbonílico o con el nitrógeno amida), la elección de la región duplicada, así como de la naturaleza de la función central (Barrish et al., 1994, Bone et al., 1991, Erickson, 1993, Wang et al., 1995), se generan unidades estructurales estables de diversas características. Estos procedimientos han conducido a la obtención de





**Figura II.B.8:** Procedimiento seguido para la obtención de inhibidores simétricos a partir de sustratos e inhibidores peptídicos.

compuestos caracterizados por constantes de inhibición en el intervalo nanomolar (Kempf et al.,



1995; Kempf et al., 1991). En general, la potencia de estos compuestos se encuentra determinada tanto por su habilidad para establecer enlaces de hidrógeno con el centro catalítico como por el espaciado adecuado y la correcta disposición de los grupos carbonílicos y de las cadenas laterales en posición P1 y P1' (Kempf, 1994).

En general, la observación de las estructuras tridimensionales (Bone et al., 1991; Erickson, 1993; Erickson et al., 1990; Kempf et al., 1993) de los complejos de estos compuestos con la enzima muestra que, aún a pesar de la orientación inversa de una de sus mitades en relación con los inhibidores basados en sustratos, existe una gran similitud en cuanto a los puentes de hidrógeno establecidos con los grupos del centro activo. Esta observación es, en cierto modo, indicativa del papel fundamental que la habilidad de los compuestos asimétricos para adoptar conformaciones "simétricas" tiene en la determinación de la afinidad. El grupo hidroxilo central del inhibidor interacciona con ambos aspárticos catalíticos; los dos grupos carbonílicos adyacentes establecen un puente de hidrógeno con la molécula de agua que actúa como puente entre el inhibidor y el "flap", mientras que las aminas a ambos lados de la unidad central interaccionan con las Gly<sup>27</sup> de ambos monómeros.

#### *b.-Diseño de novo de inhibidores no peptídicos.*

La elucidación de la estructura tridimensional de la proteasa del virus V.I.H.-1, así como de sus complejos con un elevado número de inhibidores, ha permitido profundizar en la comprensión de las relaciones estructura-función que determinan tanto la especificidad por los sustratos como la afinidad y potencia de los distintos inhibidores. De este modo, el diseño de nuevos compuestos con grupos funcionales específicos relacionados espacialmente entre sí de modo que presenten cierta complementariedad con las superficies hidrofóbicas e hidrofílicas de la proteína ha permitido el desarrollo de una nueva generación de inhibidores. Por lo tanto, mediante la maximización de las interacciones de van der Waals y la optimización del número y geometría de los enlaces de hidrógeno establecidos entre el compuesto diseñado y la enzima se han conseguido inhibidores de gran potencia, especificidad y estabilidad a la degradación, que constituyen una sólida alternativa terapéutica.

En general, una de las estrategias más comúnmente empleadas es el diseño de compuestos hidrofóbicos, preferentemente de pequeño tamaño y alto contenido en ciclos aromáticos, de forma que su elevada rigidez contribuye al incremento de la afinidad gracias a la minimización de la pérdida de entropía conformacional como consecuencia del proceso de unión. Adicionalmente, la incorporación de grupos específicos en situaciones estratégicas que permitan el desplazamiento de moléculas de agua adicionales que de otro modo estarían presentes en el complejo con el consecuente beneficio entrópico, o bien que contribuyan simplemente al incremento de su solubilidad, constituyen técnicas comúnmente utilizadas.

La naturaleza de estos inhibidores es, por tanto, muy variada, destacando los compuestos



macrocíclicos como cosalanos (Cushman et al., 1994; Cushman et al., 1995; Xie et al., 1995), cumarinas (Thaisrivongs et al., 1995), piranos (Lunney et al., 1994; Romines et al., 1995a; Skulnick et al., 1997; Vara Prasad et al., 1995), ureas cíclicas (Lam et al., 1994; Nugiel et al., 1997; Sham et al., 1996), derivados de penicilina y sulfonamidas (Backbro et al., 1997; Kitchin et al., 1994; Skulnick et al., 1995), y otros compuestos cíclicos (Ghosh et al., 1994; Hulten et al., 1997; Podlogar et al., 1994; Romines et al., 1995b).

#### **II.B.4.5.-Desarrollo de resistencia a fármacos.**

Uno de los principales obstáculos encontrados en el proceso de desarrollo de tratamientos efectivos contra el S.I.D.A. ha venido determinado por la aparición, durante la administración de los diversos agentes terapéuticos disponibles, de nuevas cepas del virus que presentan resistencia a muchos de estos fármacos (Condra et al., 1995). Este fenómeno, aunque observado inicialmente con compuestos como AZT o nevirarpine, dirigidos contra la retrotranscriptasa, ha adquirido asimismo gran importancia en el desarrollo de inhibidores de la proteasa (Ridky & Leis, 1995).

La aparición de mutaciones en el virus que confieren resistencia a los diversos fármacos parece tener su origen en la baja fidelidad de la retrotranscriptasa, cuyos errores, no eliminados, dada la inexistencia de capacidad correctora en la polimerasa de ADN del virus, se incorporan al ADN vírico con anterioridad a la integración en el genoma de la célula huésped. Como consecuencia, en presencia de drogas que seleccionan preferentemente ciertos tipos de mutaciones, este proceso, en principio de naturaleza completamente aleatoria, se traduce en la aparición de cepas resistentes (Condra et al., 1995; Ridky & Leis, 1995).

##### **II.B.4.5.1.-Localización y naturaleza de las mutaciones que confieren resistencia.**

En el caso de la proteasa las mutaciones encontradas en los virus resistentes aislados en pacientes bajo tratamiento con sus inhibidores constituyen un juego de aminoácidos muy característico. Estas mutaciones se encuentran con frecuencia localizadas en zonas directamente relacionadas con los distintos subsitios de unión (Cameron et al., 1993; Collins et al., 1995) aunque se han descrito ciertas mutaciones en posiciones no implicadas directamente en el proceso de unión de los sustratos e inhibidores.

Algunas de las mutaciones más frecuentes se encuentran situadas en la base del centro activo, como es el caso de los aminoácidos en posición 32 (implicado en los subsitios S2 y S2'), 8 (subsitios S4 y S4') y 84 (subsitios S1, S3, S3', S1'). Otras, por el contrario, forman parte de los lazos externos, altamente flexibles, como las mutaciones en posición 46 (Ho et al., 1994; Kaplan et al., 1994), 47 y 48 que forman parte del flap que, tras la unión del sustrato a la enzima, cubre el sitio catalítico, constituyendo la parte superior de los subsitios S4 y S4'.

La mayoría de las cadenas laterales de los residuos en estas posiciones se encuentran orientadas hacia el interior de la molécula, de modo que la substitución de un aminoácido por

otro se traduce en la alteración de sus interacciones con el ligando. Existen, sin embargo, ciertas excepciones, como es el caso de Met<sup>46</sup> cuya cadena lateral se encuentra dirigida hacia el disolvente, por lo que el efecto de una posible mutación en esta posición no es trivial, pudiendo estar relacionado, según se desprende de diversos estudios de dinámica molecular, con la disminución de la flexibilidad del "flap" (Collins et al., 1995; Ho et al., 1994).

Los distintos subsitios de unión muestran diferentes preferencias por los distintos tipos de aminoácidos determinadas fundamentalmente por la maximización de las interacciones, en su mayoría de van der Waals, establecidas con las cadenas laterales del inhibidor (Cameron et al., 1993; Grinde et al., 1992a), mostrando generalmente un comportamiento independiente entre sí (Cameron et al., 1994; Grinde et al., 1992b). Dado que la mayoría de los inhibidores han sido diseñados con el fin de maximizar las interacciones a lo largo del sitio de unión, posibles modificaciones de los distintos subsitios pueden afectar profundamente su afinidad por la proteína. La aparición de mutaciones, por tanto, puede conducir a una variación en la especificidad por los sustratos (Cameron et al., 1994; Grinde et al., 1992b) con la consecuente disminución en la capacidad catalítica de la enzima (Gulnik et al., 1995), así como a una menor afinidad de los inhibidores por la proteína.

Las mutaciones encontradas en sitios importantes para el establecimiento de la especificidad son generalmente muy conservativas, implicando normalmente la pérdida o ganancia de un grupo metileno, como es el caso de las mutaciones I84V, V82I, V82A, I47V, V32I, siendo M46F una de las pocas excepciones, mientras que las situadas regiones de la proteína alejadas del centro activo muestran cambios más dramáticos.

En general, tras el tratamiento con diversas drogas, los mutantes aislados en los primeros estadios del tratamiento presentan eficiencias catalíticas mermadas con respecto a la proteína nativa (Gulnik et al., 1995), mientras que, en tratamientos prolongados se ha observado la aparición de nuevos cambios en la secuencia, inoocuos en cuanto a resistencia se refiere, pero que, en cierto modo, compensan el efecto de las mutaciones originales, restaurando en gran medida la capacidad catalítica de la proteína.

#### **II.B.4.5.2.-Estudios cinéticos y estructurales.**

Conviene hacer notar que una condición necesaria para la aparición de cepas resistentes es el desarrollo de mutaciones con capacidad para disminuir la afinidad de la proteína por el inhibidor, manteniendo su actividad catalítica a niveles aceptables para la supervivencia del virus. De esta premisa se desprende la necesaria existencia de diferencias fundamentales, bien en el número, bien en la naturaleza de las interacciones que determinan la afinidad de unión de inhibidores y sustratos. Dicha modulación diferencial de las propiedades enzimáticas de la proteína no es más que el reflejo de la relaciones estructura/función en los distintos mutantes, por lo que varios laboratorios han abordado tanto el estudio de sus propiedades cinéticas (Gulnik et

al., 1995; Lin et al., 1995), como de sus estructuras cristalográficas (Baldwin et al., 1995b; Chen et al., 1995; Hong et al., 1996; Hoog et al., 1996; Skulnick et al., 1997; Skulnick et al., 1995).

La posibilidad de que cualquier mutación confiera resistencia a un inhibidor en concreto no está determinada únicamente por la disminución en la sensibilidad de la enzima a dicho inhibidor ( $K_{i, \text{mut}}/K_{i, \text{WT}}$ ), sino que su capacidad para mantener un nivel de actividad catalítica suficiente para la supervivencia del virus es asimismo un factor crítico. De este modo, dadas dos enzimas mutantes que presentan la misma afinidad por un inhibidor en concreto, aquella que posea mayor eficiencia catalítica ( $k_{\text{cat}}/K_m$ ) será seleccionada preferentemente en presencia de dicho inhibidor.

Así, la definición de la “vitalidad” de una mutación de acuerdo con la expresión:

$$\text{vitalidad} = \frac{(K_i k_{\text{cat}} / K_m)_{\text{mutante}}}{(K_i k_{\text{cat}} / K_m)_{\text{silvestre}}} \quad (\text{II.B.1})$$

permite la comparación los efectos de las diversas mutaciones en presencia de un inhibidor en particular. De este modo, que mientras mayor sea la vitalidad de una mutación, mayores son las posibilidades de que ésta sea seleccionada bajo la presión de un inhibidor determinado (Gulnik et al., 1995).

En términos generales, la mayoría de los mutantes presentan una menor eficiencia catalítica ( $k_{\text{cat}}/K_m$ ) que la proteína silvestre debido fundamentalmente al incremento en  $K_m$ , mientras que la constante catalítica no se ve drásticamente afectada por la mayoría de las mutaciones (Gulnik et al., 1995; Lin et al., 1995). En cuanto a la interacción con los inhibidores, se ha observado que el efecto de una mutación concreta en la constante de inhibición ( $K_i$ ) depende de su situación en la molécula, variando asimismo con la naturaleza de la substitución. Adicionalmente, la disminución originada en la afinidad por cada mutación varía asimismo de un inhibidor a otro, estando determinada por la estructura del inhibidor así como por los posibles cambios conformacionales inducidos en la proteína a consecuencia de la mutación (Gulnik et al., 1995; Lin et al., 1995).

A pesar de que la información estructural disponible actualmente es aún muy limitada, se dispone de las estructuras tridimensionales de los complejos de algunos inhibidores con enzimas en las cuales se han incorporado algunas de las mutaciones resistentes más frecuentemente observadas *in vivo* (Baldwin et al., 1995b; Chen et al., 1995; Hong et al., 1996). El análisis de estas estructuras no ha revelado grandes cambios estructurales en el complejo a consecuencia de la mutación, por lo que, aunque algunos autores han atribuido la disminución de la afinidad a la pérdida o modificación de algunas de las interacciones concretas establecidas entre la enzima y el inhibidor (Chen et al., 1995; Hong et al., 1996), otros sugieren la existencia de cambios conformacionales muy sutiles, extendidos a lo largo de todo el sitio de unión y que, globalmente, conducen a una menor sensibilidad de la proteína por el inhibidor (Baldwin et al., 1995b).

#### II.B.4.5.3.-Resistencia múltiple.

La proteasa del virus V.I.H.-1 ha demostrado una elevada facilidad para la incorporación de mutaciones en su secuencia, de modo que, en principio, el virus tiene la capacidad de desarrollar resistencia a cualquiera de los inhibidores diseñados hasta el momento, habiéndose descrito la aparición de mutantes resistentes al menos para quince de ellos (Condra et al., 1995; el-Farrash et al., 1994; Ho et al., 1994; Jacobsen et al., 1995; Kaplan et al., 1994; King et al., 1995; Markowitz et al., 1995; Moutouh et al., 1996; Otto et al., 1993; Roberts, 1995).

Dado que la mayoría de estos compuestos han sido desarrollados como análogos no hidrolizables del sustrato, específicamente diseñados para la inhibición de la proteína nativa mediante la optimización de las interacciones en el centro catalítico (Ridky & Leis, 1995; Rose et al., 1996b), no es sorprendente la aparición de fenómenos de resistencia múltiple, de modo que los fenotipos resistentes desarrollados en presencia de un fármaco determinado generalmente presentan resistencia a otros inhibidores (Condra et al., 1995; Ridky & Leis, 1995).

Aún así, se ha demostrado la existencia de ciertas diferencias en los patrones mutacionales desarrollados por distintos inhibidores, habiéndose incluso descrito la sensibilización adicional de algunos de los mutantes hacia otros fármacos con respecto a la proteína silvestre (Roberts, 1995). Por tanto, la aplicación de tratamientos combinados, que contienen varios inhibidores de distinta naturaleza, podría evitar, o al menos debilitar, el desarrollo de cepas resistentes del virus, resultando más efectivos desde el punto de vista terapéutico que los agentes individuales.

### ***III.-Metodologías de Cálculo de Energía Libre***

### **III.-METODOLOGÍAS DE CÁLCULO DE ENERGÍA LIBRE: PARAMETRIZACIÓN ESTRUCTURAL DE LA ENERGÉTICA DE PLEGAMIENTO DE PROTEÍNAS Y UNIÓN DE LIGANDOS.**

#### **III.1.-CONSIDERACIONES GENERALES.**

##### **III.1.1.-Interés del cálculo de Energías libres.**

El problema del reconocimiento molecular, tanto en procesos intermoleculares (entre los que destacan: la unión de cofactores, sustratos, e inhibidores a enzimas, la identificación de antígenos por sus correspondientes anticuerpos, la interacción entre hormonas y sus receptores, o las asociaciones entre proteínas y ácidos nucleicos) así como intramoleculares (constituyendo el plegamiento de una proteína un ejemplo especialmente notable) se ha convertido en una cuestión de extraordinaria relevancia. El gran interés despertado por este tipo de fenómenos, que ha sobrepasado los límites estrictamente académicos, se debe fundamentalmente a sus innumerables implicaciones tanto biotecnológicas como terapéuticas, ya que la comprensión de los factores que determinan el reconocimiento molecular posibilitaría el desarrollo de estrategias racionales para el diseño de proteínas con nuevas funciones y estabilidades modificadas y fármacos con afinidades predeterminadas por receptores específicos.

La descripción termodinámica de estos procesos es de gran interés para la elucidación de los factores que determinan el reconocimiento molecular. Desde un punto de vista microscópico, y de acuerdo con los principios de la termodinámica estadística, las poblaciones de cada uno de los estados accesibles a un sistema molecular en equilibrio son proporcionales a la diferencia en energía libre con respecto a un estado de referencia. De este modo, una vez conocida su energía libre, queda determinada la probabilidad de encontrar el sistema en un estado específico. Consecuentemente, la energía libre de Gibbs constituye un parámetro fundamental en la descripción termodinámica de sistemas de interés biológico, dado que controla la dirección en que se desarrollan los procesos espontáneos, la posición de los equilibrios químicos, así como la magnitud de las barreras energéticas asociadas al estado de transición en procesos cinéticos.

Asimismo, el gran desarrollo experimentado durante los últimos años por las diversas técnicas de análisis estructural se ha traducido en un notable incremento de la información estructural disponible para los distintos sistemas biológicos, y ha avivado el interés por el estudio de las relaciones existentes entre la estructura, estabilidad y función de los biopolímeros. La elucidación de estas cuestiones proporcionaría la clave, tanto para el diseño racional de fármacos, como para el desarrollo de las técnicas de ingeniería de proteínas, que requieren necesariamente la capacidad de predecir los efectos que cualquier modificación estructural tendría en la estabilidad y función del sistema. Por todas estas razones, el desarrollo de métodos que posibiliten el cálculo de energías libres en sistemas moleculares complejos a partir de sus coordenadas espaciales se ha



convertido en una de las cuestiones centrales en el campo de la química computacional (Beveridge & DiCapua, 1989; King, 1993; McCammon, 1991; Pohorille & Pratt, 1986; Reynolds et al., 1992; Straatsma & McCammon, 1992; Straatsma et al., 1993; van Gunsteren et al., 1993).

### III.1.2.-Cálculo mecánico-estadístico de magnitudes termodinámicas.

La mecánica estadística proporciona la conexión entre la descripción microscópica de un sistema en términos de las coordenadas espaciales y temporales de sus componentes y las distintas funciones termodinámicas, de carácter macroscópico. Una vez conocido el hamiltoniano del sistema,  $H(p, r)$ , que proporciona la descripción de las interacciones entre sus partículas en función de sus coordenadas generalizadas ( $r$ ) y momentos conjugados ( $p$ ), es posible, mediante la determinación de las probabilidades de cada uno de los estados accesibles al sistema y de la función de partición ( $Z$ ), calcular cualquiera de las magnitudes termodinámicas características.

De este modo, la energía interna ( $E[N, V, T]$ ), la entropía ( $S[N, V, T]$ ), o la energía libre de Helmholtz ( $F[N, V, T]$ ) de un sistema cualquiera pueden obtenerse de acuerdo con las siguientes expresiones:

$$E(N, V, T) = \int \int H(p, r) \left\{ \frac{\exp[-H(p, r)/k_B T]}{Z} \right\} dp dr = \langle H(p, r) \rangle \quad (\text{III.1})$$

$$\begin{aligned} F(N, V, T) &= -k_B T \ln \left\{ \frac{\int \int \exp[-H(p, r)/k_B T] dp dr}{h^{3N} N!} \right\} = \\ &= +k_B T \ln \langle \exp[+H(p, r)/k_B T] \rangle \end{aligned} \quad (\text{III.2})$$

$$\begin{aligned} S(N, V, T) &= -\frac{\delta F}{\delta T} = k_B \ln Z + \frac{\int \int H(p, r) \exp[-H(p, r)/k_B T] dp dr}{h^{3N} N! Z T} = \\ &= \frac{\langle H(p, r) \rangle}{T} - k_B \ln \langle \exp[+H(p, r)/k_B T] \rangle \end{aligned} \quad (\text{III.3})$$

en las que  $k_B$  representa la constante de Boltzmann,  $T$  la temperatura absoluta,  $h$  la constante de Planck,  $N$  el número de partículas,  $V$  el volumen del sistema, y  $Z$  la función de partición definida como:

$$Z(N, V, T) = \frac{\int \int \exp[-H(p, r)/k_B T] dp dr}{h^{3N} N!} \quad (\text{III.4})$$

La energía libre de Gibbs puede obtenerse de acuerdo con la siguiente expresión, que se

distingue de la correspondiente a la energía libre de Helmholtz por la presencia del término adicional PV en el Hamiltoniano y la necesidad de integración sobre todo el volumen accesible al sistema:

$$G(N, T, P) = -k_B T \ln \left\{ \frac{\int \int \int \exp [-H(p,r) + PV] / k_B T \, dp \, dr \, dV}{h^{3N} N!} \right\} \quad (\text{III.5})$$

El cálculo mecánico estadístico de las distintas magnitudes termodinámicas requiere, por tanto, la evaluación de las integrales sobre todo el espacio de fase y, por tanto, el conocimiento de los niveles energéticos del sistema; es decir, del valor del Hamiltoniano en cualquier punto del espacio conformacional accesible a éste. La aplicación de esta metodología al estudio de sistemas moleculares complejos, como disoluciones de biopolímeros, presenta, por lo tanto, dos dificultades fundamentales: la determinación de la expresión analítica del Hamiltoniano y el muestreo del espacio configuracional accesible al sistema.

### III.1.2.1.- Muestreo configuracional.

Este último problema se ha abordado frecuentemente mediante el empleo de técnicas computacionales que facilitan la exploración del espacio configuracional, siendo las simulaciones de dinámica molecular y Monte Carlo los métodos más ampliamente utilizados (Siepmann, 1993; Straatsma & McCammon, 1992). El método de Monte Carlo se basa en la generación de una amplia distribución conformacional, en la que cada conformación se pondera de acuerdo con su correspondiente exponente de Boltzmann, y a partir de la cual es posible determinar los valores medios de aquellas propiedades del sistema dependientes exclusivamente de las coordenadas. Las técnicas de dinámica molecular, por otra parte, consisten en la evaluación de las magnitudes promedio mediante el estudio de la evolución del sistema a lo largo de un cierto periodo de tiempo, por lo que este tratamiento es únicamente adecuado para sistemas ergódicos y tiempos de simulación suficientemente prolongados (Berendsen, 1991).

En ambos casos, el muestreo del espacio de fase se realiza proporcionalmente a los exponentes de Boltzmann de cada conformación, por lo que la búsqueda conformacional se lleva a cabo preferentemente en las regiones de baja energía. De este modo, mientras que el cálculo de los valores medios de las propiedades del sistema como la entalpía (en las que las principales contribuciones provienen de las conformaciones cercanas a los mínimos energéticos) constituye un proceso asequible, la determinación de las magnitudes termodinámicas de naturaleza estadística, como la entropía o las energías libres de Gibbs y Helmholtz, presentan grandes dificultades. En estos casos, y como se puede apreciar en las ecuaciones III.2 y III.3, cualquier

región del espacio configuracional contribuye a la determinación del valor de estas magnitudes, lo cual imposibilita, dada la dificultad de muestreo de las regiones de alta energía, la determinación de sus valores absolutos a partir de cualquier simulación finita (Beveridge & DiCapua, 1989; King, 1993; van Gunsteren et al., 1993).

### III.1.2.2.-Funciones de energía potencial.

En principio, la resolución de la ecuación de Schroedinger independiente del tiempo proporcionaría la información necesaria para la descripción del sistema en términos de sus niveles energéticos. En la práctica, sin embargo, el estudio de sistemas complejos, como son las macromoléculas biológicas, requiere un enfoque diferente ya que, en este caso, además de no existir una solución exacta de la ecuación de Schroedinger, los cálculos mecanocuánticos aproximados de las superficies potenciales son asimismo inasequibles. Se requiere, por tanto, el empleo de funciones potenciales empíricas, que representen de forma simplificada las interacciones entre las partículas del sistema (Cornell et al., 1991; Gelin, 1993).

Como consecuencia, la capacidad del hamiltoniano utilizado para reproducir las características de la función de energía potencial "real" es un factor determinante de la exactitud con que puede llevarse a cabo el cálculo de las distintas magnitudes termodinámicas. En el caso de la energía libre y de la entropía esta dependencia es aún más significativa, dado que, como se ha descrito en el apartado anterior, su cálculo requiere a su vez el muestreo, ponderado mediante los exponentes de Boltzmann, del espacio configuracional accesible al sistema.

La aproximación de Born-Oppenheimer, que permite la separación de los movimientos electrónicos y nucleares, constituye el fundamento de las funciones potenciales empíricas. De esta forma, dado que cualquier modificación en las posiciones de los núcleos atómicos se supone acompañada por la reorganización instantánea de los electrones, la energía molecular puede calcularse como función de las coordenadas nucleares. Como consecuencia, la energía potencial media de los distintos estados electrónicos correspondientes a una configuración nuclear cualquiera se puede representar mediante una función potencial, obtenida generalmente a partir de la parametrización de resultados experimentales y cálculos mecanocuánticos llevados a cabo con pequeñas moléculas modelo.

En general, las funciones potenciales empíricas más ampliamente utilizadas constan de diversos términos que describen las contribuciones de las interacciones covalentes y no covalentes a la energía del sistema, generalmente en función de ciertas coordenadas internas características. Aunque, en general, las distintas funciones potenciales proporcionan una descripción adecuada de las contribuciones covalentes a la energía de un sistema molecular determinado, el tratamiento de las interacciones no covalentes presenta enormes dificultades. Esto se traduce en la existencia de grandes limitaciones en cuanto a la aplicabilidad de estas funciones al estudio de macromoléculas biológicas, cuyas estructuras tridimensionales, funciones, y propiedades dinámicas se encuentran

fundamentalmente determinadas por las interacciones no covalentes, establecidas tanto intra como intermolecularmente, y entre las que son particularmente importantes aquellas en las que intervienen las moléculas de disolvente. La dificultad de incorporación de estas interacciones al potencial radica fundamentalmente en la ausencia de un modelo adecuado para la descripción del disolvente, que permitiera la reproducción de sus propiedades más relevantes, por un lado, y en los enormes requerimientos computacionales que conlleva la realización de simulaciones en las que se incorporen sus moléculas explícitamente por otro.

### **III.1.3.-Aproximaciones semiempíricas al cálculo de energías libres.**

Aunque las interacciones covalentes constituyen, en términos absolutos, una importante contribución a la energía total del sistema, los procesos de interés en el estudio de macromoléculas biológicas, como son el plegamiento de proteínas o sus interacciones con otras macromoléculas y ligandos, no implican generalmente modificaciones relevantes en la estructura covalente del sistema. Por el contrario, estos procesos están controlados fundamentalmente por la diferencia en el establecimiento de interacciones no covalentes entre los diferentes estados. Además, puesto que la mayoría de dichos procesos se desarrollan en disolución, las interacciones establecidas entre la macromolécula y el disolvente originan contribuciones decisivas en la determinación de las distintas magnitudes termodinámicas.

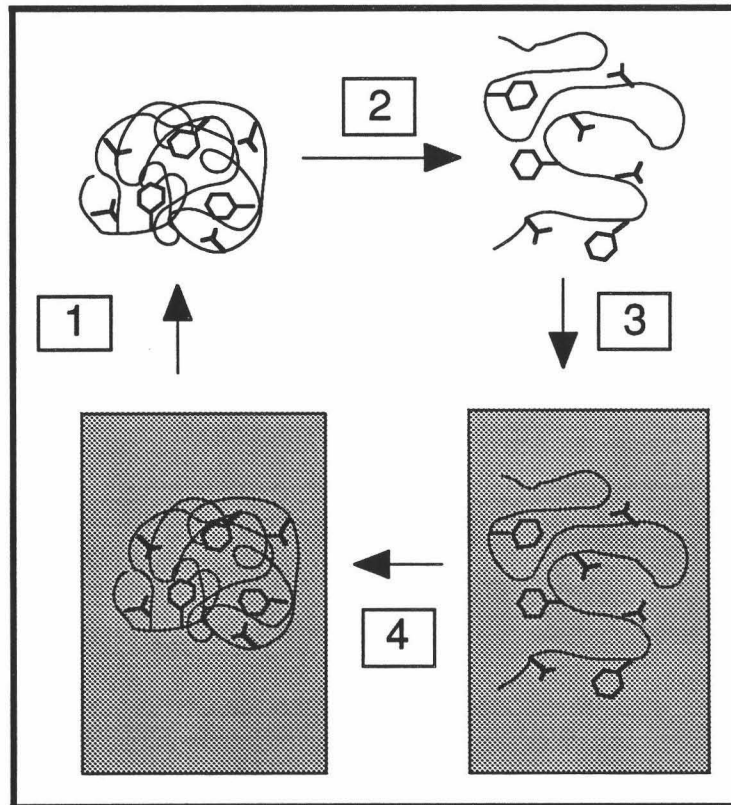
Así, aunque, teóricamente, la aplicación del tratamiento mecánico estadístico a cualquier sistema macromolecular permitiría su descripción termodinámica completa, la extrema dificultad que presenta el cálculo detallado de las interacciones no covalentes, en especial las establecidas con el disolvente, ha inducido el desarrollo de numerosos análisis fenomenológicos con el fin de describir energéticamente los procesos de plegamiento de proteínas así como de sus interacciones con ligandos u otras macromoléculas biológicas en términos estructurales.

#### **III.1.3.1.-Ciclos termodinámicos clásicos.**

El uso de ciclos termodinámicos en los que se representa el proceso bajo estudio mediante una serie de etapas independientes, como el que se muestra en la figura III.1 para el caso concreto del plegamiento de una proteína, ha constituido uno de los enfoques clásicos para abordar, en la práctica, el cálculo de magnitudes termodinámicas a partir de datos estructurales (Makhatadze & Privalov, 1993a; Makhatadze & Privalov, 1993b; Oobatake & Ooi, 1992). En este ciclo, las etapas horizontales representan el proceso de plegamiento/desplegamiento de la proteína en el vacío (parte superior de la figura) y en disolución (parte inferior), mientras que las etapas verticales a derecha e izquierda representan, respectivamente, la hidratación del estado nativo y del desnaturalizado.

En estos ciclos, dada la imposibilidad de medir directamente las energías de solvatación de

solutos del tamaño de las macromoléculas biológicas (Ben-Naim, 1994), la única etapa accesible experimentalmente es la correspondiente al proceso en disolución (etapa 4 en la figura



*Figura III.1: Ciclo termodinámico clásico representativo del proceso de plegamiento/desplegamiento de proteínas. La etapa 4 representa la reacción de plegamiento en disolución acuosa, las etapas 1 y 3 los procesos de hidratación/deshidratación de la proteína plegada y desplegada respectivamente, y la etapa 2 la reacción de desplegamiento en vacío.*

III.1). Los parámetros energéticos correspondientes a la etapa 2, que representa la ruptura de interacciones internas en el vacío, pueden obtenerse, como ya se ha discutido en las secciones anteriores a partir de cálculos de mecánica molecular utilizando una función potencial adecuada y asignando un valor de 1 a la constante dieléctrica del medio (Lazaridis et al., 1995). Asimismo, también se ha intentado su estimación sustrayendo a los resultados obtenidos experimentalmente para una magnitud termodinámica determinada los efectos de hidratación calculados a partir de compuestos modelo (Makhatadze & Privalov, 1993a; Makhatadze & Privalov, 1993b; Makhatadze & Privalov, 1995).

#### III.1.3.1.1.-Hidratación: Hipótesis de aditividad de grupos.



Dada la imposibilidad de medir directamente las energías de solvatación de macromoléculas biológicas y, en concreto, de proteínas, se ha recurrido con frecuencia a la expresión de la energía libre de hidratación de la molécula en términos de las energías de sus componentes individuales (Ben-Naim, 1994), estimados a partir de estudios experimentales de transferencia de compuestos modelo a la disolución acuosa. Este procedimiento implica la suposición de un comportamiento aditivo o superposición lineal de las contribuciones de cada uno de los distintos grupos constituyentes de la macromolécula. De este modo, la contribución de cada grupo funcional a la magnitud termodinámica de interés debe ser homogénea en toda la macromolécula y equivalente a la observada en los compuestos modelo.

Con el fin de simular el proceso de transferencia de las cadenas laterales de los distintos aminoácidos, especialmente las de carácter hidrofóbico, desde el interior de la proteína a su superficie, se han realizado numerosos estudios experimentales de procesos de transferencia de compuestos modelo, desde una fase inicial de naturaleza variada (vacío o fase gaseosa (Wolfenden et al., 1981), disolvente orgánico (Fauchère & Pliska, 1983; Kauzmann, 1959; Nozaki & Tanford, 1971; Tanford, 1962), o aminoácidos cristalinos) hasta el medio acuoso. La elección del medio de origen que represente con mayor fidelidad las condiciones en que las cadenas laterales se encuentran en el interior de la proteína, junto con la hipótesis de aditividad, constituyen los principales obstáculos para la interpretación de los resultados obtenidos con estos procedimientos.

#### III.1.3.1.2.-Estimación de las contribuciones a la hidratación a través de accesibilidad al disolvente.

La relación lineal observada entre las energías libres de solvatación procedentes de estos estudios de transferencia de compuestos modelo y sus superficies accesibles al disolvente han permitido, mediante la estimación de los coeficientes de proporcionalidad obtenidos a partir del ajuste de los datos experimentales, la predicción de energías libre e incluso entalpías y entropías de solvatación de pequeños compuestos apolares. De este modo, utilizando estos coeficientes, y adoptando la hipótesis de aditividad, se recurre a la estimación de las magnitudes termodinámicas correspondientes a las etapas de hidratación en el ciclo termodinámico descrito en la figura III.1. Para ello, sólo sería necesario calcular los cambios en área accesible al disolvente que experimentan los grupos de la proteína como consecuencia del proceso bajo estudio.

Las dos contribuciones de mayor importancia en los procesos de solvatación de macromoléculas biológicas son las interacciones de van der Waals, que favorecen el empaquetamiento de las moléculas de agua alrededor de su superficie, y el efecto hidrofóbico, de naturaleza fundamentalmente entrópica, que se refleja en la tendencia de los grupos apolares a minimizar su contacto con el disolvente con el fin de optimizar su interacción con éste. De este modo, algunos modelos incluyen ambos efectos en el potencial mediante la definición de una energía superficial, proporcional al área accesible al disolvente. Como consecuencia, una vez



determinados los parámetros de proporcionalidad (generalmente a partir de datos experimentales de transferencia de compuestos modelo desde el vacío a la fase acuosa), el cálculo del área accesible de cada grupo o átomo en la macromolécula permite estimar estas contribuciones de corto alcance a la solvatación (Sharp, 1993; van Gunsteren et al., 1994).

Aspectos como el número y tipo de átomos considerados en el cálculo de las accesibilidades, la definición analítica de las áreas, el tipo de función potencial a la que se encuentran asociados, así como la naturaleza de las moléculas y la procedencia de los datos experimentales utilizados durante el proceso de parametrización, contribuyen a la amplia variedad de métodos para el cálculo de energías de hidratación existentes en la actualidad (van Gunsteren et al., 1994). Las parametrizaciones a nivel atómico desarrolladas por Eisenberg y McLachlan (Eisenberg & McLachlan, 1986) a partir de datos de transferencia entre agua y octanol, y la desarrollada posteriormente por Wesson y Eisenberg (Wesson & Eisenberg, 1992) a partir de datos de transferencia desde la fase gaseosa destacan entre los más ampliamente utilizados.

Resulta evidente que el estado de referencia considerado en cada caso tiene grandes implicaciones en el significado de los resultados obtenidos para las energías de solvatación. En el caso de utilizarse datos de transferencia entre la fase gaseosa y el medio acuoso para la determinación de los parámetros de proporcionalidad, se obtienen valores que contienen exclusivamente las contribuciones correspondientes a la hidratación, por lo que es necesario incluir en la función potencial algún término adicional que permita considerar las contribuciones procedentes de las interacciones internas. Por el contrario, si se utilizan datos de transferencia de compuestos modelo desde una fase de referencia líquida o sólida no es necesario considerar dichos términos explícitamente, aunque, en estos casos, surge la controversia sobre el estado de referencia que mejor reproduce las características del interior de una proteína.

#### III.1.3.1.3.-*Problemas de estos enfoques.*

Aunque el uso de los ciclos como el descrito en la figura III.1 para la evaluación de los parámetros termodinámicos en disolución es conceptualmente correcta, su evaluación presenta serios problemas, por lo que es difícil esperar que posean la precisión necesaria para la predicción de estabilidades o para su aplicación al desarrollo de técnicas de diseño molecular.

##### *a.-Fenómenos compensatorios entre magnitudes termodinámicas.*

La existencia de fenómenos de compensación entre las distintas contribuciones que determinan las magnitudes finales de cada uno de los parámetros termodinámicos constituye uno de los principales inconvenientes de este tipo de enfoques. Así, la energía libre de Gibbs se encuentra determinada por los cambios entálpicos y entrópicos, que se oponen entre sí, de modo que la diferencia entre ambos da lugar a cambios en energía de Gibbs de mucha menor magnitud que las contribuciones individuales. Dichos cambios en entalpía y entropía se encuentran a su vez

determinados por otros términos que se compensan entre sí, de modo que la entalpía correspondiente al establecimiento de interacciones internas posee signo contrario a la entalpía de hidratación, y la entropía relacionada con los cambios en la hidratación se ve compensada por la entropía conformacional. Como resultado, los cambios globales en cada uno de los parámetros termodinámicos son, en general, de mucha menor magnitud que los correspondientes a sus componentes individuales.

En el caso concreto de la entalpía de desnaturalización de una proteína, por ejemplo, la principal dificultad surge como consecuencia de la gran diferencia existente entre la magnitud de la entalpía de desnaturalización observada experimentalmente y las contribuciones entálpicas correspondientes a la ruptura de interacciones no covalentes y a la hidratación. Estas dos contribuciones son de signos opuestos y alrededor de un orden de magnitud mayores que la entalpía total, por lo que el menor error en la estimación de cualquiera de ellas se vería amplificado enormemente en la evaluación de la entalpía de desnaturalización. De acuerdo con el análisis realizado por Hilser y colaboradores (Hilser et al., 1996), con objeto de conseguir un error de  $\pm 6\%$  en la temperatura de desnaturalización, la entalpía asociada al proceso debe estimarse con un error menor del 2%, lo que implicaría la predicción de la entalpía de hidratación con menos de un 0.1% de error.

#### *b.-Comportamiento no aditivo de las contribuciones individuales.*

El segundo inconveniente que presentan estos métodos tiene su origen en la imposibilidad de medir directamente la energética de transferencia de una proteína desde la fase gaseosa a la disolución acuosa. Volviendo al ejemplo de la entalpía, la estimación con exactitud de la entalpía de hidratación a partir de datos experimentales procedentes de estudios con compuestos modelo es una tarea realmente difícil, si no imposible, si se consideran los errores experimentales que conllevan los datos de compuestos modelo y el hecho de que estos sistemas sólo representan aproximadamente la situación en el interior de las proteínas.

Conviene hacer notar que estos errores no incluyen las incertidumbres asociadas a la extrapolación de los resultados obtenidos con compuestos modelos a proteínas mediante la suposición de aditividad en las contribuciones de los distintos grupos. En este sentido, la aproximación de aditividad, necesaria para estos análisis, se ha cuestionado frecuentemente (Cabani et al., 1981; Lazaridis et al., 1995). Para compuestos apolares, en los que las principales interacciones entre sus componentes son de corto alcance y, por lo tanto, de carácter muy local, la suposición de aditividad no parece ser demasiado drástica. Sin embargo, existen evidencias de que esta hipótesis falla cuando se aplica a moléculas que contienen varios grupos polares, entre los que se establecen interacciones de largo alcance (Cabani et al., 1981; Lazaridis et al., 1995; Wolfenden, 1983). Por lo tanto, en sistemas macromoleculares que contengan numerosos grupos polares próximos entre sí (como es el caso de las proteínas y, más concretamente, de su esqueleto

peptídico) es de esperar desviaciones significativas de las reglas de aditividad.

## **III.2.-PARAMETRIZACIÓN ESTRUCTURAL DE LA ENERGÉTICA DE PLEGAMIENTO DE PROTEÍNAS Y UNIÓN DE LIGANDOS.**

### **III.2.1.-Características generales.**

#### **III.2.1.1.-Parametrización independiente de las distintas magnitudes termodinámicas.**

Aunque la energía libre de Gibbs es claramente la magnitud de mayor interés, la determinación experimental, así como la predicción estructural, de las demás magnitudes termodinámicas permite el acceso a información complementaria de gran importancia para el diseño molecular. El estudio termodinámico de las interacciones entre una proteína y sus ligandos proporciona, no sólo el valor de la energía de Gibbs ( $\Delta G$ ), sino los cambios en entalpía ( $\Delta H$ ), entropía ( $\Delta S$ ) y capacidad calorífica a presión constante ( $\Delta C_p$ ) asociados a la unión entre ambos. Por lo tanto, además de ser responsables de la dependencia con la temperatura de las constantes de afinidad, los cambios en entalpía, entropía, y capacidad calorífica reflejan el balance existente entre las distintas fuerzas moleculares que contribuyen a la determinación de la energía libre de Gibbs correspondiente al proceso de unión. De este modo, la comparación de los valores predichos a partir de un análisis termodinámico estructural para estas cuatro magnitudes termodinámicas ( $\Delta G$ ,  $\Delta H$ ,  $\Delta S$  y  $\Delta C_p$ ) con los obtenidos experimentalmente permite determinar si los cálculos realizados captan correctamente la naturaleza y magnitud de las fuerzas implicadas en el proceso de plegamiento o de asociación entre moléculas.

#### **III.2.1.2.-Parametrización basada en datos termodinámicos de proteínas.**

Numerosos autores han considerado el uso de datos experimentales de pequeños compuestos modelo como una posible solución cuantitativa al problema del plegamiento de proteínas y de sus interacciones con diversos ligandos. Como se ha expuesto anteriormente, según este planteamiento, la simple adición de los parámetros termodinámicos correspondientes a grupos químicos específicos en moléculas orgánicas de pequeño tamaño permitiría estimar cuantitativamente la energética de los procesos en que intervienen proteínas. Sin embargo, estos procedimientos han sido objeto de fuertes controversias, centradas fundamentalmente en la validez de la hipótesis de aditividad de las distintas contribuciones, así como en la naturaleza del interior de la proteína.

La parametrización aquí descrita, y utilizada a lo largo de este trabajo, tiene su fundamento en la idea de que el comportamiento de las distintas macromoléculas biológicas, y, en concreto, de las proteínas, no puede racionalizarse a partir de la suma de las interacciones de

grupos individuales en pequeñas moléculas orgánicas, de naturaleza muy diferente a los sistemas en que estamos interesados. De este modo, y como se ha comentado anteriormente, aunque algunas contribuciones, como es el caso de los efectos asociados a la hidratación de grupos apolares o los cambios en capacidad calorífica, se ajustan aparentemente a las reglas de aditividad, otras, sin embargo, no presentan un comportamiento aditivo. Estas desviaciones de la aditividad son especialmente notables en el caso de la hidratación de grupos polares, de las interacciones intramoleculares, o de los cambios en entropía (Cabani et al., 1981; D'Aquino et al., 1996; Lazaridis et al., 1995).

Por lo tanto, se considera que la utilización de datos experimentales obtenidos directamente con proteínas o péptidos para el establecimiento de las correlaciones energía/estructura constituye una aproximación más adecuada al problema de la predicción cuantitativa de la energética de estos sistemas. De este modo, aunque estas correlaciones no proporcionen una explicación de los orígenes y naturaleza de cada una de las contribuciones, son una herramienta de gran utilidad que posibilita la predicción de forma precisa de la energética de proteínas a partir de datos estructurales.

### **III.2.1.3.-La parametrización utilizada representa un enfoque fenomenológico.**

En principio, el desarrollo de una metodología que permita la predicción de estabilidades o afinidades de unión, así como el desarrollo de estrategias de diseño molecular, no requiere necesariamente el conocimiento de las últimas causas de cada una de las fuerzas implicadas en el proceso. Por el contrario, para poder realizar estas tareas es suficiente con disponer de la capacidad para predecir con precisión los parámetros termodinámicos asociados al proceso de plegamiento o a las interacciones con ligandos.

La metodología utilizada en este trabajo, desarrollada durante los últimos años en el laboratorio del Dr. Freire, está basada en un juego de ecuaciones empíricas que permiten relacionar los parámetros termodinámicos de un proceso determinado con las características estructurales del sistema, expresadas como un juego de coordenadas atómicas. Estas ecuaciones se han obtenido a partir del análisis estadístico de la base de datos termodinámicos de sistemas para los cuales se dispone de información estructural de alta resolución. Como se demostrará a continuación, estas ecuaciones paramétricas captan la magnitud de los distintos términos energéticos, sin considerar las contribuciones de las distintas fuerzas elementales de forma explícita.

### **III.2.2.- Energía libre de Gibbs.**

#### **III.2.2.1.-Partición general de la energía libre de Gibbs.**

El objetivo fundamental de la parametrización estructural de la energética, descrita en este

capítulo, es la predicción de los cambios en energía libre de Gibbs asociados a procesos en los que intervienen proteínas (ya sean procesos de plegamiento/desplegamiento, unión de ligandos a enzimas, o interacciones proteína/proteína) a partir de consideraciones estructurales. Para ello es necesario, además de disponer de información estructural de alta resolución para el sistema en cuestión, recurrir a las expresiones termodinámicas básicas que relacionan los distintos componentes de la energía libre de Gibbs.

La energía de Gibbs asociada a un proceso determinado puede expresarse como suma de diferentes contribuciones, de modo que existen diversas posibilidades de partición que pueden diferir entre sí en cuanto al número y naturaleza de los términos considerados. En general, la partición adoptada debe venir determinada, principalmente, por la exactitud con que las distintas contribuciones puedan expresarse y determinarse a partir de parámetros estructurales.

De este modo, la descomposición más general de la energía de Gibbs es:

$$\Delta G = \Delta G_{\text{gen}} + \Delta G_{\text{ion}} + \Delta G_{\text{trans}} + \Delta G_{\text{otros}} \quad (\text{III.6})$$

donde  $\Delta G_{\text{gen}}$  contiene las contribuciones asociadas a la formación de las estructuras secundaria y terciaria, es decir, interacciones de van der Waals, enlaces de hidrógeno, hidratación y entropía conformacional. El segundo término,  $\Delta G_{\text{ion}}$ , hace referencia a los efectos electrostáticos y de ionización, mientras que  $\Delta G_{\text{trans}}$  incluye la contribución originada por la pérdida de grados de libertad traslacionales como consecuencia de las asociaciones intermoleculares. El último término,  $\Delta G_{\text{otros}}$ , normalmente incorpora otras contribuciones, características de cada sistema específico, que deben ser consideradas individualmente.

Algunos de estos términos, como es el caso de los efectos de ionización o las contribuciones traslacionales, se calculan con mayor facilidad y exactitud en términos de energía libre, mientras que, para llevar a cabo la estimación de las contribuciones genéricas, es conveniente la considerar individualmente sus componentes entálpicos y entrópicos.

### III.2.2.2.-Energía de Gibbs genérica.

La denominada energía libre de Gibbs genérica contiene el grueso de las contribuciones relacionadas con los fenómenos estructurales y de hidratación. En este caso, el cálculo se lleva a cabo mediante la estimación individual de sus componentes entálpicos y entrópicos en términos de parámetros estructurales, como pueden ser los cambios en áreas polares y apolares accesibles al disolvente ( $\Delta \text{ASA}_{\text{pol}}$  y  $\Delta \text{ASA}_{\text{ap}}$ ) asociados al proceso en cuestión, o la densidad de empaquetamiento de la estructura proteica, determinada por la distribución de las distancias interatómicas entre los distintos tipos de átomos.

La energía libre genérica queda completamente especificada a cualquier temperatura, una vez determinados los cambios en entalpía ( $\Delta H$ ), entropía ( $\Delta S$ ) y capacidad calorífica ( $\Delta C_p$ ) a una



temperatura de referencia ( $T_R$ ):

$$\Delta G(T) = \Delta H(T) - T \cdot \Delta S(T) \quad (\text{III.7})$$

$$\Delta H(T) = \Delta H(T_R) + \int_{T_R}^T \Delta C_p \cdot dT \quad (\text{III.8})$$

$$\Delta S(T) = \Delta S(T_R) + \int_{T_R}^T \Delta C_p \cdot d \ln T \quad (\text{III.9})$$

De acuerdo con estas ecuaciones, y debido a la dependencia de los cambios en entalpía y entropía con la temperatura, éstos deben evaluarse a una temperatura de referencia y posteriormente extrapolarse a la temperatura deseada mediante el cambio en capacidad calorífica.

La elección de dichas temperaturas de referencia es fundamentalmente una cuestión de conveniencia, de modo que las temperaturas escogidas representan normalmente condiciones en las que, o bien la adquisición de datos experimentales es especialmente sencilla y precisa, o bien alguna de las propiedades termodinámicas presenta un comportamiento excepcional. Así, por ejemplo, para la evaluación de los cambios en entalpía, parametrizados fundamentalmente a partir de datos de desnaturalización de proteínas, se ha elegido como temperatura de referencia la correspondiente a la mediana del conjunto de temperaturas de desnaturalización, obtenidas experimentalmente para las proteínas de la base de datos. De este modo, el uso de esta temperatura (60 °C) minimiza los errores experimentales procedentes de las extrapolaciones a temperaturas muy lejanas de la de transición. En el caso de la entropía, por el contrario, las temperaturas de referencia (62 °C y 112 °C) representan los puntos en los que las entropías de hidratación son nulas, facilitando, por tanto, la evaluación de las distintas contribuciones, conformacionales y de hidratación, al cambio entrópico total.

### III.2.2.3.-Cálculo de las áreas accesibles al disolvente.

La parametrización de las contribuciones genéricas a las distintas funciones termodinámicas se establece en términos de distintos parámetros estructurales, entre los que se cuentan los cambios en áreas polares y apolares accesibles al disolvente asociados al proceso de desplegamiento de las proteínas o a la unión de sus ligandos. Es, por tanto, necesario estimar estas áreas en los distintos estados implicados en dichos procesos; es decir, en los estados nativo y desnaturalizado, en el primer caso, y en el complejo y la proteína y el ligando libres, en el segundo.

#### III.2.2.3.1.-Algoritmo de Lee y Richards.



Los valores de áreas polares y apolares accesibles al disolvente (para una definición detallada del área accesible al disolvente, así como de los distintos algoritmos disponibles para su determinación ver Apéndice A.1) se obtienen mediante la aplicación del algoritmo de Lee y Richards (Lee & Richards, 1971) a las estructuras tridimensionales de alta resolución de cada una de los sistemas. El algoritmo utilizado consiste en la generación de secciones bidimensionales a lo largo de uno de los ejes principales de la molécula, representada como un conjunto de esferas de radio equivalente a la suma de los radios de van der Waals de cada átomo y el radio asignado al disolvente o esfera sonda. De este modo, la superficie accesible se obtiene mediante la suma de los productos de los perímetros, definidos por la intersección entre los planos de corte y la molécula, por el espaciado entre los planos de corte.

Por tanto, una vez especificados los valores de los radios de van der Waals para cada tipo de átomo, el espesor de corte transversal y el radio de la esfera sonda que representa la molécula del disolvente, es posible obtener, a partir de las coordenadas atómicas, el área accesible al disolvente para cada uno de los átomos de la molécula. Las accesibilidades deben ser posteriormente clasificadas de acuerdo con su naturaleza. En nuestro caso, se han definido tres tipos atómicos, de modo que cualquier átomo de carbono se considera apolar, y los demás polares, con la excepción de los oxígenos pertenecientes a los grupos hidroxilo de los aminoácidos Ser y Thr, cuyas áreas se clasifican en una categoría especial (ver sección III.2.3.3.2 para más detalles).

La exactitud del algoritmo depende, lógicamente, del valor escogido para el espaciado entre los planos de corte, y alcanza un valor límite conforme se disminuye la separación entre ellos. De este modo, se ha sugerido el empleo de espesores de corte transversal comprendidos entre 0.25 y 0.5 Å como valores óptimos, ya que representan un compromiso entre la exactitud numérica y el tiempo de computación (Richards, 1985).

Aunque inicialmente, tanto para el desarrollo de la parametrización como para sus primeras aplicaciones, se hizo uso del programa ACCESS (Scott R. Pressnell, University of California, San Francisco), en la actualidad los cálculos de accesibilidades se llevan a cabo mediante el empleo de una serie de librerías C, desarrolladas recientemente en el grupo del Dr. Freire (Hilser et al., 1997), y que facilitan la integración del cálculo del área accesible en proyectos de mayor envergadura. En todos los casos, los cálculos se realizan adoptando un valor para el radio del disolvente de 1.4 Å y una espesor de corte transversal de 0.25 Å, así como el juego de radios efectivos de van der Waals mostrados en la tabla III.1.

#### III.2.2.4.2.-Áreas accesibles de conformaciones no estructuradas.

Dado que las distintas contribuciones genéricas a la energía libre se expresan en términos de los cambios en accesibilidad asociados al proceso bajo estudio, el cálculo de las áreas polares y apolares accesibles al disolvente, tanto en el estado inicial como final, constituye una etapa

imprescindible. De este modo, por ejemplo, para la calcular la disminución en la accesibilidad al disolvente de un resto cualquiera en la proteína a consecuencia del proceso de desplegamiento sería necesario evaluar las áreas tanto en el estado nativo como desnaturalizado.

**Tabla III.1:** Radios de van der Waals utilizados para el cálculo de las superficies accesibles al disolvente mediante el algoritmo de Lee y Richards.\*

(Å)			
N	1.65	CA	1.87
CB	1.70	CD	1.87
		CG	1.87
CE	1.87	CE1	1.76
CH2	1.76	CZ	1.87
		ND	1.65
		NE	1.65
		NZ	1.65
OD	1.40	OD1	1.40
OE	1.40	OE1	1.40
OG	1.40	OG1	1.40
OXT	1.40	OH	1.40
P	1.90		
		C	1.76
		CD1	1.76
		CG1	1.87
		CE2	1.76
		CZ2	1.76
		ND1	1.65
		NE1	1.65
		NH1	1.65
		OD2	1.40
		OE2	1.40
		OT1	1.40
		SD	1.85
		O	1.40
		CD2	1.76
		CG2	1.87
		CE3	1.76
		CZ3	1.76
		ND2	1.65
		NE2	1.65
		NH2	1.65
		OT2	1.40
		OH2	1.40
		SG	1.85

\* Lo nomenclatura utilizada para los distintos átomos es la utilizada en los ficheros PDB para denominar los átomos constituyentes de las distintas cadenas laterales.

No obstante, si bien la estimación de la accesibilidad de la proteína nativa no requiere más que la aplicación de alguno de los múltiples algoritmos disponibles al juego de coordenadas atómicas, resultantes de estudios estructurales de alta resolución (RMN o Difracción de Rayos X), la pobre definición estructural del estado desnaturalizado imposibilita la aplicación de este tipo de procedimientos. Como consecuencia, es inevitable recurrir al establecimiento de algún modelo molecular apropiado, que represente aproximadamente las características estructurales de la cadena polipeptídica desplegada, y permita estimar su accesibilidad (Creamer et al., 1995; Creamer et al., 1997). Un problema similar se presenta cuando se pretende predecir la energética de conformaciones parcialmente plegadas o la afinidad de unión de ligandos peptídicos, para lo que se requiere la estimación de las superficies correspondientes a cadenas polipeptídicas no estructuradas.

Entre los modelos moleculares más frecuentemente utilizados para este propósito cabe destacar, desde aminoácidos individuales bloqueados en los extremos (Sneddon & Tobias, 1992), tripéptidos extendidos de secuencia Gly-X-Gly (Lesser & Rose, 1990; Miller et al., 1987; Rose et al., 1985; Shrake & Rupley, 1973), o tripéptidos de secuencia Ala-X-Ala (Lee et al., 1994; Zielenkiewicz & Saenger, 1992), hasta cadenas polipeptídicas de secuencia idéntica a la proteína pero con las cadenas laterales y el esqueleto peptídico completamente extendidos (Livingstone et

al., 1991).

Aunque todos ellos constituyen estados de referencia válidos, que permiten la definición de las variaciones en accesibilidad con respecto al estado nativo de la proteína, ninguno proporciona un buen modelo estructural del estado desplegado. Es de esperar que la presencia de la cadena polipeptídica, de gran longitud y flexibilidad, así como la posibilidad de establecimiento de interacciones entre las cadenas laterales de los aminoácidos se traduzca en una disminución de la accesibilidad promedio de cada resto con respecto a su conformación completamente extendida (Creamer et al., 1995; Creamer et al., 1997; Zielenkiewicz & Saenger, 1992).

Inicialmente, para el desarrollo de la parametrización y sus primeras aplicaciones, se escogió el juego complejo de tripéptidos extendidos de secuencia Ala-X-Ala como modelo molecular del estado desplegado. La accesibilidad del estado desplegado de la proteína o el ligando libre se obtenía mediante el producto del área correspondiente a un aminoácido determinado en el tripéptido (mostrados en la tabla IV.A.2) por el número de veces que éste aparece en la secuencia. Aunque se halla muy extendida y ha sido adoptada por diferentes grupos de investigación, la utilización de estos tripéptidos como modelos moleculares, de acuerdo con lo discutido en el párrafo anterior, conduce a la sobrestimación de los cambios en áreas accesibles al disolvente, por lo que uno de los objetivos de esta tesis ha sido la obtención de un nuevo juego de valores, que refleje más adecuadamente las accesibilidades promedio de los distintos aminoácidos en los estados no estructurados de la proteína. En la actualidad se utilizan estos valores de accesibilidad, recogidos en la tabla IV.A.2 y cuya obtención se describe en el capítulo IV.A, para el cálculo de las áreas accesibles al disolvente correspondientes a conformaciones no estructuradas, como pueden ser el estado desplegado de una proteína o los ligandos peptídicos libres en disolución.

#### **III.2.2.4.-Base de datos empleada para la parametrización.**

La base de datos de proteínas utilizada para llevar a cabo la parametrización estructural de la energética consiste en un conjunto de sistemas que han sido adecuadamente caracterizados desde un punto de vista energético, y para lo cuales se dispone de datos estructurales de alta resolución y que se enumeran a continuación. Se muestran entre paréntesis las referencias correspondientes al análisis termodinámico experimental y entre corchetes del código bajo el cual se encuentra depositada la correspondiente estructura tridimensional en el Banco de Datos de Proteínas (PDB): Citocromo C [5CYT] (Privalov & Khechinashvili, 1974), Anhidrasa Carbónica B [2CAB] (Tatunashvili & Privalov, 1986), Quimotripsina [4CHA] (Tischenko et al., 1974),  $\alpha$  Lactoalbúmina [1ALC] (Griko et al., 1994a), Lisozima de clara de huevo [1LYM] (Khechinashvili et al., 1973), Mioglobina [1MBN] (Privalov et al., 1986), Nucleasa de estafilococo [1SNC] (Griko et al., 1988), Papaína [9PAP] (Tiktopulo & Privalov, 1978), Parvalbúmina [5CPV] (Filimonov et al., 1978), Pepsinógeno [1PSG] (Mateo & Privalov, 1981), Ribonucleasa A [7RSA] (Privalov et al.,

1973), y Tripsina [1TLD] (Tischenko & Gorodnov, 1979).

### **III.2.3.-Contribuciones genéricas a la energía de Gibbs.**

#### **III.2.3.1.-Entalpía.**

##### **III.2.3.1.1.-Consideraciones generales.**

En ausencia de contribuciones procedentes de efectos de protonación o ionización acoplados al proceso de plegamiento o a la unión de ligandos y otros efectos específicos, los cambios de entalpía asociados a estos procesos se originan fundamentalmente a consecuencia de la ruptura de interacciones internas de naturaleza no covalente (como interacciones de van der Waals, enlaces de hidrógeno o interacciones electrostáticas), y de los efectos de hidratación de grupos, inicialmente ocultos en el interior de la proteína plegada o en el complejo y que, tras el proceso de desplegamiento o disociación, pasan a estar en contacto con el disolvente.

Con anterioridad, y desde un punto de vista completamente fenomenológico, se observó la existencia de una relación de proporcionalidad entre los cambios en entalpías y capacidades caloríficas correspondientes al proceso de desplegamiento de proteínas y los cambios en las áreas polares y apolares accesibles al disolvente que dicho proceso implica (Gómez & Freire, 1995; Livingstone et al., 1991; Murphy et al., 1992; Ooi & Oobatake, 1991; Xie & Freire, 1994b). Como consecuencia, una vez corregidos los efectos de protonación, el cambio de entalpía a una temperatura de referencia determinada puede predecirse con exactitud mediante el uso de una ecuación simple del tipo:

$$\Delta H(T_R) = a \cdot \Delta ASA_{ap} + b \cdot \Delta ASA_{pol} \quad (\text{III.10})$$

donde las constantes  $a$  y  $b$  son función del grado de empaquetamiento de la proteína, como se verá más adelante.

A pesar de que, en un principio, puedan resultar sorprendente el que una expresión tan simple como la anterior haya demostrado poseer un mayor poder predictivo que otros enfoques más sofisticados, son varias las razones que permiten justificar el éxito de esta metodología. En primer lugar, conviene hacer incapié en el hecho de que este procedimiento no requiere el cálculo independiente de los términos de hidratación y los correspondientes a la ruptura de interacciones internas. Por el contrario la ecuación III.10 es el resultado de un análisis estadístico de naturaleza empírica, llevado a cabo sobre datos experimentales de entalpías de desplegamiento en disolución acuosa, por lo que no implica la substracción de dos grandes cantidades con el propósito de estimar valores un orden de magnitud inferior.

En segundo lugar, la homogeneidad estructural existente entre las distintas proteínas, que presentan composiciones atómicas, interacciones entre pares de átomos y densidades de

empaquetamiento muy similares, se ha postulado como la razón por la cual únicamente dos parámetros (a y b) son suficientes para reproducir con precisión los datos experimentales. De este modo, las contribuciones entálpicas originadas por la ruptura de las interacciones internas son proporcionales a los cambios en áreas accesibles (Freire, 1995; Griko et al., 1995; Xie et al., 1994), siendo los parámetros de proporcionalidad, a y b, semejantes para todas las proteínas.

### III.2.3.1.2.-Parametrización inicial.

La parametrización original de la entalpía en términos de áreas polares y apolares accesibles al disolvente se realizó utilizando como temperatura de referencia aquella a la que la contribución apolar a la entalpía es prácticamente nula (también denominada temperatura de convergencia),  $T_H^*$ . El valor de  $T_H^*$ , cercano a 100 °C, se obtuvo a partir del análisis de los datos experimentales disponibles para el proceso de plegamiento/desplegamiento de proteínas (Murphy & Gill, 1991).

De este modo, el cambio de entalpía asociado al desplegamiento de una proteína o a la unión de alguno de sus ligandos podría calcularse a cualquier temperatura de acuerdo con la expresión:

$$\Delta H(T) = \Delta H^*(T_H^*) + (\Delta C_{P,ap} + \Delta C_{P,pol}) \cdot (T - T_H^*) \quad (\text{III.11})$$

donde  $\Delta H^*(T_H^*)$  representa la contribución polar a la temperatura de referencia  $T_H^*$  y  $\Delta C_{P,ap}$  y  $\Delta C_{P,pol}$  son las capacidades caloríficas apolares y polares parametrizadas asimismo en función de los cambios en accesibilidades (ver sección III.2.3.3.).

El valor de  $\Delta H^*(T_H^*)$  se obtuvo mediante la parametrización, en términos de los cambios en área accesible polar, de los datos experimentales para las entalpías de desplegamiento previamente extrapolados a la temperatura de referencia. El valor obtenido para el parámetro de proporcionalidad es  $(35 \pm 3) \text{ cal} \cdot (\text{mol} \cdot \text{Å})^{-1}$ , de modo que:

$$\Delta H^*(T_H^*) = (35 \pm 3) \cdot \Delta ASA_{pol} \quad (\text{III.12})$$

Estas ecuaciones permiten la estimación de las entalpías de desplegamiento de proteínas con un error del 6 % a 60 °C, temperatura correspondiente a la mediana de las temperaturas de desnaturalización (Murphy et al., 1992; Murphy & Freire, 1992).

### III.2.3.1.3.-Temperaturas de referencia.

Como se ha comentado en el apartado anterior, inicialmente se adoptó como temperatura de referencia para la entalpía la denominada temperatura de convergencia, cercana a 100 °C, a la cual, los valores de entalpía normalizados por resto parecían adoptar un valor constante para la



mayoría de las proteínas globulares estudiadas hasta el momento. Esta temperatura corresponde, asimismo, a aquella en que la entalpía de vaporización de hidrocarburos líquidos, así como la de disolución de gases apolares, se hacen nulas (Murphy & Freire, 1992). La utilización de esta temperatura de referencia, a la cual supuestamente las contribuciones apolares desaparecen, permitiría la simplificación de las ecuaciones y la reducción del número de parámetros ajustables. No obstante, dado que no se dispone de medidas directas de  $\Delta H(T_H^*)$ , este valor ha de calcularse extrapolando los resultados experimentales, haciendo uso de la capacidad calorífica.

Con posterioridad, se consideró que la elección como referencia de la temperatura correspondiente a la mediana de las temperaturas de desnaturalización experimentales (60 °C) constituiría una mejor opción. De este modo, mediante la utilización de esta temperatura de referencia, es posible minimizar los errores procedentes de la extrapolación a altas temperaturas, que pueden llegar a ser significativos debido a la incertidumbre existente en la determinación experimental de la capacidad calorífica y a la posible dependencia de ésta con la temperatura (Privalov & Makhatadze, 1993).

De este modo, las ecuaciones que relacionan los cambios en entalpía a una temperatura determinada con las variaciones en accesibilidad tienen la forma:

$$\Delta H(T) = a(T) \cdot \Delta ASA_{ap} + b(T) \cdot \Delta ASA_{pol} \quad (\text{III.13})$$

$$\begin{aligned} \Delta H(T) = a(333.15) \cdot \Delta ASA_{ap} + b(333.15) \cdot \Delta ASA_{pol} + \Delta C_p \cdot (T - 333.15) = \\ \Delta H(333.15) + \Delta C_p \cdot (T - 333.15) \end{aligned} \quad (\text{III.14})$$

donde los parámetros  $a$  y  $b$  a la temperatura de 333.15 K (60 °C) adoptan los valores -8.44 y 31.4 cal·(mol·Å<sup>2</sup>)<sup>-1</sup>, respectivamente.

Estos nuevos parámetros proporcionan una estimación más precisa de las entalpías y sugieren un menor valor, cercano a 79 °C para la temperatura a la que se desvanecen las contribuciones apolares, similar a la temperatura de convergencia observada para dipéptidos sólidos (Murphy et al., 1992).

Bajo la mayoría de las condiciones, los valores de entalpía proporcionados por estas ecuaciones constituyen más del 90% de los cambios entálpicos correspondientes al plegamiento de proteínas, ya que los efectos asociados con protonaciones y otros efectos iónicos son relativamente pequeños. A pesar de esto, es necesario considerar explícitamente otras contribuciones adicionales, características de cada sistema, como la unión de ligandos. Estas contribuciones específicas son especialmente relevantes para aquellas transiciones que se producen a bajas temperaturas, en las que las contribuciones genéricas, descritas por estas ecuaciones, son muy pequeñas.

#### III.2.3.1.4.-Parametrización actual: consideración del grado de empaquetamiento.



*a.-Distribución de distancias interatómicas en proteínas.*

La similitud estructural entre las distintas proteínas, a la que se atribuye la existencia de proporcionalidad entre los cambios en entalpía y superficies accesibles, puede analizarse considerando la distribución de las distancias de separación entre los distintos tipos de átomos que las constituyen. Con este fin, recientemente se ha llevado a cabo un análisis de las distancias de separación correspondientes a cada tipo de par atómico (C-C, C-O, CO-C, CO-N, CO-O, O-N, O-O, y C-S) en las distintas proteínas (Hilser et al., 1996). Los autores calculan los valores medios de las distancias interatómicas para cada par ( $\langle R_{X-Y} \rangle$ ), bien considerando una distancia de corte (en concreto 7 Å y 8 Å) o ponderando las distancias individuales de acuerdo con su energía, de modo que:

$$\langle R_{X-Y} \rangle_w = \sum R_{i,j} \cdot (E_{i,j} / E_{total}) \quad (III.15)$$

donde  $R_{i,j}$  es la distancia de separación entre un par de átomos  $i$  y  $j$ ,  $E_{i,j}$  es la energía correspondiente a dicho par, y  $E_{total}$  es la energía total para todos los pares del tipo  $x-y$  en la proteína. Independientemente del criterio adoptado para la definición de la distancia promedio, los valores correspondientes a las distintas proteínas analizadas se encuentran muy agrupados, mostrando muy pequeñas dispersiones en torno a la media.

La pequeña variación observada entre las distancias de separación medias para cada pareja de átomos en las distintas proteínas refleja la ausencia de grandes diferencias entre sus energías de van der Waals promedio. De este modo, la energía correspondiente a la ruptura de interacciones internas será esencialmente constante por unidad de área de un tipo dado, por lo que puede considerarse como una función dependiente exclusivamente de los cambios en las áreas accesibles al disolvente. Es decir, siempre que la distribución de distancias de separación sea similar entre las distintas proteínas, los coeficientes de proporcionalidad, que relacionan los cambios en entalpía con las variaciones en accesibilidad, serán también similares para todas ellas. De esta forma, las entalpías asociadas con la pérdida de estructura y la consecuente hidratación de un Å<sup>2</sup> de superficie polar o apolar serán equivalentes en todas las proteínas.

Cabe esperar desviaciones de este comportamiento, por lo tanto, en aquellos casos en que la distribución de distancias difiera del promedio; es decir, en aquellos sistemas en los que la densidad de empaquetamiento se aleje del promedio observado para el estado nativo. Asimismo, es posible la aparición de desviaciones cuando la distribución estructural de los distintos tipos de átomos y contactos difiera del promedio.

De hecho, se ha observado que estos coeficientes, que permiten la estimación precisa de las entalpías del estado nativo, sobreestiman el cambio en entalpía asociado con el proceso de desplegamiento correspondiente a conformaciones más relajadas, con menor densidad de empaquetamiento (Xie et al., 1994). De este modo, las variaciones en la distribución de las

densidades de empaquetamiento y en la localización de los grupos polares y apolares entre las distintas proteínas, pueden ser los responsables de la estabilización de ciertos estados parcialmente plegados o de ciertas modificaciones del comportamiento cooperativo.

Asímismo es importante considerar la posible existencia de efectos de esta naturaleza a nivel local, que pudieran ser de gran importancia para el desarrollo de procesos de diseño molecular.

### *b.-Introducción del grado de empaquetamiento.*

Como se ha descrito los apartados anteriores, en ausencia de contribuciones específicas, como la protonación de ciertos grupos o la presencia de cofactores o iones metálicos (efectos que deben ser considerados individualmente), las entalpías genéricas se encuentran determinadas fundamentalmente por las interacciones intramoleculares y las contribuciones procedentes de los fenómenos de hidratación.

#### *b.1.-Interacciones de van der Waals.*

El cambio en entalpía procedente de la ruptura de interacciones de van der Waals puede describirse mediante un potencial de Lennard-Jones del tipo:

$$\Delta H_{LJ} = - \sum_{i \neq j} (A/R_{ij}^{12} + B/R_{ij}^6) \quad (\text{III.16})$$

o bien:

$$\Delta H_{LJ} = - \sum_{i \neq j} [A' \cdot (R_{\min}/R_{ij})^{12} + B' \cdot (R_{\min}/R_{ij})^6] \quad (\text{III.17})$$

donde  $R_{\min}$  es la distancia correspondiente al mínimo del potencial de Lennard Jones y que, al igual que los coeficientes  $A'$ ,  $B'$ , toma valores característicos para cada tipo par atómico (C-C, C-O, CO-C, CO-N, CO-O, O-N, O-O Y C-S, en nuestro caso). Dado que, como se desprende de los resultados del análisis de la distribución de distancias interatómicas en proteínas, la relación  $(R_{\min}/R_{ij})$  toma valores promedio entre 0.7 y 0.8 para todas las interacciones (Hilser et al., 1996), el término repulsivo, correspondiente a la decimosegunda potencia del cociente entre las distancias es despreciable, de modo que es posible simplificar la expresión para el cambio entálpico. De esta forma:

$$\Delta H_{LJ} = - \sum_{i \neq j} B' \cdot (R_{\min}/R_{ij})^6 \quad (\text{III.18})$$

Considerando, para el cada tipo de par atómico, la distancia promedio ponderada energéticamente,  $\langle R_{X-Y, w} \rangle$ , la expresión anterior se transforma en :

$$\Delta H_{LJ} \cong \sum_{\text{pares atómicos}} B'_k \cdot N_k \cdot (R_{\min} / \langle R \rangle)_k^6 \cong \sum_{\text{pares atómicos}} B'_k \cdot N_k \cdot U_k^6 \quad (\text{III.19})$$

donde  $N_k$  representa el número de pares de un tipo determinado existentes en la molécula, y la sumatoria se extiende a los diferentes tipos de pares atómicos considerados.

### b.2.-Parametrización conjunta.

Hilser y colaboradores (Hilser et al., 1996) proponen la posibilidad de estimar el cambio de entalpía a partir de una ecuación única, que englobe el efecto de la hidratación y de la ruptura de las interacciones no covalentes:

$$\Delta H = \sum_k \alpha_k \cdot \Delta ASA_k + \sum_k B'_k \cdot N_k \cdot U_k^6 \quad (\text{III.20})$$

donde  $\alpha_k$  representan las contribuciones unitarias a la entalpía de hidratación por  $\text{Å}^2$  de superficie expuesta.

Considerando que, como demuestran los autores, para las proteínas consideradas en la base de datos el número de contactos polares, apolares y mixtos (establecidos entre átomos polares y apolares) perdidos en el proceso de desplegamiento es proporcional a los cambios en áreas accesibles al disolvente polar, apolar y total, respectivamente, la entalpía de plegamiento de una proteína puede parametrizarse de acuerdo con la ecuación:

$$\Delta H(60) = (\alpha_{ap} + \beta_{ap} \cdot U_{ap}^6) \cdot \Delta ASA_{ap} + (\alpha_{pol} + \beta_{pol} \cdot U_{pol}^6) \cdot \Delta ASA_{pol} + \beta_{mix} \cdot U_{mix}^6 \cdot \Delta ASA_{total} \quad (\text{III.21})$$

donde los coeficientes  $\alpha_i$  y  $\beta_i$  representan los parámetros efectivos que pueden obtenerse a partir del análisis estadístico de la base de datos de proteínas. Conviene hacer notar que el término correspondiente a las interacciones mixtas no contiene la contribución debida a la hidratación (parámetro  $\alpha$ ) ya que ésta ha sido considerada en los términos polares y apolares.

Las funciones  $U$  (definidas como el cociente entre las distancias medias observadas en la proteína y el radio correspondiente al mínimo del potencial de Lennard Jones para cada tipo de interacción) se calculan utilizando los parámetros publicados por Levitt (Levitt, 1974) y las distancias correspondientes a los promedios ponderados energéticamente.

En la ecuación III.21 la función  $U_{ap}$  es el promedio entre todas las funciones  $U$  correspondientes a interacciones establecidas entre átomos apolares,  $U_{pol}$  entre átomos polares, y,

por último  $U_{\text{mix}}$  entre átomos polares y apolares. Estas últimas parecen ser interacciones de gran importancia en la determinación del valor final de la entalpía (Lazaridis et al., 1995).

Los valores de los distintos parámetros, obtenidos mediante ajuste no lineal de los datos procedentes del estudio experimental de proteínas, son:  $\alpha_{\text{ap}} = -12.96$ ;  $\beta_{\text{ap}} = 25.34$ ;  $\alpha_{\text{pol}} = 24.38$ ;  $\beta_{\text{pol}} = 16.57$ ; y  $\beta_{\text{mix}} = 16.42$ . Estos coeficientes no son más que términos paramétricos que contienen la distintas contribuciones promediadas sobre el colectivo de proteínas en la base de datos y, como tales, no pueden relacionarse directamente con una interacción específica o un proceso concreto, por lo que deben ser tratados como simples parámetros procedentes de ajustes fenomenológicos.

Dichos parámetros únicamente adoptarán valores similares a los resultantes del estudio de pequeños compuestos modelo en aquellos casos en los que la contribución de una interacción específica a la entalpía presente un comportamiento aditivo. La comparación de los valores obtenidos para proteínas con los descritos para otras moléculas utilizadas frecuentemente como modelos (Hilser et al., 1996), sugiere que, aunque las contribuciones apolares parecen cumplirlas con bastante fidelidad, las de naturaleza polar muestran desviaciones significativas de las reglas de aditividad (Cabani et al., 1981; Lazaridis et al., 1995).

Aunque el desarrollo de estas nuevas ecuaciones proporciona una mejora estadística significativa en cuanto a la capacidad para predecir las entalpías, pasando de un error del 6% al 3%, su ventaja más interesante consiste en que proporcionan la posibilidad de analizar situaciones muy variadas, desde hélices expuestas al disolvente, glóbulos fundidos, y otros estados parcialmente plegados, que presentan desviaciones en el grado de empaquetamiento con respecto a las proteínas nativas. Estas nuevas ecuaciones introducen la proporcionalidad de la entalpía con la densidad de empaquetamiento a través del parámetro estructural  $U$ , de modo que la entalpía será máxima si  $U = 1$  y mínima (y esencialmente igual a la entalpía de hidratación) si  $U = 0$ . Es interesante hacer notar que, si las funciones  $U$  adoptan los valores correspondientes a las proteínas nativas, la ecuación III.21 se reduce a:

$$\Delta H(60) = 31.4 \cdot \Delta \text{ASA}_{\text{pol}} - 8.44 \cdot \Delta \text{ASA}_{\text{ap}} \quad (\text{III.22})$$

expresión idéntica a la obtenida en el estado anterior de la parametrización, en el que no se consideran los efectos de empaquetamiento explícitamente. Esto parece confirmar la capacidad de la ecuación III.21 para reproducir las pequeñas diferencias en las distancias de separación encontradas en proteínas.

### III.2.3.2.-Entropía.

#### III.2.3.2.1.-Consideraciones generales.

La parametrización efectiva de los cambios entrópicos asociados a procesos de plegamiento y desplegamiento de proteínas o a la unión de sus ligandos requiere la separación de las contribuciones en dos términos: la entropía conformacional y la entropía asociada con los cambios de hidratación de los grupos polares y apolares de la molécula:

$$\Delta S_{\text{gen}}(T) = \Delta S_{\text{solv}}(T) + \Delta S_{\text{conf}} \quad (\text{III.22})$$

La contribución del disolvente, dependiente de la temperatura, tiene su origen fundamentalmente en el efecto hidrofóbico y, en menor medida, en la hidratación de grupos polares. Por otra parte, la entropía conformacional, que esencialmente es constante con la temperatura, procede de las variaciones en el número de conformaciones accesibles del esqueleto peptídico y de las cadenas laterales de la proteína y/o el ligando como consecuencia del plegamiento o la asociación. El plegamiento de una proteína, por ejemplo, conlleva una gran disminución en el número de grados de libertad, lo que se traduce en una gran contribución negativa al cambio en entrópico total, que debe ser compensado por otras fuerzas estabilizadoras como la deshidratación de superficies hidrofóbicas o el establecimiento de interacciones internas.

La predicción de estabildades y afinidades de unión requiere una estimación precisa de los cambios en entropía asociados al proceso bajo estudio, ya que pequeñas variaciones en la entropía se traducen en grandes cambios en estabilidad o afinidad de unión. Para una proteína globular típica, por ejemplo, un error del  $\pm 5\%$  en la estimación de la entropía tiene como resultado un error de  $\pm 15\%$  en el cálculo de la temperatura de desnaturalización.

#### III.2.3.2.2.-Entropía de hidratación.

Las contribuciones entrópicas procedentes de los efectos de solvatación pueden expresarse en función de las capacidades caloríficas, si se utilizan como referencia las temperaturas a las que las entropías de hidratación polar y apolar son nulas. De este modo:

$$\begin{aligned} \Delta S_{\text{solv}}(T) &= \Delta S_{\text{solv, ap}}(T) + \Delta S_{\text{solv, pol}}(T) = \\ &= \Delta C_{P, \text{ap}} \cdot \ln(T/T_{S, \text{ap}}^*) + \Delta C_{P, \text{pol}} \cdot \ln(T/T_{S, \text{pol}}^*) \end{aligned} \quad (\text{III.23})$$

donde  $\Delta C_{P, \text{ap}}$  y  $\Delta C_{P, \text{pol}}$  representan las capacidades caloríficas asociadas a los grupos apolares y polares respectivamente, que se calculan, como se describirá en la sección III.2.3.3., a partir de las áreas apolares y polares accesibles al disolvente.

Inicialmente, se adoptó como referencia la denominada temperatura de convergencia para la entropía, cercana a 112 °C (Baldwin, 1986; Murphy & Freire, 1992; Murphy & Gill, 1990). A esta temperatura, el cambio en entropía normalizado por número de restos adopta el mismo valor para un determinado conjunto de proteínas globulares. Estas proteínas están caracterizadas

estructuralmente por valores constantes, una vez normalizados por resto, de las superficies polares protegidas del disolvente y valores variables de las áreas apolares. El valor adoptado por  $T_{S,ap}^*$  es invariable para todos los procesos que implican la transferencia de superficies apolares al medio acuoso (Baldwin, 1986; Murphy et al., 1990). De acuerdo con estas consideraciones, el cálculo de la entropía podría llevarse a cabo mediante la expresión:

$$\Delta S(T) = \Delta S^* + (\Delta C_{P,ap} + \Delta C_{P,pol}) \cdot \ln(T/385.15) \quad (III.24)$$

donde  $\Delta S^*$  es el cambio de entropía residual evaluado a 112 °C, correspondiente a las contribuciones conformacionales y a la entropía de hidratación polar residual a esta temperatura.

Estudios recientes han demostrado que las contribuciones entrópicas correspondientes a la solvatación de grupos polares son muy pequeñas a temperatura ambiente, habiéndose propuesto como referencia para el término polar una temperatura de 62 °C (D'Aquino et al., 1996), por lo que actualmente la evaluación de la contribuciones entrópicas relacionadas con el disolvente se realiza de acuerdo con la ecuación:

$$\Delta S_{solv}(T) = \Delta C_{P,ap} \cdot \ln(T/385.15) + \Delta C_{P,pol} \cdot \ln(T/335.15) \quad (III.25)$$

Asimismo, es interesante hacer notar que, mientras la entropía de solvatación apolar parece comportarse de forma aditiva, existen evidencias que indican que las contribuciones polares, fundamentalmente procedentes de efectos electrostáticos y de protonación, parecen depender del número de grupos funcionales de naturaleza polar en la molécula y de su proximidad espacial (Cabani et al., 1981).

### III.2.3.2.3.-Entropía conformacional.

La evaluación de los cambios en entropía conformacional, que reflejan la restricción de los grados de libertad rotacionales correspondientes a los ángulos de torsión de la macromolécula y el ligando, han constituido el mayor obstáculo en el proceso de cálculo de energías libres de unión o de plegamiento. Como se ha descrito en el apartado anterior, los cambios en la entropía de hidratación pueden parametrizarse de forma efectiva como una función de los cambios en áreas polares y apolares accesibles al disolvente a través de las capacidades caloríficas (ver ecuación III.25). Sin embargo, el cambio de entropía conformacional no permite este tipo de parametrización, por lo que debe ser estimado independientemente.

#### a.-Parametrización inicial

Dado que la entropía de solvatación apolar es nula a la temperatura de referencia escogida ( $T_{S}^* = 112$  °C), la entropía residual esta temperatura, una vez corregidos los efectos electrostáticos



y de protonación, refleja las contribuciones conformacionales y estadísticas a la entropía total. Una vez normalizada por resto, esta magnitud representa la contribución promedio de un aminoácido a la entropía, así como las posibles restricciones en la entropía del estado desnaturalizado asociadas a la presencia de puentes disulfuro.

De este modo, a partir del análisis de la base de datos de proteínas globulares se obtiene un valor de  $4.3 \text{ cal} \cdot (\text{K} \cdot \text{mol} \cdot \text{res})^{-1}$  (Murphy & Freire, 1992; Murphy & Gill, 1991). Por lo tanto, la entropía conformacional correspondiente al proceso de plegamiento/desplegamiento de una proteína cualquiera podría estimarse sin más que multiplicar esta contribución individual por el número de restos en su secuencia (Murphy et al., 1992).

Sin embargo, el valor promedio obtenido a partir del análisis de la base de datos de proteínas no permite distinguir entre los aminoácidos con distintas cadenas laterales, cada uno de los cuales contribuirá de forma diferente al cambio de entropía conformacional. Esta limitación, que no representa un gran problema para el análisis de transiciones de plegamiento/desplegamiento, en las que se encuentran implicados un elevado número de aminoácidos (Murphy et al., 1992; Murphy & Freire, 1992), puede traducirse en la aparición de desviaciones significativas en la estimación de las entropías correspondientes a procesos que implican pequeñas cadenas peptídicas como la unión de ligandos o el desplegamiento de hélices aisladas.

**b.-Parametrización actual: evaluación independiente de los distintos términos entrópicos.**

Como se ha comentado en el apartado anterior, la capacidad de evaluar con exactitud las distintas contribuciones a la entropía conformacional es un requerimiento imprescindible para la estimación de las estabildades y afinidades de unión y, como consecuencia, para el desarrollo de metodologías que permitan el diseño molecular. La estimación precisa de los cambios en entropía conformacional requiere el cálculo de las contribuciones entrópicas correspondientes a cada uno de los enlaces simples con capacidad de rotación existentes en la molécula bajo estudio, para ambos estados, inicial y final, implicados en el proceso en cuestión. Estos cálculos requieren el desarrollo de algoritmos que permitan la cuantificación de las entropías conformacionales utilizando las coordenadas atómicas de las moléculas implicadas.

Dada la homogeneidad existente en las proteínas, compuestas exclusivamente por veinte aminoácidos, es posible abordar la evaluación independiente de las contribuciones aportadas por cada uno de ellos. Con objeto de facilitar el cálculo de estas entropías, es conveniente considerar de forma independiente las contribuciones procedentes de los ángulos diedros del esqueleto peptídico y las correspondientes a las distintas cadenas laterales, de modo que:

$$\Delta S_{\text{conf}} = \Delta S_{\text{esq.}} + \Delta S_{\text{c. lat.}} \quad (\text{III.26})$$

donde  $\Delta S_{\text{esq}}$  representa la entropía adquirida por el esqueleto polipeptídico como consecuencia del desplegamiento o la disociación, y  $\Delta S_{\text{c. lat.}}$  la correspondiente a las cadenas laterales.

### **b.1.-Entropía conformacional de las cadenas laterales.**

La entropía configuracional de una cadena lateral puede evaluarse a partir de su perfil energético, calculado utilizando los ángulos diedros de dicha cadena como las únicas variables independientes. En principio, los cambios en entropía configuracional,  $S_{\text{config}}$ , se encuentran determinados, tanto por las contribuciones conformacionales ( $S_{\text{conform}}$ ), como por las vibracionales ( $S_{\text{vib}}$ ). Sin embargo, en procesos como el plegamiento de una proteína y la unión de ligandos peptídicos, en los que los estados vibracionales no se ven alterados durante el proceso (de modo que la entropía vibracional es igual en el estado inicial y final ( $\Delta S_{\text{vib}} = 0$ )) la entropía configuracional puede calcularse considerando únicamente las contribuciones conformacionales (Lee et al., 1994).

$$\Delta S_{\text{config}} = (S_{\text{conform, f}} + S_{\text{vib, f}}) - (S_{\text{conform, i}} + S_{\text{vib, i}}) \cong S_{\text{conform, f}} - S_{\text{conform, i}} \quad (\text{III.27})$$

Con objeto de facilitar el cálculo de las entropías conformacionales, se consideran tres estados posibles para cada una de las cadenas laterales en la proteína. La cadena lateral puede encontrarse enterrada en el interior de la proteína, casi completamente inmovilizada y, por tanto, con una entropía conformacional mínima o casi nula. Asimismo, la cadena lateral puede pertenecer a una región plegada y encontrarse expuesta al disolvente, con lo cual el número de grados de libertad se incrementaría de forma significativa, poseyendo, por tanto, una mayor entropía con respecto al caso anterior. Por último, en el estado desplegado, así como en las regiones desestructuradas de la proteína, las cadenas laterales poseen la máxima entropía conformacional.

Por lo tanto, para describir las transiciones entre estos tres estados, es necesario evaluar dos términos:  $\Delta S_{\text{bu} \rightarrow \text{ex}}$ , asociado a la transferencia de una cadena lateral enterrada en el interior de la proteína a su superficie, y  $\Delta S_{\text{ex} \rightarrow \text{u}}$ , correspondiente al incremento de entropía experimentado por una cadena lateral expuesta al disolvente cuando el esqueleto peptídico pasa de la conformación nativa al estado desnaturalizado.

#### **b.1.1.- $\Delta S_{\text{bu} \rightarrow \text{ex}}$**

Suponiendo que cualquier cadena lateral en el interior de la proteína o del complejo adopta una conformación fija (es decir, posee una entropía conformacional nula), el cambio entrópico asociado a su transferencia a la superficie requeriría exclusivamente la estimación de la entropía conformacional correspondiente a dicha cadena expuesta al disolvente cuando el esqueleto peptídico se encuentra plegado en una conformación determinada.

De este modo, los cambios entrópicos correspondientes a esta etapa para cada una de las cadenas laterales se estiman mediante los valores mostrados en la primera columna de la tabla III.2, que han sido calculados por Lee y colaboradores (Lee et al., 1994) a partir de los perfiles energéticos en conformación helicoidal. Estos autores han llevado a cabo el cálculo de las entropías conformacionales correspondientes a las cadenas laterales de todos los aminoácidos en posición central de una hélice estándar de secuencia (Ala)<sub>4</sub>-X-(Ala)<sub>4</sub>. Los perfiles energéticos (calculados considerando los ángulos diedros de las cadenas laterales como las únicas variables independientes y haciendo uso de las funciones potenciales y parámetros descritos por Weiner y colaboradores (Weiner et al., 1984) y Jorgensen y Tirado Rives (Jorgensen & Tirado-Rives, 1988)) permiten la estimación de la probabilidad de cada conformación y, por tanto, el cálculo de las entropías mediante la ecuación:

$$S = -R \cdot \sum_i p_i \cdot \ln p_i \quad (\text{III.27})$$

donde R es la constante de los gases y  $p_i$  representa la probabilidad de cada conformación, definida como:

$$p_i = \frac{\exp(-E_i/RT)}{\sum \exp(-E_i/RT)} = \frac{\exp(-E_i/RT)}{Z} \quad (\text{III.28})$$

donde Z es la función de partición conformacional.

Los valores obtenidos de este modo para las probabilidades de los distintos conformeros muestran una excelente concordancia con las poblaciones estimadas a partir de las estructuras tridimensionales de proteínas (Blaber et al., 1994), lo que sugiere estas últimas no son más que un reflejo de la distribución de Boltzmann para las energías de los conformeros. Asimismo, las entropías obtenidas para las distintas cadenas laterales son muy similares a las obtenidas por Creamer y Rose a partir de simulaciones de Monte Carlo (Creamer & Rose, 1992).

### b.1.2.- $\Delta S_{\text{ex} \rightarrow \text{u}}$

El segundo término considerado en los cálculos,  $\Delta S_{\text{ex} \rightarrow \text{u}}$ , refleja los cambios en la entropía de una cadena lateral cualquiera como consecuencia del cambio conformacional sufrido por el esqueleto peptídico en el proceso de plegamiento; es decir, del paso de una conformación fija correspondiente a una estructura secundaria determinada a un estado altamente desestructurado como el desnaturalizado. El cálculo exhaustivo de las entropías correspondientes a las cadenas laterales en este último caso presenta una gran dificultad ya que requeriría la evaluación de los perfiles energéticos para todas las conformaciones del esqueleto peptídico que contribuyen de algún modo a la definición del estado desplegado.

Creamer y Rose (Creamer & Rose, 1992) han llevado a cabo la estimación aproximada, mediante cálculos de Monte Carlo de estas entropías en el estado desplegado, de modo que, a partir de estos valores y los obtenidos anteriormente para las cadenas laterales expuestas, se evalúa este segundo término para cada uno de los diferentes aminoácidos. Los valores obtenidos se muestran asimismo, en la segunda columna de la tabla III.2.

Tabla III.2: Entropías conformacionales de los distintos aminoácidos.\*

Aminoácido	$\Delta S_{bu \rightarrow ex}$ cal·(K·mol) <sup>-1</sup>	$\Delta S_{ex \rightarrow u}$ cal·(K·mol) <sup>-1</sup>	$\Delta S_{bb}$ cal·(K·mol) <sup>-1</sup>
Ala	0.00	0.00	4.1
Arg	7.11	-0.84	3.4
Asn	3.29	2.24	3.4
Asp	2.00	2.16	3.4
Cys	3.55	0.61	3.4
Gln	5.02	2.12	3.4
Glu	3.53	2.27	3.4
Gly	0.00	0.00	6.5
His	3.44	0.79	3.4
Ile	1.74	0.67	2.18
Leu	1.63	0.25	3.4
Lys	5.86	1.02	3.4
Met	4.55	0.58	3.4
Phe	1.40	2.89	3.4
Ser	3.68	0.55	3.4
Thr	3.31	0.48	3.4
Trp	2.74	1.15	3.4
Tyr	2.78	3.12	3.4
Val	0.12	1.29	2.18

\*Tabla reproducida de (D'Aquino et al., 1996). Los valores de  $\Delta S_{bu \rightarrow ex}$  y  $\Delta S_{ex \rightarrow u}$  proceden de (Lee et al., 1994). Con el propósito de aproximar el valor de  $\Delta S_{bb}$ , los aminoácidos se han clasificado en cuatro grupos de acuerdo con la naturaleza de sus cadenas laterales: los que no poseen carbono beta (Gly), los que poseen sólo carbono beta (Ala), los que poseen cadenas ramificadas en el carbono beta (Val e Ile) y los que poseen carbono gamma.

### b.2.-Entropía conformacional del esqueleto peptídico.

El cálculo de los valores de entropía conformacional asociados a la transición del esqueleto peptídico de un aminoácido desde su conformación nativa al estado desnaturalizado puede llevarse a cabo mediante un procedimiento similar al descrito en el apartado anterior para la entropía de las cadenas laterales. Al igual que entonces, suponiendo que en el estado plegado el esqueleto peptídico adopta una única conformación, con lo que su contribución entrópica será

nula, el cambio entrópico asociado al proceso de desplegamiento puede estimarse una vez conocida la entropía del esqueleto de dicho aminoácido en el estado desnaturalizado.

Con este fin, D'Aquino y colaboradores (D'Aquino et al., 1996) han obtenido recientemente los perfiles energéticos en función de los ángulos diedros del esqueleto peptídico de varios aminoácidos en dipéptidos. A partir de estos datos es posible realizar el cálculo directo de la distribución de probabilidad de los diferentes conformeros y, como consecuencia, de la entropía conformacional del esqueleto peptídico. Los resultados obtenidos, recogidos en la tercera columna de la tabla III.2, reflejan la influencia que las distintas cadenas laterales ejercen sobre la movilidad del esqueleto peptídico del correspondiente aminoácido.

Así, se puede observar cómo las mayores restricciones conformacionales provienen de la presencia del carbono beta, responsable de la notable reducción en la entropía conformacional del esqueleto peptídico de Ala con respecto a Gly (la diferencia entre ambos del orden de  $2.46 \text{ kcal} \cdot (\text{K} \cdot \text{mol})^{-1}$ ). La extensión de las cadenas laterales más allá del carbono beta introduce asimismo ciertas restricciones estéricas, aunque de mucha menor magnitud que las impuestas por el carbono beta, de modo que las cadenas laterales ramificadas en el carbono beta son las responsables de las mayores reducciones en la entropía del esqueleto peptídico.

Los valores resultantes de estos cálculos son muy similares a los obtenidos experimentalmente, mediante el análisis de las curvas de desnaturalización térmica de mutantes de la cremallera de leucinas GCN4, para el cambio en entropía conformacional del esqueleto peptídico entre Gly y Ala (D'Aquino et al., 1996). Asimismo, los valores recogidos en la tabla III.2 son muy parecidos a los descritos por Leach y Scheraga (Leach et al., 1966), procedentes del análisis de la fracción de conformaciones del esqueleto peptídico permitidas estéricamente para cada tipo de aminoácido.

### **b.2.1.-Efecto de la longitud de la cadena**

Los valores mostrados en la tabla III.2 proceden del análisis conformacional de los distintos aminoácidos en un dipéptido, de modo que únicamente incluyen los efectos de volumen excluido correspondientes a los aminoácidos adyacentes. Sin embargo, el número de conformaciones inaccesibles debido a efectos estéricos aumenta paralelamente a la longitud de la cadena polipeptídica, alcanzando un valor asintótico para cadenas suficientemente largas. Por lo tanto, cuando la entropía del esqueleto de la cadena polipeptídica se calcula como la suma de las contribuciones entrópicas de los aminoácidos individuales, es necesario considerar un término corrector que incluye las restricciones de volumen excluido y tiene la forma:

$$S_{\text{corr}} = (R/n) \cdot \ln F_{\text{acc}}(n) \quad (\text{III.29})$$

donde  $R$  es la constante de los gases y  $F_{acc}$  representa la fracción de conformaciones permitidas para una cadena peptídica de  $n$  restos.  $S_{corr}$  es una función que varía lentamente con el número de restos en la cadena, tomando valores iniciales de  $-0.09 \text{ cal}\cdot(\text{K}\cdot\text{mol}\cdot\text{res})^{-1}$  para  $n = 10$  y alcanzando un valor asintótico de  $-0.128 \text{ cal}\cdot(\text{K}\cdot\text{mol}\cdot\text{res})^{-1}$  para  $n = 440$ . El valor de  $S_{corr}$  puede calcularse para cada caso, pero en el caso de una proteína globular típica tiene un valor cercano a  $-0.121 \text{ cal}\cdot(\text{K}\cdot\text{mol}\cdot\text{res})^{-1}$ . Aunque la corrección por resto es pequeña, para una proteína constituida por 100 aminoácidos, la corrección total es del orden de  $-12 \text{ cal}\cdot(\text{K}\cdot\text{mol})^{-1}$ , que representa una contribución a la energía libre de  $-3.6 \text{ kcal}\cdot\text{mol}^{-1}$  a  $25^\circ\text{C}$ .

Asímismo, la entropía conformacional de las cadenas peptídicas desplegadas varía ligera, aunque significativamente, con la densidad de carga de la cadena peptídica, de modo que, en general, una elevada densidad de carga, tanto repulsiva como atractiva, disminuye la entropía conformacional de un péptido no estructurado.

### II.2.3.3.- Capacidad calorífica.

#### III.2.3.3.1.-Consideraciones generales.

La capacidad calorífica a presión constante  $C_p$  es una medida de las fluctuaciones en la magnitud de la entalpía de un sistema determinado. De este modo, desde un punto de vista termodinámico-estadístico, la capacidad calorífica se define como la derivada con respecto a la temperatura, evaluada a presión constante, de la entalpía promedio del sistema, de acuerdo con la ecuación:

$$C_p = (\partial \langle H \rangle / \partial T)_p = \left\{ \left[ \sum_{i=0}^N H_i^2 \cdot \exp(-G_i/RT) / Q \right] - \left[ \sum_{i=0}^N H_i \cdot \exp(-G_i/RT) / Q \right]^2 \right\} / RT^2 \quad \text{(III.30)}$$

$$= \{ \langle H^2 \rangle - \langle H \rangle^2 \} / RT$$

donde  $N$  es el número de estados accesibles al sistema;  $R$  la constante de los gases;  $\langle H \rangle$  representa el valor promedio de la entalpía del sistema obtenido como:

$$\langle H \rangle = \sum_{i=0}^N H_i \cdot \exp(-G_i/RT) / Q \quad \text{(III.31)}$$

y  $Q$  es la función de partición definida como la sumatoria de los pesos estadísticos:

$$Q = \sum_{i=0}^N \exp(-G_i/RT) \quad \text{(III.32)}$$



La ecuación III.30 es una de las expresiones estándar de la termodinámica estadística y refleja el hecho de que la capacidad calorífica a presión constante es directamente proporcional al segundo momento de la distribución entálpica alrededor de su valor promedio, o, en términos más simples, a la dispersión característica de la distribución de los valores de entalpía.

Diversos estudios teóricos sobre la capacidad calorífica de proteínas sugieren que las fluctuaciones entálpicas en estos sistemas proceden fundamentalmente de las rotaciones internas, de las fluctuaciones conformacionales de baja frecuencia, así como de los modos vibracionales de alta frecuencia, asociados a la elongación y flexión de los enlaces covalentes (Benson, 1968). Sin embargo, aunque otros efectos como las interacciones no covalentes o la hidratación de los grupos localizados en la superficie molecular constituyen sólo una pequeña fracción de la capacidad calorífica total de una proteína, estos términos son los principales responsables de los cambios en capacidad calorífica asociados a los procesos de desnaturalización o unión de ligandos.

La capacidad calorífica desempeña un papel fundamental en la determinación de la energética de estabilización de las estructuras de proteínas y de la afinidad de unión de sus ligandos. Dado que el proceso de desplegamiento de una proteína conlleva un incremento significativo de la capacidad calorífica, esta magnitud tiene una gran influencia en la determinación de la energía de Gibbs de estabilización (Privalov & Khechinashvili, 1974). Así, el cambio en capacidad calorífica  $\Delta C_p$  afecta tanto a la entalpía  $[\Delta H(T)]$  como a la entropía  $[\Delta S(T)]$ , magnitudes que determinan la energía de Gibbs, de forma que:

$$\Delta H(T) = \Delta H(T_R) + \int_{T_R}^T \Delta C_p \cdot dT \quad (\text{III.33})$$

$$\Delta S(T) = \Delta S(T_R) + \int_{T_R}^T \Delta C_p \cdot d \ln T \quad (\text{III.34})$$

donde  $\Delta H(T_R)$  y  $\Delta S(T_R)$  representan los cambios entálpicos y entrópicos a la temperatura de referencia  $T_R$ .

En promedio, para una pequeña proteína globular, el término correspondiente a la capacidad calorífica modifica la entalpía de estabilización entre 150 y 250 kcal·mol<sup>-1</sup> en el intervalo de temperatura de 0 a 100 °C, mientras que la entropía se ve modulada entre 450 y 780 cal·(K·mol)<sup>-1</sup> en el mismo intervalo de temperaturas. El cambio en capacidad calorífica es asimismo el responsable de fenómenos como la desnaturalización fría o la existencia de una temperatura de máxima estabilidad para las proteínas. Resulta evidente, por tanto, que la capacidad para calcular con suficiente precisión los cambios en capacidad calorífica a partir de datos estructurales es un requisito indispensable para el desarrollo de metodologías de cálculo de

magnitudes termodinámicas, con capacidad para describir adecuadamente los procesos de plegamiento de proteínas y unión de ligandos.

### III.2.3.3.2.-Parametrización inicial.

Hace ya algunos años que diversos grupos atribuyeron el origen de los cambios de capacidad calorífica que acompañan el proceso de desplegamiento de una proteína a la exposición al medio acuoso de las cadenas laterales apolares, ocultas en el interior de la proteína en el estado nativo (Privalov & Khechinashvili, 1974). El papel de los grupos polares en la determinación de los cambios de capacidad calorífica, cuyas contribuciones son de signo opuesto a las procedentes de los grupos apolares (Murphy & Gill, 1991; Privalov & Makhatadze, 1992), se estableció con posterioridad, aceptándose en la actualidad la existencia de ambos tipos de contribuciones.

De este modo, varios laboratorios han demostrado que el cambio en capacidad calorífica asociado con el proceso de plegamiento y desplegamiento de proteínas puede parametrizarse de forma sencilla en términos de áreas polares y apolares accesibles al disolvente, de modo que las ecuaciones paramétricas obtenidas proporcionan una estimación precisa de las capacidades caloríficas experimentales (Murphy et al., 1992; Oobatake & Ooi, 1992; Privalov & Makhatadze, 1992; Spolar et al., 1992; Xie & Freire, 1994a).

Así, el cambio observado en la capacidad calorífica de una proteína como consecuencia de un cambio conformacional o de la unión de algún ligando puede expresarse en función de los cambios en accesibilidad al disolvente según la ecuación:

$$\Delta C_p = a \cdot \Delta ASA_{ap} + b \cdot \Delta ASA_{pol} = 0.45 \cdot \Delta ASA_{ap} - 0.26 \cdot \Delta ASA_{pol} \quad (\text{III.35})$$

donde los coeficientes 0.44 y -0.26 representan las contribuciones elementales por Å<sup>2</sup> de superficie apolar y polar expuesta al medio acuoso en cal·(K·mol)<sup>-1</sup>, obtenidos a partir del análisis estadístico de los datos termodinámicos de las proteínas de la base de datos (Murphy et al., 1992; Murphy & Freire, 1992; Xie et al., 1994; Xie & Freire, 1994b). La ecuación anterior refleja el hecho de que la exposición al disolvente de superficies apolares conduce a un incremento en la capacidad calorífica del sistema, mientras que la solvatación de superficies polares tiene como consecuencia la disminución de la capacidad calorífica total.

En contra de lo supuesto inicialmente, la hidratación de los grupos hidroxilo de ciertas cadenas laterales de naturaleza alifática como Ser y Thr, parece contribuir de forma positiva al cambio de capacidad calorífica (Murphy & Laynez, comunicación personal), por lo que su contribución se introduce en la ecuación anterior como un término adicional de la forma:

$$\Delta C_{p,OH} = 0.17 \cdot \Delta ASA_{OH} \quad (\text{III.36})$$

Como se ha demostrado en numerosos trabajos, esta ecuación predice el cambio en capacidad calorífica de desplegamiento con una precisión ( $\pm 9\%$ ) cercana al error experimental en la determinación de  $\Delta C_p$ . Al igual que ocurre con los términos entálpicos, las contribuciones procedentes de efectos específicos como los asociados a los cambios en el estado de protonación o la unión diferencial de ligandos o desnaturalizantes deben considerarse explícitamente y de forma individual (Gómez & Freire, 1995; Gómez et al., 1995).

### III.2.3.3.3.-Parametrización actual: Dependencia con la temperatura.

Recientemente se ha demostrado que existe cierta dependencia de los cambios de capacidad calorífica correspondientes al proceso de plegamiento/desplegamiento de una proteína con la temperatura (Privalov & Khechinashvili, 1974; Privalov & Makhatadze, 1990). La disminución de la magnitud de  $\Delta C_p$ , observada a altas temperaturas y ocasionalmente también a temperaturas bajas (Privalov & Makhatadze, 1990), se debe a que, mientras la capacidad calorífica del estado nativo aumenta aproximadamente de forma lineal con la temperatura, la correspondiente al estado desnaturalizado, que de por sí varía con la temperatura a menor velocidad que la del nativo, presenta una pendiente decreciente a altas temperaturas.

La predicción del comportamiento termodinámico de las proteínas, tanto en los procesos de desplegamiento como en los de asociación, requiere la parametrización precisa de los correspondientes cambios en capacidades caloríficas y de su dependencia con la temperatura. Las capacidades caloríficas molares parciales, así como su dependencia con la temperatura, se han determinado calorimétricamente para varias proteínas tanto en el estado nativo como en el desnaturalizado (Griko et al., 1994a; Griko et al., 1994b; Makhatadze et al., 1994; Makhatadze et al., 1993; Privalov & Makhatadze, 1990; Xie et al., 1994). Estos datos termodinámicos, junto con la información estructural de alta resolución disponible para la mayoría de estas proteínas, proporcionan una excelente oportunidad para explorar las bases estructurales de los cambios de capacidad calorífica.

En este contexto, se ha presentado recientemente el primer intento, llevado a cabo en el laboratorio del Dr. Freire, para modelar en una única función matemática el comportamiento de la capacidad calorífica de los distintos estados de la proteína, incorporando la dependencia con la temperatura de esta magnitud termodinámica (Gómez et al., 1995).

#### a.-Contribuciones a la capacidad calorífica de proteínas.

En general, puede considerarse que la capacidad calorífica molar parcial de una proteína en disolución acuosa está originada por dos contribuciones independientes: un término correspondiente a la hidratación, que refleja las interacciones entre la proteína y el disolvente, y un segundo término o capacidad calorífica intrínseca, que incluye las fluctuaciones entálpicas de la estructura covalente, así como las contribuciones correspondientes a las interacciones no

covalentes presentes en la proteína en cuestión (Yang & Rupley, 1979). De este modo, en términos generales, la capacidad calorífica de una proteína cualquiera puede expresarse como:

$$C_p = C_{p,a} + C_{p,b} + C_{p,c} + C_{p,p} \quad (\text{III.37})$$

El primer término en la expresión anterior,  $C_{p,a}$ , o capacidad calorífica primaria, contiene las contribuciones atómicas y covalentes a la capacidad calorífica total y, por definición depende únicamente de la composición de la proteína; es decir, de su secuencia aminoacídica, siendo, por tanto, independiente del estado conformacional de ésta. El segundo término,  $C_{p,b}$ , refleja las contribuciones de las interacciones no covalentes establecidas en la proteína, mientras que el tercero,  $C_{p,c}$  corresponde a la capacidad calorífica de hidratación, originada por las interacciones entre ésta y el disolvente. El último término,  $C_{p,p}$  representa la capacidad calorífica debida al estado de protonación de las cadenas laterales con grupos ionizables como His, Asp, Glu, Arg o Lys.

De acuerdo con este esquema, y desde un punto de vista experimental, el término  $C_{p,a}$  correspondería a la capacidad calorífica de la cadena polipeptídica no estructurada en estado anhidro, mientras que la capacidad calorífica de la proteína anhidra en estado nativo es igual a la suma de términos  $C_{p,a}$  y  $C_{p,b}$ . En cuanto a los cambios de capacidad calorífica asociados con diversos cambios conformacionales de la proteína (procesos de plegamiento y desplegamiento), o con las asociaciones de proteínas y ligandos en disolución, es obvio que únicamente contienen contribuciones procedentes de los términos  $C_{p,b}$ ,  $C_{p,c}$ , y  $C_{p,p}$ , ya que estos procesos no implican modificación alguna, ni en la masa, ni en la estructura primaria de la proteína, de modo que:

$$\Delta C_p = \Delta C_{p,b} + \Delta C_{p,c} + \Delta C_{p,p} \quad (\text{III.37})$$

#### *b.-Parametrización global de la capacidad calorífica.*

Con objeto de refinar las ecuaciones paramétricas correspondientes a la capacidad calorífica, Gómez y colaboradores (Gómez et al., 1995) se basan en que la capacidad calorífica de una proteína en cualquier conformación puede expresarse, en ausencia de efecto de protonación o tras su corrección, como:

$$C_p = C_{p,a} + C_{p,b} + C_{p,c} \quad (\text{III.38})$$

de modo que esta ecuación podría utilizarse para llevar a cabo el ajuste global de mínimos cuadrados de la base de datos de proteínas, con independencia de su estado conformacional. De este modo, la misma ecuación, con el mismo juego de parámetros, debe ser capaz de predecir la

dependencia lineal con la temperatura de la capacidad calorífica del estado nativo y la dependencia no lineal de la correspondiente al estado desnaturalizado.

Con el fin de establecer ciertas restricciones funcionales a las ecuaciones que describirán cada uno de los términos en la expresión anterior, es conveniente analizar con cierto detalle las características estructurales relacionadas con cada uno de los términos.

### b.1.-Contribuciones covalentes.

La capacidad calorífica de una proteína puede expresarse de forma general como la suma varias contribuciones correspondientes a los distintos grados de libertad (translacionales, rotacionales, vibracionales y electrónicos):

$$C_P = C_{P,trans} + C_{P,rot} + C_{P,vib} + C_{P,elect} \quad (III.39)$$

En el caso de proteínas y otras macromoléculas biológicas, que constan de un elevado número de átomos y, por tanto, contienen un gran número de grados de libertad vibracionales ( $3N - 6$ ), la contribución vibracional juega un papel importante en la determinación de la capacidad calorífica. De este modo, en el intervalo de temperaturas de interés biológico (entre 0 y 100 °C), en el que las contribuciones electrónicas pueden considerarse despreciables (Blinder, 1975), la capacidad calorífica primaria de una proteína se debe fundamentalmente a las frecuencias de vibración características de los modos de flexión y elongación de cada enlace de valencia así como a los correspondientes a las rotaciones internas (Janz, 1958).

La capacidad calorífica primaria, es decir la correspondiente a la estructura covalente de la proteína, representa, en promedio, entre el 91 y 96% de la capacidad calorífica total de una proteína anhidra. Los valores resultantes de la suma de las contribuciones de cada uno de sus grupos constituyentes, tanto a nivel de aminoácidos, como a nivel de enlaces y átomos, son muy similares a los obtenidos experimentalmente, existiendo únicamente pequeñas diferencias procedentes de las contribuciones no covalentes (Gómez et al., 1995).

La capacidad calorífica de una proteína anhidra, una vez normalizada por peso, presenta valores muy similares para todas las proteínas ( $0.298 \pm 0.003 \text{ cal} \cdot (\text{K} \cdot \text{g})^{-1}$ ). Asimismo, la dependencia con la temperatura (lineal en el intervalo de interés) muestra el mismo comportamiento, siendo, en promedio de  $9.77 \pm 0.2 \times 10^{-4} \text{ cal} \cdot (\text{K}^2 \cdot \text{g})^{-1}$  para todas las proteínas de la base de datos (Gómez et al., 1995).

Estas observaciones parecen indicar que, en promedio, la capacidad calorífica primaria de una proteína puede considerarse proporcional a su peso molecular, de modo que puede describirse mediante una expresión de la forma:

$$C_{P,a} = [\nu + \omega \cdot (T - 298.15)] \cdot M. \quad (III.40)$$

donde T es la temperatura absoluta, M es el peso molecular de la proteína y los coeficientes  $\omega$  y  $\nu$  son parámetros ajustables, cuyo valor debe obtenerse del análisis de los datos experimentales. En general, en sistemas como péptidos pequeños, o proteínas cuya composición difiera notablemente de la observada como promedio en las distintas proteínas globulares, es de esperar que estas constantes de proporcionalidad adopten valores diferentes.

### *b.2.-Contribución de las interacciones internas.*

Como se ha comentado en el apartado anterior, las diferencias observadas entre las capacidades caloríficas de las proteínas en estado anhidro determinadas experimentalmente y las calculadas a partir de las contribuciones individuales de sus grupos constituyentes son de pequeña magnitud. Estas diferencias, constituirían, en principio, una medida de las contribuciones no covalentes, procedentes de la estructura secundaria de la proteína (Eftink et al., 1983; Gómez et al., 1995).

En general, es de esperar que las contribuciones procedentes de las interacciones no covalentes dependan de la densidad de empaquetamiento de la proteína, de modo que pueden considerarse, como primera aproximación, proporcionales a la superficie total oculta al disolvente en el estado nativo y pueden describirse de acuerdo con una expresión del tipo:

$$C_{P,b} = [p + q \cdot (T - 298.15)] \cdot BSA_{Total} \quad (III.41)$$

donde p y q representan las constantes de proporcionalidad y  $BSA_{Total}$  el área accesible total, tanto apolar como polar, protegida del acceso del disolvente en la estructura nativa. La pequeña magnitud de las contribuciones no covalentes imposibilita el desarrollo de un análisis más detallado, que incluya los efectos del grado de empaquetamiento.

### *b.3.-Hidratación.*

Como se ha comentado anteriormente, la capacidad calorífica de una proteína nativa en disolución presenta una dependencia con la temperatura casi completamente lineal en el intervalo de temperaturas accesible experimentalmente (entre 0 y 50 °C). Asimismo, además de las contribuciones covalentes, que representan cerca de un 85% de la capacidad calorífica total en disolución, la hidratación de la proteína contribuye de forma positiva a la capacidad calorífica. Por lo tanto, la capacidad calorífica de una proteína en disolución es generalmente de mayor magnitud que la correspondiente al estado anhidro. Sin embargo, el incremento relativo entre ambos estados difiere para las distintas proteínas estudiadas, lo cual sugiere la posibilidad de que estos incrementos dependan de la composición superficial específica de cada proteína.



De igual modo, la dependencia con la temperatura de la proteína nativa en disolución es de mayor magnitud que la observada para la proteína anhidra y para el estado desplegado. Estas contribuciones adicionales a la capacidad calorífica total proceden de la solvatación de los átomos situados en la superficie de la proteína, compuesta por regiones polares y apolares en proporciones variables. Para una proteína globular estándar, alrededor del 55% de la superficie expuesta en el estado nativo es de naturaleza apolar, lo cual explica cuantitativamente la contribución positiva de la hidratación.

A diferencia de lo expuesto para el estado nativo, la dependencia con la temperatura de la capacidad calorífica del estado desnaturalizado de una proteína no es lineal, sino que, para temperaturas comprendidas en el intervalo entre 0 y 100 °C, puede aproximarse mediante un polinomio de segundo orden (Griko et al., 1994a; Martínez et al., 1994; Xie et al., 1994). En general, la capacidad calorífica del estado desplegado es de mayor magnitud que la correspondiente al estado nativo, aunque la diferencia entre ambos disminuye con la temperatura. Estas diferencias en capacidad calorífica y en su dependencia con la temperatura pueden atribuirse a la hidratación de los grupos de la proteína que se encuentran en el interior de la proteína en el estado nativo y que, como consecuencia del desplegamiento, entran en contacto con el disolvente, de modo que el estado desnaturalizado expone una mayor proporción de grupos apolares al medio. Asimismo, y en menor medida, las diferencias en capacidades caloríficas podrían proceder de la ruptura de interacciones no covalentes como consecuencia de la pérdida de las estructuras secundaria y terciaria, consideradas en  $C_{p,b}$ .

En principio, es de esperar que la capacidad calorífica de hidratación sea proporcional a la magnitud de las áreas polares y apolares accesibles al disolvente, del mismo modo que los cambios en capacidad calorífica asociados con las transiciones conformacionales (Murphy et al., 1992; Spolar et al., 1992). De este modo, debe ser posible expresar la capacidad calorífica asociada a los fenómenos de hidratación de cualquier estado conformacional mediante una única función cuadrática de la temperatura cuyos coeficientes se obtengan como combinaciones lineales de las áreas polares y apolares accesibles al disolvente en la conformación pertinente (Gómez et al., 1995):

$$C_{p,c} = a(T) \cdot ASA_{ap} + b(T) \cdot ASA_{pol} \quad (\text{III.41})$$

$$a(T) = a_1 + a_2 \cdot (T - 298.15) + a_3 \cdot (T - 298.15)^2 \quad (\text{III.42})$$

$$b(T) = b_1 + b_2 \cdot (T - 298.15) + b_3 \cdot (T - 298.15)^2 \quad (\text{III.43})$$

#### **b.4.-Efectos de protonación.**

En general, las contribuciones a la capacidad calorífica debidas a fenómenos de ionización de cadenas laterales pueden describirse mediante la ecuación:

$$C_{P, \text{prot}} = \sum_i \left( F_{\text{prot}, i} \cdot \Delta C_{P, \text{prot}, i} + F_{\text{prot}, i} \cdot (1 - F_{\text{prot}, i}) \cdot \frac{\Delta H_{\text{prot}, i}^2}{R \cdot T^2} \right) \quad (\text{III.44})$$

donde la sumatoria incluye cada uno de los grupos ionizables existentes en la proteína.  $\Delta C_{P, \text{prot}}$  representa la diferencia en capacidad calorífica de los distintos estados de protonación de una cadena determinada,  $\Delta H_{\text{prot}}$  es la correspondiente entalpía de protonación y  $F_{\text{prot}}$  es el grado de protonación.

El primer término en la ecuación anterior refleja la contribución debida a la existencia de diferencias entre las capacidades caloríficas de los dos posibles estados de ionización de una cadena lateral, y depende de la magnitud de esta diferencia así como del grado de protonación. En general, la diferencia en las contribuciones de los distintos estados de ionización de las cadenas laterales con grupos ionizables (Arg, Asp, Glu, Lys, His) son de pequeña magnitud (unas 4 cal·(K·mol)<sup>-1</sup> para el grupo imidazol y unas 30 cal·(K·mol)<sup>-1</sup> para el grupo carboxilo) en comparación con la capacidad calorífica total de la proteína.

El segundo término contiene las contribuciones a la capacidad calorífica originadas como consecuencia de las fluctuaciones térmicas del grado de protonación, por lo que es proporcional a la entalpía de protonación efectiva, y adopta su valor máximo para grados de protonación del 50%; es decir, cuando el pH de la solución coincide con el pK<sub>a</sub> del grupo ionizable. Por ejemplo, para un resto de histidina en una solución no tamponada en la que pH = pK<sub>a</sub>, la contribución máxima debida a estos fenómenos es del orden de 70 cal·(K·mol)<sup>-1</sup>, ya que en estas condiciones,  $\Delta H_{\text{prot}} = -7 \text{ kcal} \cdot \text{mol}^{-1}$ . Sin embargo, bajo condiciones estándar, se espera que la contribución total originada por los efectos de protonación sea pequeña, debido a que tanto la capacidad calorífica como la entalpía de protonación asociadas a la absorción o liberación de un protón se encuentran compensadas por las correspondientes capacidades caloríficas y entalpías del tampón.

#### b.5.-Ajuste global de la capacidad calorífica.

En ausencia, o tras la corrección, de los efectos de protonación, la expresión:

$$C_P = C_{P,a} + C_{P,b} + C_{P,c} \quad (\text{III.44})$$

donde cada uno de los términos presenta las formas funcionales descritas en los apartados previos, refleja la capacidad calorífica de cualquier estado conformacional de una proteína determinada. De acuerdo con esto, Gómez y colaboradores (Gómez et al., 1995), han llevado a cabo el ajuste no lineal de mínimos cuadrados global de la base de datos de proteínas completa, con

independencia del estado conformacional, utilizando las ecuaciones descritas anteriormente y que resumimos a continuación:

$$C_{P,a} = [\nu + \omega \cdot (T - 25)] \cdot P.Mol. \quad (III.45)$$

$$C_{P,b} = [p + q \cdot (T - 25)] \cdot BSA_{Total} \quad (III.46)$$

$$C_{P,c} = a(T) \cdot ASA_{ap} + b(T) \cdot ASA_{pol} \quad (III.47)$$

donde

$$a(T) = a_1 + a_2 \cdot (T - 25) + a_3 \cdot (T - 25)^2 \quad (III.48)$$

$$b(T) = b_1 + b_2 \cdot (T - 25) + b_3 \cdot (T - 25)^2 \quad (III.49)$$

De este modo, los autores encuentran que una única ecuación, con un mismo juego de parámetros, es capaz de predecir la dependencia lineal con la temperatura de la capacidad calorífica del estado nativo y la dependencia no lineal del estado desnaturalizado. En la tabla III.3 se recoge el juego de parámetros resultantes del ajuste de la base de datos.

**Tabla III.3: Parámetros resultantes del ajuste global de la base de datos a la función que describe la capacidad calorífica.**

$\nu$	0.28	[cal·(K·g) <sup>-1</sup> ]	$\omega$	$9.75 \times 10^{-4}$	[cal·(K <sup>2</sup> ·g) <sup>-1</sup> ]
$p$	$8.7 \times 10^{-3}$	[cal·(K·mol·Å <sup>2</sup> ) <sup>-1</sup> ]	$q$	$6.43 \times 10^{-4}$	[cal·(K <sup>2</sup> ·mol·Å <sup>2</sup> ) <sup>-1</sup> ]
$a_1$	0.45	[cal·(K·mol·Å <sup>2</sup> ) <sup>-1</sup> ]	$b_1$	-0.265	[cal·(K·mol·Å <sup>2</sup> ) <sup>-1</sup> ]
$a_2$	$2.63 \times 10^{-4}$	[cal·(K <sup>2</sup> ·mol·Å <sup>2</sup> ) <sup>-1</sup> ]	$b_2$	$2.85 \times 10^{-4}$	[cal·(K <sup>2</sup> ·mol·Å <sup>2</sup> ) <sup>-1</sup> ]
$a_3$	$-4.2 \times 10^{-5}$	[cal·(K <sup>3</sup> ·mol·Å <sup>2</sup> ) <sup>-1</sup> ]	$b_3$	$4.31 \times 10^{-5}$	[cal·(K <sup>3</sup> ·mol·Å <sup>2</sup> ) <sup>-1</sup> ]

Los resultados obtenidos de este análisis (Gómez et al., 1995) indican que la capacidad calorífica absoluta de las distintas conformaciones de una proteína puede calcularse de forma precisa a partir de parámetros estructurales en un amplio intervalo de temperatura. Como se ha discutido en trabajos anteriores (Privalov & Makhatadze, 1990), la capacidad calorífica del estado desplegado puede obtenerse a partir de las contribuciones individuales de las cadenas laterales de los distintos aminoácidos y de los enlaces peptídicos. Sin embargo, la capacidad calorífica del

estado nativo o de los estados parcialmente plegados en disolución no presentan esta aditividad, de modo que no pueden ser determinados a partir de la secuencia, aunque, sí a partir de su estructura tridimensional.

Los cambios de capacidad calorífica asociados a los procesos de plegamiento y desplegamiento de proteínas están determinados por dos contribuciones fundamentales: la hidratación de los grupos polares y apolares que inicialmente se encuentran ocultos en el interior de la proteína y que se exponen al disolvente como consecuencia del desplegamiento, y la ruptura de interacciones no covalentes características del estado nativo.

La capacidad calorífica asociada a la hidratación apolar es de signo positivo y, de acuerdo con los parámetros obtenidos del ajuste global, disminuye desde  $0.45 \text{ cal}\cdot(\text{K}\cdot\text{mol}\cdot\text{Å}^2)^{-1}$  a  $25 \text{ °C}$  hasta  $0.23 \text{ cal}\cdot(\text{K}\cdot\text{mol}\cdot\text{Å}^2)^{-1}$  a  $100 \text{ °C}$ . Por otro lado, la capacidad calorífica de hidratación polar es negativa, variando desde  $-0.25 \text{ cal}\cdot(\text{K}\cdot\text{mol}\cdot\text{Å}^2)^{-1}$  a  $25 \text{ °C}$  hasta casi cero a  $100 \text{ °C}$ . Es interesante hacer notar que los valores obtenidos a  $25 \text{ °C}$  para ambos son similares a los determinados anteriormente sin considerar la dependencia con la temperatura (Murphy et al., 1992; Xie & Freire, 1994a). Las interacciones no covalentes contribuyen positivamente a la capacidad calorífica del estado nativo, por lo que su ruptura en el proceso de desplegamiento representa una contribución negativa al cambio en capacidad calorífica.

### III.2.4.-Contribuciones electrostáticas a la energía de Gibbs.

#### III.2.4.1.-Contribuciones electrostáticas.

Tanto la estabilidad estructural de una proteína, como las constantes de afinidad de numerosos ligandos, presentan frecuentemente una fuerte dependencia con el pH, por lo que es conveniente prestar una especial atención a la evaluación de las contribuciones de naturaleza electrostática a la energía libre de Gibbs. La energía electrostática asociada a una carga determinada en la proteína puede evaluarse como el trabajo requerido para trasladar dicha partícula cargada desde un estado de referencia, aislado de la macromolécula, hasta su posición real en la proteína. En general, son dos los factores principales que determinan las contribuciones electrostáticas a la energía de Gibbs: las interacciones coulombicas y los efectos de hidratación.

Las interacciones coulombicas, establecidas entre dos partículas cargadas, representan la influencia de un átomo cargado sobre la energía potencial de un segundo, y pueden calcularse como el trabajo requerido para situar una carga en una posición determinada en la macromolécula, en contra del potencial electrostático originado por las demás cargas existentes en el sistema:

$$W_{ij} = 332 \int_0^{z_i} \frac{z_j \cdot dq}{r_{ij} \cdot A} = \frac{332 \cdot z_i \cdot z_j}{r_{ij} \cdot A} \quad (\text{III.50})$$

donde  $r_{i,j}$  representa la distancia de separación entre las cargas  $i$  y  $j$  en Å,  $Z_i$  y  $Z_j$  son las cargas que interaccionan entre sí, y  $A$  es el parámetro de atenuación, que incorpora los efectos dieléctricos y de apantallamiento.

Para evaluar las contribuciones procedentes de los fenómenos de hidratación, por el contrario, es necesario estimar la energía propia o interacciones carga-medio, que representan el efecto que la carga de un átomo tiene sobre su propio potencial. Este término, que representa la energía almacenada en el campo eléctrico de un ión, puede calcularse a partir del trabajo total correspondiente al proceso de carga de una partícula de radio  $r_i$ :

$$W_{ii} = 332 \cdot \int_0^{Z_i} \frac{q \cdot dq}{r_i \cdot A} = \frac{332 \cdot Z_i^2}{2r_i \cdot A} \quad (\text{III.51})$$

Las contribuciones electrostáticas se calculan directamente como energías libres, sin recurrir a la estimación individual de sus componentes entálpicos y entrópicos, de acuerdo con el procedimiento descrito por García-Moreno (García-Moreno, 1995; García-Moreno et al., 1997), utilizando la ecuación:

$$\Delta G_{el} = \sum_i \frac{332 \cdot Z_i^2}{2 \cdot r_i} \cdot \left( \frac{1}{A} - \frac{1}{A_{ref}} \right) + 332 \cdot Z_i \cdot \sum_j \frac{Z_j}{A \cdot r_{ij}} \quad (\text{III.52})$$

donde  $Z$  es la carga,  $r_i$  el radio de la partícula cargada,  $r_{ij}$  la separación entre ambas cargas y  $A$  y  $A_{ref}$  los parámetros de atenuación en la molécula y en el estado de referencia, respectivamente, que incorporan los efectos dieléctricos y de apantallamiento.

La constante dieléctrica de una proteína es un factor clave en la determinación de la magnitud de las interacciones electrostáticas en proteínas. La elevada constante dieléctrica del disolvente acuoso tiene su origen en su naturaleza líquida y en su elevada densidad de dipolos. El interior de una proteína, por el contrario, está caracterizado por una constante dieléctrica de mucha menor magnitud, dado que está constituido por un número elevado de átomos apolares, que se encuentran agrupados y protegidos del disolvente debido al efecto hidrofóbico, y de grupos polares, en su mayoría dispuestos en orientaciones semirígidas y parcialmente neutralizados mediante el establecimiento de enlaces de hidrógeno.

El valor de la constante dieléctrica en el interior de la proteína es un parámetro crítico para la evaluación de los efectos electrostáticos a partir de cálculos estructurales que consideren un modelo continuo del disolvente. En nuestro caso, el valor concreto de la constante dieléctrica efectiva, utilizada en los cálculos procede del ajuste de los datos de la Temperatura de transición

estructural ( $T_m$ ) de diversas proteínas en función del pH y es cercano a 16. La magnitud de la constante dieléctrica obtenida de este modo, se encuentra en concordancia con los resultados experimentales descritos por García-Moreno y colaboradores (García-Moreno, 1995; García-Moreno et al., 1997), obtenidos mediante el análisis de las alteraciones observadas en los pKas de grupos ionizables ocultos en el interior de la proteína. En todos los casos, tanto las cargas asignadas a cada grupo ionizable como las constantes dieléctricas efectivas, se encuentran ponderados de acuerdo con su grado de exposición al disolvente.

### III.2.4.2.-Efectos de protonación.

Los efectos de protonación pueden incorporarse una vez conocidos los valores de pKa de los grupos que cambian de estado de ionización como consecuencia del plegamiento o del proceso de asociación. En proceso de plegamiento/desplegamiento de proteínas estos cambios en el grado de protonación entre ambos estados conformacionales son bastante comunes, encontrándose el equilibrio frecuentemente asociado a la protonación o desprotonación de uno o más grupos en la proteína. De ser este el caso, la contribución de los efectos de protonación a la energía libre de Gibbs de cualquier estado conformacional  $j$  puede expresarse como:

$$K_b(\text{pH}) = K_{b,\text{int}} \cdot \frac{10^{(\text{pK}_a - \text{pH})}}{(1 + 10^{(\text{pK}_a - \text{pH})})} \quad (\text{III.53})$$

donde  $\text{pK}_{j,i}$  representa el pK del grupo ionizable  $i$  en el estado  $j$  y  $\text{pK}_{N,i}$  el pK del grupo  $i$  en el estado de referencia. Por tanto, de acuerdo con esta ecuación, la contribución adicional de la protonación a  $\Delta G$  a cualquier pH puede calcularse si se conoce el pKa de cada grupo ionizable para la conformación considerada.

En numerosos procesos de unión, la asociación del ligando a la proteína se encuentra acoplada a la protonación obligatoria de uno o más grupos, bien en la proteína, o bien en el ligando. En este caso, la constante de afinidad global se define como:

$$\Delta G_{\text{ion}} = -R \cdot T \cdot \ln \left( \prod_i \frac{(1 + 10^{\text{pK}^{j,i} - \text{pH}})}{(1 + 10^{\text{pK}^{N,i} - \text{pH}})} \right) \quad (\text{III.54})$$

donde  $K_{b,\text{int}}$  es la constante de unión bajo condiciones experimentales en las que el protón se encuentra unido a la proteína previamente a la unión del ligando. De acuerdo con esta ecuación, a cualquier pH, la contribución adicional de los efectos de protonación a la energía libre de Gibbs puede calcularse en términos del pKa del grupo de acuerdo con la ecuación:



$$\Delta G_{\text{ion}} = -R \cdot T \cdot \ln \frac{10^{(\text{pK}_a - \text{pH})}}{(1 + 10^{(\text{pK}_a - \text{pH})})} \quad (\text{III.55})$$

De forma análoga pueden derivarse las ecuaciones que describen las contribuciones a la afinidad de unión de la ionización obligatoria de uno o más grupos como consecuencia de la unión, de modo que:

$$K_b(\text{pH}) = K_{b, \text{int}} \cdot \frac{1}{(1 + 10^{(\text{pK}_a - \text{pH})})} \quad (\text{III.56})$$

donde ahora  $K_{b, \text{int}}$  es la constante de unión bajo condiciones en las que el grupo se encuentra ionizado con anterioridad a la unión del ligando. De forma similar, a cualquier pH, las contribuciones adicionales a la energía libre debidas a la protonación se describen como:

$$\Delta G_{\text{ion}} = R \cdot T \cdot \ln (1 + 10^{(\text{pK}_a - \text{pH})}) \quad (\text{III.57})$$

### **III.2.5.-Contribuciones traslacionales a la energía de Gibbs.**

La asociación de dos o más moléculas se traduce en una reducción en el número de grado de libertad traslacionales disponibles para cada una de ellas. La magnitud exacta de este término (Murphy et al., 1994; Kauzmann, 1959; Novotny et al., 1989; Janin, 1995) e incluso la necesidad de su consideración (Gilson et al., 1997) han sido motivo de discusión. En el proceso de desarrollo de la parametrización estructural de la energética de proteínas utilizada en este trabajo, se observó que el valor que mejor se ajusta a los resultados experimentales es la denominada entropía crítica, definida por Kauzmann (Kauzmann, 1959).

De este modo, el cambio en los grados de libertad traslacionales debido a la reducción en el número de partículas en disolución se refleja en la entropía de mezcla, de modo que el cambio entrópico correspondiente a la formación de 1 mol de complejo en una reacción de asociación de estequiometría :



viene dado por la expresión:

$$\Delta S_{\text{mix}} = (i + j + \dots + k - 1) \cdot R \cdot \ln X + R \cdot (i \cdot \ln i + j \cdot \ln j + \dots + k \cdot \ln k) \quad (\text{III.59})$$

siempre y cuando la disolución sea lo suficientemente diluida para poder ser considerada como ideal. En esta ecuación, X representa la fracción molar del complejo. Si las concentraciones que determinan el equilibrio se definen en unidades molares, entonces el cambio en entropía corresponde a la formación de un mol de producto en un litro de disolución. En este caso,  $X = 1/55.5$  y, por tanto, en una reacción de unión simple de la forma  $A + B \rightarrow P$ ,  $\Delta S_{\text{mix}} = -8 \text{ cal} \cdot (\text{K} \cdot \text{mol})^{-1}$  mientras que para una reacción de la forma  $A + 2B \rightarrow P$   $\Delta S_{\text{mix}} = -13.2 \text{ cal} \cdot (\text{K} \cdot \text{mol})^{-1}$  por complejo formado.

Algunos autores han tratado el cambio en los grados de libertad traslacionales asociados a la unión en términos de la ecuación de Sackur-Tetrode (Searle et al., 1992; Finkelstein & Janin, 1989), que se evalúa a partir de los estados cuánticos de la partícula en la caja, asumiendo que ésta posee volumen nulo y no presenta interacciones intermoleculares. Para moléculas con masas moleculares similares a las proteínas, la disminución en entropía traslacional se predice cercana a  $-25 \text{ kcal} \cdot \text{mol}^{-1}$  para la formación de un complejo bimolecular, valor tres veces mayor que el término correspondiente a la entropía crítica.

Experimentalmente, el cambio en entropía traslacional podría evaluarse si se conoce la energética de un sistema dimérico que pudiese estudiarse bajo condiciones disociativas y no disociativas (cross-link covalente) siempre que las interacciones intermoleculares no variaran en ambos casos:

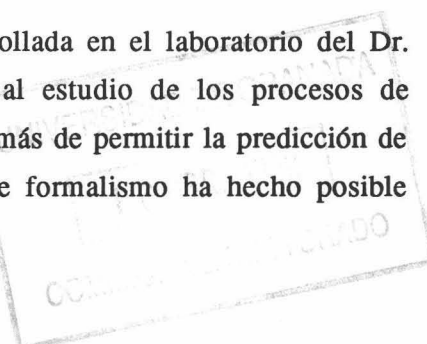
$$A_{\text{pleg}} - A_{\text{pleg}} \approx 2 A_{\text{despleg}} \quad (\text{III.60})$$

$$A_{\text{pleg}} - A_{\text{pleg}} \approx A_{\text{despleg}} - A_{\text{despleg}} \quad (\text{III.61})$$

El estudio llevado a cabo por Antonson y colaboradores (Antonson et al., 1995) sobre la estabilidad de un dominio coiled-coil de laminina en presencia y ausencia de un puente disulfuro constituye una buena aproximación a la situación descrita en el párrafo anterior. En este caso, el cambio en entalpía es similar para ambas especies reflejando el origen entrópico de las diferencias en energía libre. Los autores observan que bajo las condiciones estudiadas (todas las actividades fijas a 1 M), la diferencia en energía libre de Gibbs para las dos reacciones es  $2.63 \text{ kcal} \cdot \text{mol}^{-1}$ , valor muy cercano a la entropía crítica de Kauzmann.

### III.2.6.-Diversas aplicaciones.

La parametrización estructural de la energética, desarrollada en el laboratorio del Dr. Freire durante los últimos años, se ha aplicado con éxito al estudio de los procesos de plegamiento/desplegamiento de proteínas. En este contexto, además de permitir la predicción de las magnitudes termodinámicas asociadas con el proceso, este formalismo ha hecho posible

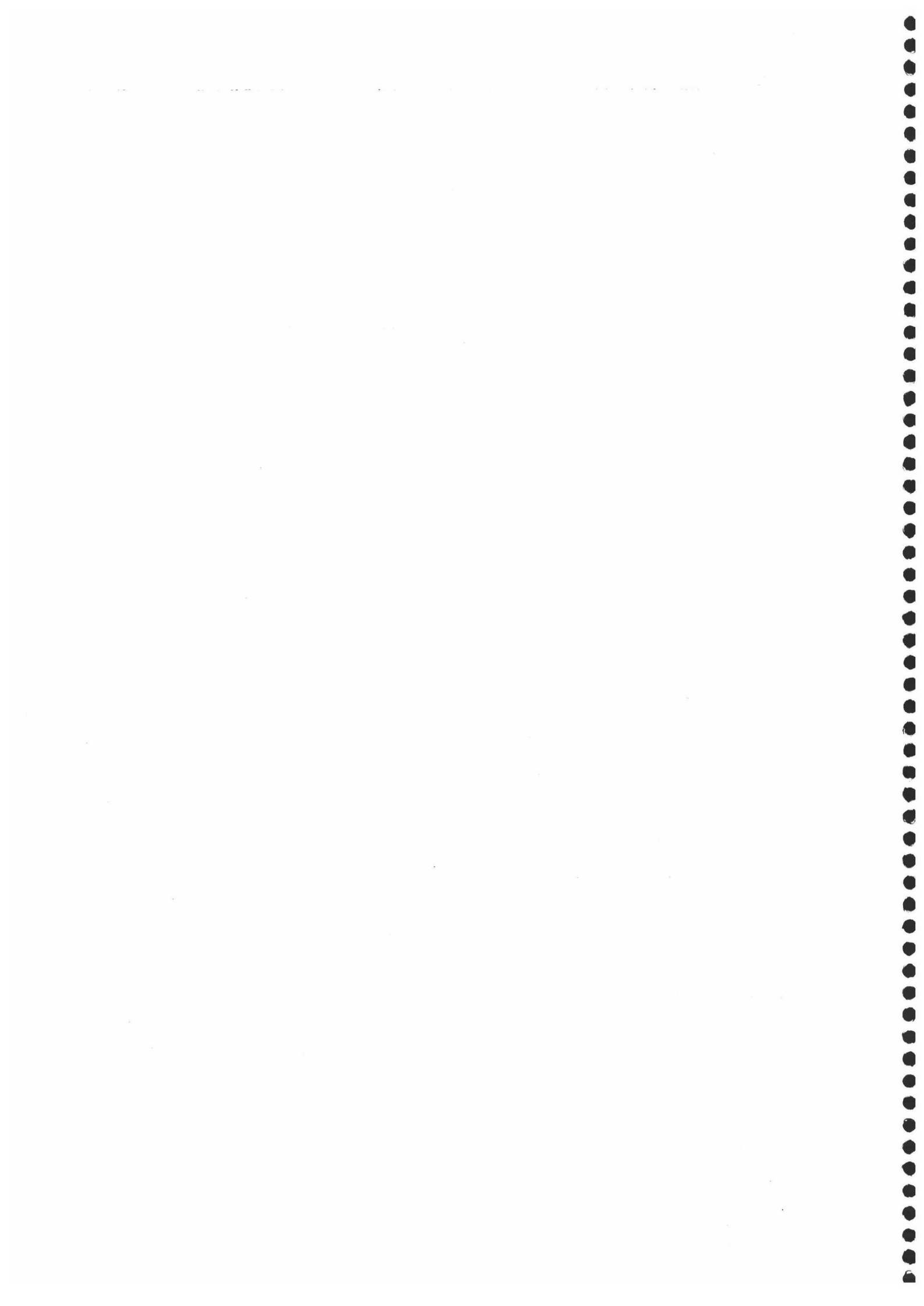


realizar el análisis de las bases moleculares de la cooperatividad en estos sistemas (Freire et al., 1993; Freire & Murphy, 1991; Murphy et al., 1992; Xie & Freire, 1994a; Xie et al., 1993). En este contexto, se han identificado las principales propiedades, tanto termodinámicas como estructurales, que caracterizan a los posibles intermedios termodinámicos en el proceso de plegamiento (Freire & Xie, 1994; Xie et al., 1994; Xie & Freire, 1994a; Xie & Freire, 1994b).

Asímismo, además de poseer la capacidad para reproducir las magnitudes termodinámicas correspondientes al proceso global de plegamiento/desplegamiento, esta parametrización ha permitido estudiar los efectos de fenómenos asociados a aminoácidos particulares, como las propensiones helicoidales (Luque et al., 1996), objeto de este trabajo y que se describe con detalle en la memoria en la sección IV.A, o el cálculo de las constantes de estabilidad por resto, que permiten predecir los patrones de los factores de protección al intercambio hidrógeno/deuterio, determinados por resonancia magnética nuclear (Hilser & Freire, 1996a; Hilser & Freire, 1996b; Hilser et al., 1997).

En cuanto a los procesos de unión de ligandos, se ha abordado el estudio de procesos como la asociación Angiotensina II/anticuerpo (Murphy et al., 1993) o la unión del inhibidor peptídico Pepstatina A (Gómez & Freire, 1995) y otros inhibidores de renina (Gómez et al., 1997) a la proteasa aspártica Endothiapepsina. Estos procedimientos permiten identificar los restos de la enzima y del ligando que contribuyen en mayor grado a la afinidad de unión, facilitando el desarrollo de estrategias de diseño molecular, como se describirá en la sección IV.B, de este trabajo.

## **IV.-RESULTADOS Y DISCUSIÓN**



#### IV.A.-OBTENCIÓN DE UNA ESCALA TERMODINÁMICO-ESTRUCTURAL DE PROPENSIÓN HELICOIDAL. OPTIMIZACIÓN DE LA PARAMETRIZACIÓN ESTRUCTURAL DE LA ENERGÉTICA DE PLEGAMIENTO DE PROTEÍNAS Y UNIÓN DE LIGANDOS.

##### IV.A.1.-Consideraciones generales.

El área accesible al disolvente en cada uno de los estados implicados en el proceso bajo estudio es una de las principales variables estructurales consideradas en la parametrización estructural de la energética de plegamiento de proteínas y de la unión de sus ligandos, desarrollada durante los últimos años en el laboratorio del Dr. Freire y descrita en la sección III.2. Las diferencias en accesibilidad se encuentran directamente relacionadas con los cambios en la hidratación y, de modo indirecto, con la ruptura y/o establecimiento de interacciones internas (contactos de van der Waals y enlaces de hidrógeno) que tienen lugar durante el proceso bajo estudio. Generalmente dichos cambios de accesibilidad se expresan en términos de áreas accesibles al disolvente (ASA) (Connolly, 1983; Lee & Richards, 1971; Shrake & Rupley, 1973), definidas en el apéndice A.1, donde se proporciona una descripción más detallada de las distintas definiciones de las superficies moleculares así como de los algoritmos de cálculo de mayor interés.

Diversos grupos han demostrado la existencia de cierta correlación entre los cambios en áreas accesibles y las distintas magnitudes termodinámicas (Gómez et al., 1995; Murphy et al., 1992; Murphy & Freire, 1992; Ooi et al., 1987; Spolar et al., 1992; Spolar & Record Jr, 1994), de modo que éstas varían proporcionalmente a los cambios en accesibilidad asociados al proceso en cuestión. Sin embargo, aún a pesar de su evidente interés, la superficie accesible al disolvente, especialmente la correspondiente a las conformaciones desplegadas, no es una cantidad definida de forma unívoca, por lo que su estimación se convierte en el factor limitante de la precisión con que las magnitudes termodinámicas pueden calcularse a partir de datos estructurales.

En esta sección se describe la aplicación de la parametrización estructural de la energética de plegamiento de proteínas y unión de ligandos al cálculo de las propensiones helicoidales de cada uno de los aminoácidos. Dado que, en este caso, es necesario evaluar el efecto de simples substituciones de un aminoácido por otro en la estabilidad de hélices peptídicas o de proteínas, la precisión con que se realice el cálculo del área accesible al disolvente en ambas conformaciones es una cuestión crítica para el éxito de los resultados.

Las áreas accesibles al disolvente correspondientes a las conformaciones nativas se calculan generalmente mediante la aplicación de alguno de los algoritmos de cálculo (ver apéndice A.1.) a la estructura tridimensional de la macromolécula, obtenida mediante cristalografía de rayos X o resonancia magnética nuclear. Para estimar la accesibilidad de conformaciones no estructuradas, como las correspondientes al estado desplegado de la proteína,



o a ligandos peptídicos libres en disolución, es necesario recurrir a modelos estructurales estándar, que generalmente corresponden a cadenas polipeptídicas en conformación completamente extendida. En ambos casos, el cálculo se realiza a partir de un única estructura estática, aún a pesar de que la accesibilidad al disolvente de una cadena polipeptídica en disolución debe encontrarse modulada por sus posibles cambios conformacionales. Estos implican, desde las fluctuaciones conformacionales de la estructura nativa, generalmente provocados por los movimientos de las cadenas laterales situadas en su superficie, hasta los producidos por la rotación de los ángulos diedros del esqueleto peptídico y de las cadenas laterales en el estado desplegado.

En general, bajo condiciones de equilibrio, la superficie accesible al disolvente de un aminoácido en una cadena polipeptídica no es un único valor constante, sino una cantidad promedio determinada, por las contribuciones de todas las posibles conformaciones adoptadas por la molécula de acuerdo con su probabilidad. De este modo, admitiendo para la energía una distribución de Boltzmann entre los distintos estados conformacionales, la probabilidad de que una cadena polipeptídica adopte una conformación determinada,  $P(\phi, \psi, \chi_i)$ , es una función de los ángulos diedros correspondientes al esqueleto peptídico, así como de los ángulos de torsión de las cadenas laterales, de acuerdo con la expresión:

$$P(\phi, \psi, \chi_i) = \frac{\exp[-E(\phi, \psi, \chi_i)/kT]}{\sum \exp[-E(\phi, \psi, \chi_i)/kT]} \quad (\text{IV.1})$$

donde la sumatoria en el denominador define la función de partición conformacional,  $Q$ ,  $k$  es la constante de Boltzmann y  $T$  la temperatura absoluta del sistema.

Por lo tanto, una vez conocido el perfil energético en función de cada uno de los ángulos diedros, el promedio de Boltzmann de la superficie accesible al disolvente,  $\langle \text{ASA} \rangle$ , puede estimarse de acuerdo con la ecuación estándar:

$$\langle \text{ASA} \rangle = \sum \text{ASA}(\phi, \psi, \chi_i) \cdot P(\phi, \psi, \chi_i) = \frac{\sum \text{ASA}(\phi, \psi, \chi_i) \cdot \exp[-E(\phi, \psi, \chi_i)/kT]}{\sum \exp[-E(\phi, \psi, \chi_i)/kT]} \quad (\text{IV.2})$$

Es de esperar que la utilización de estos valores de accesibilidad promedio permitan mejorar la precisión con que puede llevarse a cabo el cálculo de las distintas magnitudes termodinámicas, lo cual presenta un especial interés si, como ocurre en nuestro caso, se pretende evaluar pequeños cambios en accesibilidades, como es el caso de los asociados a mutaciones en hélices expuestas al disolvente, que permitirían el cálculo de las propensiones helicoidales de cada uno de los aminoácidos.

## IV.A.2.-Métodos

### IV.A.2.1.-Cálculo de áreas accesibles promedio en conformación helicoidal.

En el caso de conformaciones nativas, en las cuales las posibilidades de rotación (tanto de los ángulos diedros de la cadena principal como de los correspondientes a las cadenas laterales enterradas en el interior de la proteína) se encuentran muy restringidas, puede considerarse, como primera aproximación, que las posibles fluctuaciones en la accesibilidad al disolvente están fundamentalmente ocasionadas por los cambios en las conformaciones adoptadas por las cadenas laterales que se encuentran localizadas en la superficie molecular.

Numerosos estudios conformacionales han demostrado la existencia de cierta correlación entre la localización y extensión de las regiones conformacionales accesibles a una cadena lateral y la conformación adoptada por el esqueleto peptídico. Esta dependencia es especialmente notable en el caso de conformaciones helicoidales, para las cuales se producen las mayores restricciones conformacionales en los ángulos  $\chi_1$  y  $\chi_2$  de la cadena lateral como consecuencia de la localización de los átomos del esqueleto (McGregor et al., 1987).

Con objeto de poder realizar posteriormente el cálculo de las propensiones helicoidales de los distintos aminoácidos con suficiente precisión, se procedió al cálculo del promedio de Boltzmann de las áreas accesibles al disolvente de cada uno de los aminoácidos en conformación helicoidal. Para ello se requiere la generación de una distribución significativa de conformaciones estéricamente accesibles al sistema, así como el consiguiente cálculo de la energía correspondiente a cada una de ellas, lo que permitirá, de acuerdo con las ecuaciones anteriormente expuestas, la estimación de la función de partición y de sus respectivas probabilidades.

#### IV.A.2.1.1.-*Construcción de las estructuras helicoidales.*

Todos los cálculos se han llevado a cabo utilizando hélices estándar de nueve restos. Las hélices, de secuencia  $(Ala)_4-X-(Ala)_4$ , donde X representa cada uno de los veinte aminoácidos con la excepción de Pro, se construyeron utilizando el módulo Biopolymers del programa InsightII (Biosym) y fijando los ángulos diedros de la cadena principal a los valores  $\phi = -65^\circ$  y  $\psi = -40^\circ$ .

Para cada una de estas hélices se procedió a la generación de un extenso juego de conformaciones mediante la modificación sistemática de los ángulos diedros de las cadenas laterales. En el caso de cadenas laterales con un único ángulo diedro (Cys y Val)<sup>2</sup> se ha empleado un intervalo de variación de  $1^\circ$  entre  $0^\circ$  y  $360^\circ$ , generándose, por tanto, 360 conformaciones diferentes. Para aquéllas con dos ángulos diedros el número de conformaciones generadas es 1296, al considerarse intervalos de variación para los ángulos  $\chi_1$  y  $\chi_2$  de  $10^\circ$ . En el caso de

<sup>2</sup> En el caso de los aminoácidos Ser y Thr se ha considerado la rotación del hidrógeno Gamma con respecto al oxígeno Beta, por lo que se definen dos ángulos diedros efectivos para sus cadenas laterales.

aminoácidos con cadenas laterales de mayor longitud, los ángulos posteriores a  $\chi_2$  se mantienen constantes a un valor correspondiente a la conformación completamente extendida ( $\chi_3, \chi_4, \chi_5 = 180^\circ$ ).

#### IV.A.2.1.2.-Obtención de los perfiles de accesibilidad.

Para cada una de las conformaciones así generadas, se procede al cálculo del área superficial accesible al disolvente mediante la aplicación del algoritmo de Lee y Richards (Lee & Richards, 1971), utilizando un radio para la molécula de disolvente de 1.4 Å y un espesor corte transversal de 0.25 Å. De esta forma, se obtienen los perfiles de accesibilidad para cada uno de los aminoácidos en una hélice alfa en función de los ángulos diedros de su cadena lateral.

No todas las conformaciones generadas de este modo son posibles, debido a impedimentos de tipo estérico; por lo que aquellas conformaciones que impliquen colisiones entre las esferas de van der Waals de algunos de sus átomos (es decir, distancias interatómicas menores que la suma de los radios de van der Waals correspondientes a los átomos implicados) son rechazadas. El juego de radios empleado para llevar a cabo la selección de conformaciones se muestra en la Tabla IV.A.1. En este punto, a partir únicamente de los perfiles de accesibilidad, es posible el cálculo de la media aritmética de la accesibilidad del aminoácido en posición central, considerando exclusivamente las conformaciones estéricamente permitidas.

**Tabla IV.A.1:** Radios de van der Waals utilizados para identificar las conformaciones con problemas estéricos.

(Å)			
<b>C carbonílico</b>	1.35	<b>C alifático</b>	1.35
<b>N</b>	1.27	<b>S</b>	1.02
		<b>O</b>	1.20

#### IV.A.2.1.3.-Obtención de los perfiles de probabilidad.

Una vez estimada la accesibilidad al disolvente característica de cada una de las conformaciones adoptadas por la cadena lateral, el cálculo de su probabilidad requiere la asignación de su correspondiente energía. Para ello se ha recurrido a los perfiles energéticos calculados previamente por Lee et al. (Lee et al., 1994), que fueron obtenidos utilizando los parámetros y funciones potenciales propuestos por Weiner y colaboradores (Weiner et al., 1984) y por Jorgensen y Tirado-Rives (Jorgensen & Swenson, 1985), e incluyendo interacciones electrostáticas entre cargas parciales, para cuyo cálculo se asignó un valor de 78 a la constante dieléctrica del disolvente.

Los cálculos energéticos fueron realizados con hélices  $(Ala)_4$ -X- $(Ala)_4$ , idénticas a las empleadas para la estimación de las accesibilidades, siendo el procedimiento seguido para la generación de conformaciones completamente paralelo al empleado para el cálculo de las áreas. Por tanto, se dispone del valor de área accesible y de energía para cada una de las conformaciones generadas de acuerdo con el procedimiento descrito anteriormente.

Una vez obtenidos los valores de energía, la estimación de los pesos estadísticos, o factores de Boltzmann, de cada una de las conformaciones, así como de la función de partición, permite el cálculo de las probabilidades, definidas como:

$$P_i = \frac{\exp[-E_i(\chi_1, \chi_2)/kT]}{Q} = \frac{\exp[-E_i(\chi_1, \chi_2)/kT]}{\sum_i \exp[-E_i(\chi_1, \chi_2)/kT]} \quad (IV.3)$$

donde  $E_i$  representa la energía de cada conformación,  $k$  la constante de Boltzmann,  $T$  la temperatura absoluta del sistema, y  $Q$  la función de partición conformacional.

**IV.A.2.1.4.-Cálculo de los promedios de Boltzmann para las áreas accesibles en conformación helicoidal.**

Por definición, el promedio de Boltzmann para el área accesible al disolvente de cada uno de los aminoácidos puede obtenerse como el producto de la accesibilidad de cada conformación por su probabilidad, de acuerdo con la expresión:

$$\langle ASA \rangle = \sum_i ASA_i(\chi_1, \chi_2) \cdot P_i = \frac{\sum_i ASA_i(\chi_1, \chi_2) \cdot \exp[-E_i(\chi_1, \chi_2)/kT]}{\sum_i \exp[-E_i(\chi_1, \chi_2)/kT]} \quad (IV.4)$$

Las desviaciones estándar correspondientes a los valores promedio obtenidos de esta manera se calculan como:

$$\sigma = \sqrt{\frac{\langle ASA^2 \rangle - \langle ASA \rangle^2}{n - 1}} \quad (IV.5)$$

donde  $n$  representa el número de conformaciones consideradas y  $\langle ASA^2 \rangle$  se obtiene como:

$$\langle ASA^2 \rangle = \sum_i ASA_i(\chi_1, \chi_2)^2 \cdot P_i = \frac{\sum_i ASA_i(\chi_1, \chi_2)^2 \cdot \exp[-E_i(\chi_1, \chi_2)/kT]}{\sum_i \exp[-E_i(\chi_1, \chi_2)/kT]} \quad (IV.6)$$

#### **IV.A.2.1.5.-Influencia de cadenas laterales vecinas en la accesibilidad de un aminoácido determinado.**

Con objeto de estimar la influencia que las cadenas laterales de los aminoácidos vecinos pudieran tener sobre los valores promedio de las áreas accesibles al disolvente en conformación helicoidal, calculados como se ha descrito en la sección anterior, se ha aplicado el mismo procedimiento a las hélices de secuencia Leu-Asn-Ala-Ala-Lys-X-Glu-Leu-Asp-Lys-Ala-Ile, correspondientes a la hélice 39-50 de la proteína Lisozima T4. Para este sistema se dispone, además de la estructura tridimensional de alta resolución de la proteína nativa (2LZM.pdb y 3LZM.pdb), de las estructuras resultantes tras la mutación del resto Ser en posición 44 por varios aminoácidos (S44R: 113l.pdb, S44N: 111l.pdb, S44E: 217l.pdb, S44I: 108l.pdb, S44K: 109l.pdb, S44L: 110l.pdb, S44F: 137l.pdb, S44P: 112l.pdb, S44T: 114l.pdb, S44W: 216l.pdb, S44V: 115l.pdb, S44A: 1164.pdb). A pesar de que los cambios globales en accesibilidad y estabilidad de esta hélice son distintos de los correspondientes a las hélices modelo, las diferencias relativas en esta posición son comparables con las calculadas para las hélices estándar, ya que el sitio de mutación se encuentra en el centro de la hélice y completamente expuesto al disolvente (Blaber et al., 1994).

#### **IV.A.2.2.-Áreas accesibles al disolvente en conformaciones desplegadas.**

La accesibilidad de cada aminoácido en el estado desplegado de una proteína se ha estimado tradicionalmente como su área accesible al disolvente en diversos modelos estándar, que supuestamente reflejan de forma aproximada la situación encontrada en una proteína desnaturalizada, o en cadenas polipeptídicas no estructuradas. Entre los más frecuentemente utilizados pueden destacarse los tripéptidos de secuencia Ala-X-Ala o Gly-X-Gly en conformación completamente extendida.

Como se ha discutido en la sección III.2.4, estas representaciones constituyen una pobre representación del estado desplegado de una proteína debido a que, además de ignorar los efectos que la presencia del resto de la cadena peptídica pueda tener en la accesibilidad de un aminoácido determinado, proporcionan una descripción estática del sistema. En este contexto, algunos estudios recientes han demostrado que la aplicación de estos modelos conduce a la sobrestimación de los cambios en áreas accesibles al disolvente asociados a procesos de plegamiento de proteínas (Creamer et al., 1995; Creamer et al., 1997).

El valor de la superficie accesible al disolvente de los distintos aminoácidos en el estado desplegado de una proteína depende tanto de los ángulos de las cadenas laterales como de los correspondientes al esqueleto peptídico. Aunque, en principio, la accesibilidad de un aminoácido concreto puede verse condicionada por la conformación adoptada por los demás restos de la

cadena, puede considerarse, en una primera aproximación, que, en ausencia de estructura secundaria residual, la conformación de cada aminoácido en el estado desplegado es independiente de la adoptada por el resto de la cadena. Como consecuencia, es posible considerar que el área accesible al disolvente de un aminoácido determinado en el estado desplegado dependerá fundamentalmente de los valores adoptados por sus propios ángulos diedros.

Con objeto de estimar los cambios en accesibilidad de los distintos aminoácidos como consecuencia de las fluctuaciones conformacionales de su esqueleto peptídico, se procedió al cálculo del área accesible al disolvente de cada aminoácido en la posición central de un tripéptido de secuencia Ala-X-Ala en función de sus ángulos diedros  $\phi$  y  $\psi$ .

#### IV.A.2.2.1.-Construcción de los tripéptidos.

Los tripéptidos se construyeron en conformación completamente extendida utilizando el módulo Biopolymers del programa InsightII (Biosym). Los ángulos diedros correspondientes al esqueleto peptídico de los aminoácidos situados en los extremos del tripéptido se fijan a  $180^\circ$  y todas las cadenas laterales se construyen con los valores correspondientes a las conformaciones completamente trans.

#### IV.A.2.2.2.-Cálculo de perfiles de accesibilidad en el estado desplegado.

Para cada uno de los tripéptidos construidos se procedió a la generación de diferentes conformaciones mediante la modificación sistemática de los ángulos diedros del esqueleto peptídico del aminoácido central a intervalos de  $10^\circ$ , siguiendo un procedimiento similar al descrito anteriormente para las conformaciones helicoidales. Al igual que en dicho caso, no todas las conformaciones generadas son posibles debido a los impedimentos estéricos impuestos tanto por el esqueleto de la cadena peptídica como por las cadenas laterales, por lo que las conformaciones que presentan solapamientos entre las esferas de van der Waals son rechazadas.

Conviene señalar la elevada sensibilidad que la localización y extensión de las regiones conformacionales permitidas presentan con respecto a la asignación de los radios de van der Waals de cada tipo de átomo. En nuestro caso, se han utilizado los radios propuestos por Leach y colaboradores (Leach et al., 1966), que se muestran en la tabla IV.A.2, para llevar a cabo la selección de las conformaciones permitidas.

**Tabla IV.A.2:** Radios de van der Waals utilizados para identificar las conformaciones con problemas estéricos.

(Å)			
C carbonílico	1.35	C alifático	1.35
N	1.27	S	1.02
		O	1.20



### IV.A.3.-Resultados y discusión.

#### IV.A.3.1.-Cálculo de áreas accesibles promedio en conformación helicoidal.

##### IV.A.3.1.1.-Perfiles de accesibilidad y probabilidad.

El procedimiento seguido para llevar a cabo el cálculo de las áreas accesibles al disolvente promedio en la conformación helicoidal se ilustra en la figura IV.A.1 para el caso concreto del aminoácido Fenilalanina. La gráfica en el panel superior izquierdo representa el perfil de energía, calculado por Lee y colaboradores (Lee et al., 1994) en función de los ángulos  $\chi_1$  y  $\chi_2$  de la cadena lateral. Los valores de energía se representan en forma de curvas de nivel con respecto al mínimo en energía, espaciadas a intervalos de 2 kcal·mol<sup>-1</sup>. En el panel superior derecho se muestra el correspondiente perfil de probabilidad, donde las líneas de contorno representan las conformaciones con valores de probabilidad 10<sup>-5</sup>, 10<sup>-4</sup>, 10<sup>-3</sup>, 3 · 10<sup>-3</sup>, 5 · 10<sup>-3</sup>, 7 · 10<sup>-3</sup>, 10<sup>-2</sup>, 2 · 10<sup>-2</sup>, 3 · 10<sup>-2</sup>, 4 · 10<sup>-2</sup> y 5 · 10<sup>-2</sup>. En la parte inferior de la figura se representan los perfiles de áreas accesibles polares y apolares correspondientes a dicho aminoácido.

##### IV.A.3.1.2.-Áreas accesibles promedio.

En la tabla IV.A.3 se muestran los valores de áreas accesibles apolares (sección superior) y polares (sección inferior) para los distintos aminoácidos en la posición central de la hélice. La primera columna recoge el área accesible al disolvente de la conformación correspondiente al mínimo energético para cada uno de los aminoácidos. La segunda columna muestra los promedios de Boltzmann y sus correspondientes desviaciones estándar, mientras que en la tercera se recogen los valores correspondientes a la media aritmética resultante de promediar las accesibilidades de todas las conformaciones permitidas estéricamente. Los mismos valores se representan en la figura IV.A.II, en la que se puede observar la similitud existente entre los valores de accesibilidad en el mínimo de energía potencial (representados mediante barras negras) y los correspondientes promedios de Boltzmann (barras blancas). Asimismo, las pequeñas diferencias existentes entre el promedio ponderado energéticamente (promedios de Boltzmann) y los valores de la media aritmética (correspondientes a las barras rayadas de la figura), indican la ausencia de grandes variaciones en accesibilidad dentro de los límites del espacio conformacional accesible a las distintas cadenas laterales.

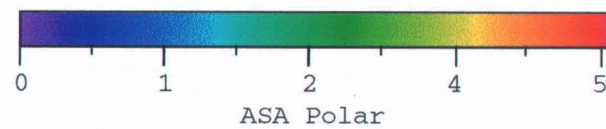
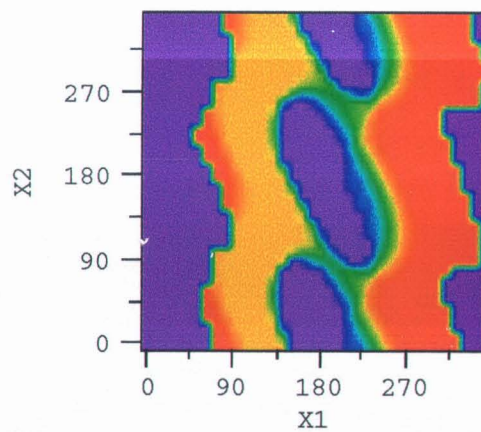
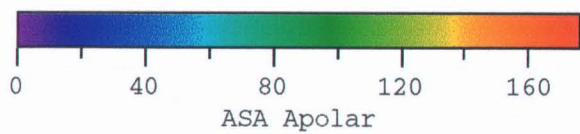
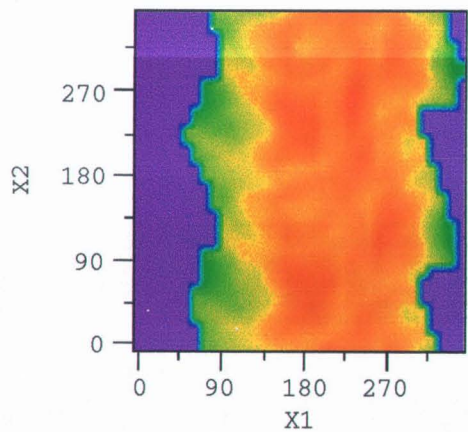
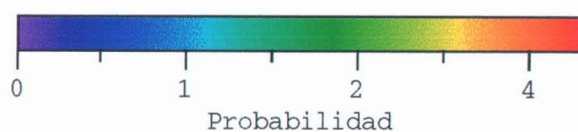
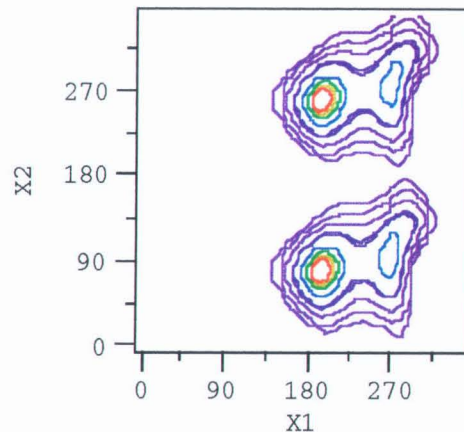
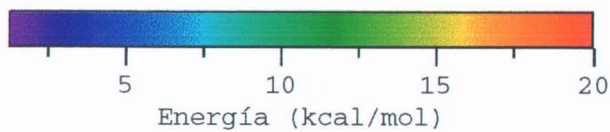
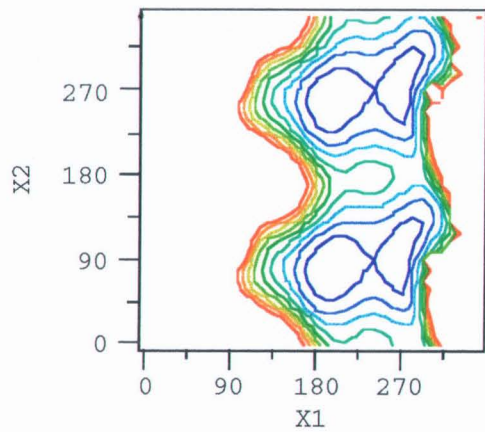
La última columna incluida en la tabla IV.A.1 muestra los promedios de Boltzmann, obtenidos a partir de las estructuras correspondientes a los distintos mutantes de Lisozima T4. Como se puede observar, estos datos indican la ausencia de diferencias significativas en cuanto a la accesibilidad promedio de cada aminoácido en la conformación helicoidal como consecuencia de la presencia de cadenas laterales más o menos voluminosas en posiciones cercanas en la hélice.

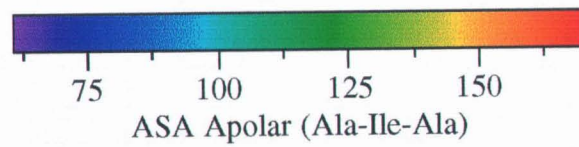
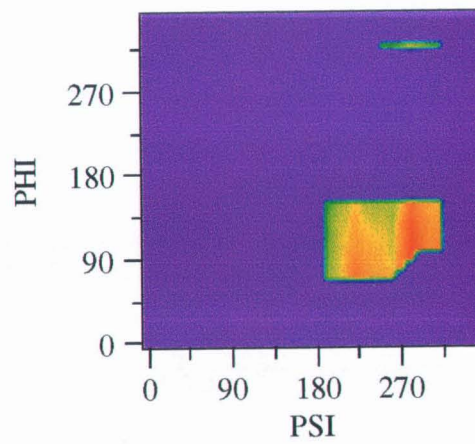
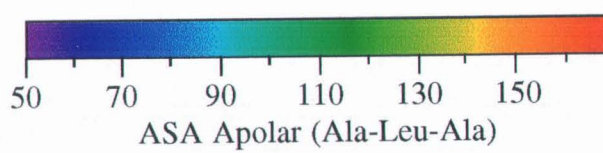
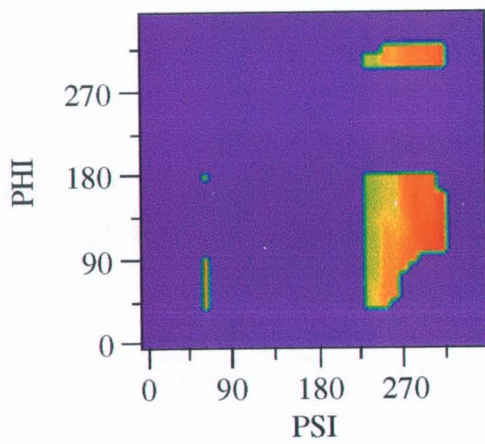
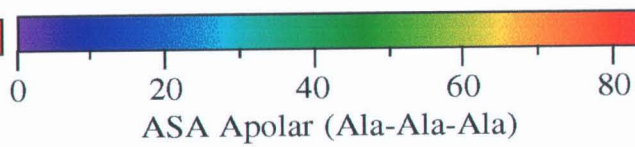
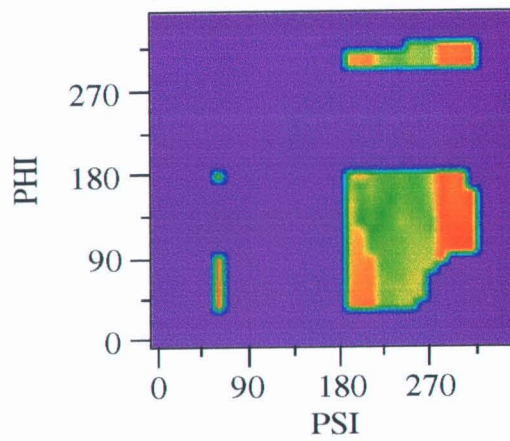
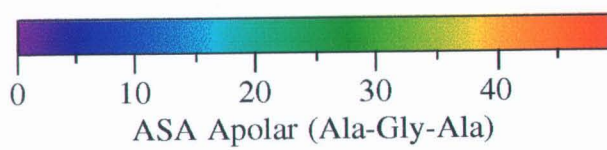
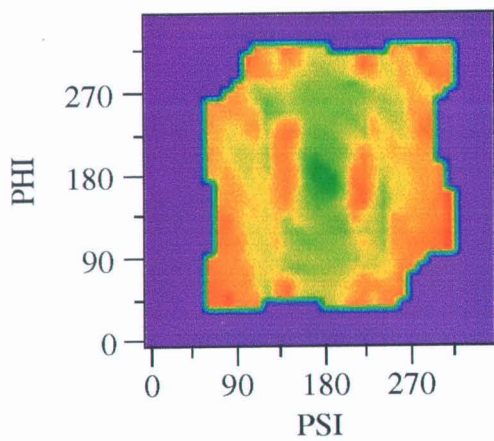
Tabla IV.A.3: Valores promedio de ASA para los distintos aminoácidos en la posición central de una hélice alfa.

Aminoácido	ASA <sub>ap</sub> , min	ASA <sub>ap</sub> , Boltz	ASA <sub>ap</sub> , num	ASA <sub>ap</sub> , Lisoz
Ala	65.50	65.50 ± 0.00	65.50 ± 0.00	51.29
Arg	70.23	70.21 ± 0.13	71.04 ± 7.44	75.37
Asn	27.62	28.59 ± 0.12	33.92 ± 6.51	31.00
Asp	33.16	37.09 ± 0.14	41.44 ± 5.22	-----
Cys	42.64	50.00 ± 0.13	53.65 ± 8.97	-----
Glu	49.34	62.94 ± 0.28	60.24 ± 8.80	39.46
Gly	39.80	39.80 ± 0.00	39.80 ± 0.00	31.84
His	63.83	80.05 ± 0.28	79.99 ± 9.91	-----
Ile	112.35	116.64 ± 0.17	109.47 ± 10.92	106.32
Leu	126.91	122.16 ± 0.16	101.70 ± 14.92	105.2
Lys	96.46	96.45 ± 0.12	90.73 ± 11.72	80.94
Met	93.23	97.98 ± 0.32	97.38 ± 15.14	-----
Phe	162.00	160.83 ± 0.18	145.12 ± 16.50	127.82
Ser	36.74	35.48 ± 0.09	39.45 ± 6.71	-----
Thr	67.77	63.89 ± 0.30	57.86 ± 12.44	-----
Trp	176.42	165.88 ± 0.39	57.05 ± 18.87	140.8
Tyr	137.24	133.22 ± 0.18	116.37 ± 15.53	-----
Val	97.52	85.17 ± 0.71	93.81 ± 11.13	94.15

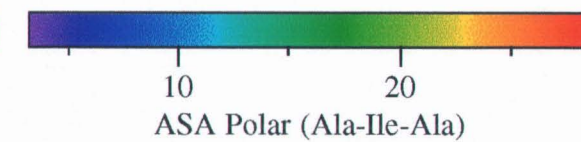
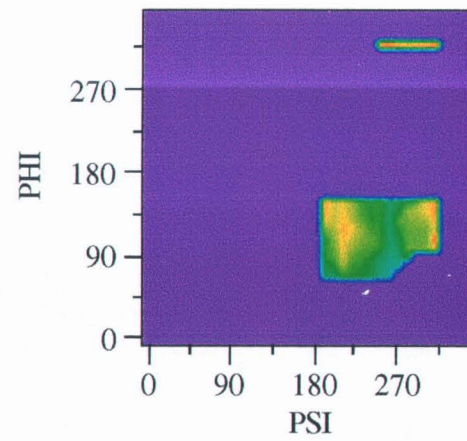
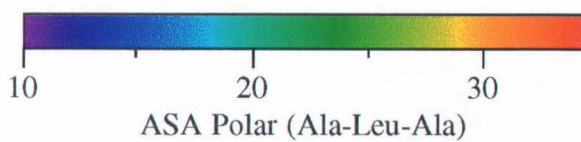
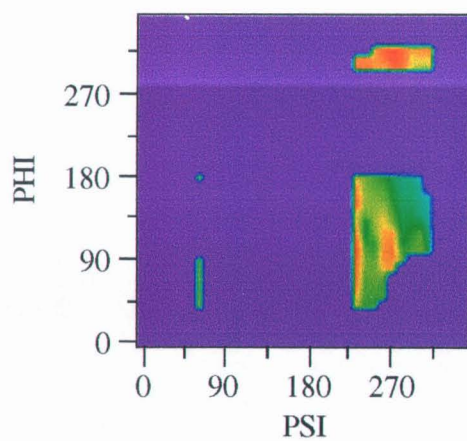
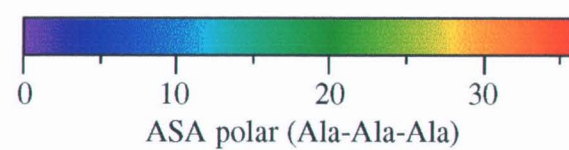
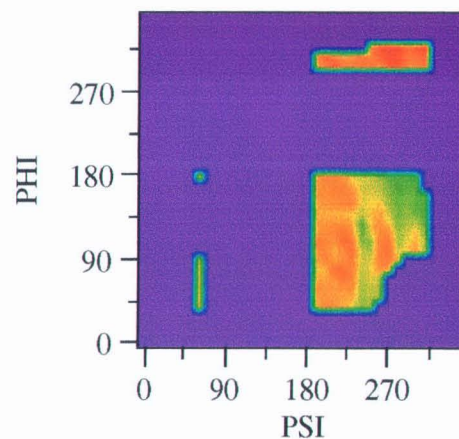
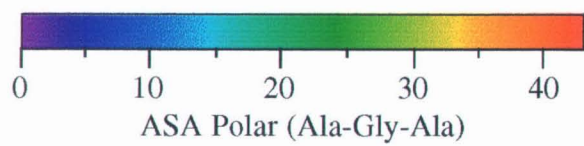
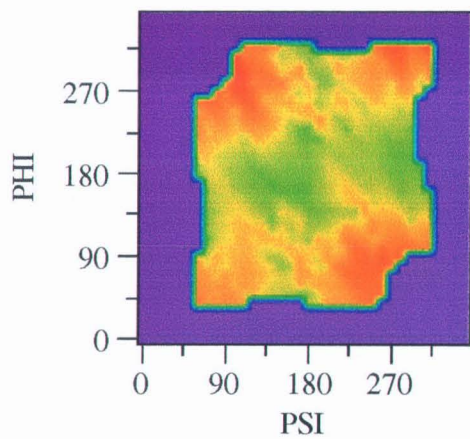
  

	ASA <sub>pol</sub> , min	ASA <sub>pol</sub> , Boltz	ASA <sub>pol</sub> , num	ASA <sub>pol</sub> , Lisoz
Ala	4.80	4.80 ± 0.00	4.80 ± 0.00	4.49
Arg	116.93	117.12 ± 0.14	101.77 ± 17.99	84.59
Asn	79.04	76.32 ± 0.23	57.96 ± 16.34	65.74
Asp	69.39	65.56 ± 0.15	49.24 ± 14.14	-----
Cys	59.44	63.92 ± 0.31	51.81 ± 17.72	-----
Gln	77.39	79.31 ± 0.12	64.37 ± 17.06	-----
Glu	67.46	61.28 ± 0.29	55.17 ± 13.91	60.66
Gly	9.10	9.10 ± 0.00	9.10 ± 0.00	5.28
His	63.42	57.2 ± 0.21	48.19 ± 11.76	-----
Ile	3.17	3.10 ± 0.02	3.21 ± 1.49	0.00
Leu	5.02	3.88 ± 0.01	3.40 ± 1.92	3.15
Lys	51.80	48.39 ± 0.15	47.39 ± 10.57	41.09
Met	41.63	40.29 ± 0.14	29.87 ± 13.48	-----
Phe	2.46	2.55 ± 0.04	3.27 ± 1.92	0.00
Ser	34.88	36.33 ± 0.01	4.82 ± 0.36	-----
Thr	14.06	16.36 ± 0.01	4.24 ± 0.69	-----
Trp	32.25	32.22 ± 0.01	28.39 ± 6.82	20.89
Tyr	43.76	43.14 ± 0.04	42.17 ± 4.03	-----
Val	3.86	3.37 ± 0.01	4.08 ± 0.76	0.00









observar la notable variación en la extensión de las regiones del espacio conformacional permitidas en el tripéptido Ala-Ala-Ala con respecto al que contiene Gly en la posición central. La presencia de cadenas laterales de mayor tamaño, aunque restringe adicionalmente las posibilidades de rotación de los ángulos  $\phi$  y  $\psi$ , tiene efectos mucho más leves en la libertad conformacional del esqueleto peptídico, incluso en aquellos casos en que existen ramificaciones en la cadena lateral.

Los promedios numéricos de las accesibilidades polar y apolar de cada aminoácido, obtenidos considerando únicamente las conformaciones permitidas estéricamente, se recogen en la tabla IV.A.3, junto con los valores de sus respectivas desviaciones estándar y el área del correspondiente aminoácido en el tripéptido en conformación completamente extendida. Se puede observar que las variaciones en las áreas accesibles al disolvente entre las distintas conformaciones son pequeñas, siendo la desviación estándar promedio del 15% para el área apolar y del 16% para la polar.

#### IV.A.3.3.-Escala de propensión helicoidal.

Dada la pequeña magnitud de las variaciones en accesibilidad al disolvente entre las distintas conformaciones del tripéptido, se decidió intentar refinar los valores de áreas accesibles al disolvente de los distintos aminoácidos en el estado desplegado. Con este fin se procedió a la aplicación de la parametrización estructural de la energética de plegamiento de proteínas al estudio de las propensiones helicoidales de los distintos aminoácidos, que han sido medidas experimentalmente en sistemas de naturaleza muy diversa (desde péptidos sintéticos hasta proteínas) como se ha descrito en la sección II.A.3.

Estos cálculos se realizaron con un doble objetivo: por un lado, comprobar la posibilidad de predecir los valores de  $\Delta\Delta G$ , procedentes de los distintos estudios experimentales, utilizando un único juego de valores de áreas accesibles al disolvente para los aminoácidos en el estado desplegado y, por otro, comprobar si los valores empleados, capaces de reproducir los datos experimentales, se encuentran comprendidos dentro del intervalo de accesibilidades que pueda presentar cada aminoácido en el espacio conformacional permitido.

Los valores de  $\Delta\Delta G$  experimentales considerados proceden del análisis de la influencia que la substitución de cada uno de los aminoácidos tiene en la estabilidad del sistema. Los sistemas considerados incluyen dos proteínas (Lisozima T4, en la que las substituciones se han llevado a cabo en posición 44 (Blaber et al., 1994), y Barnasa, de la que se dispone de datos experimentales para los mutantes en posición 32 (Horovitz et al., 1992)) y dos péptidos sintéticos (una cremallera de leucinas de 29 aminoácidos de longitud, diseñada por O'Neil y DeGrado (O'Neil & DeGrado, 1990), y un péptido helicoidal monomérico de 21 restos, estabilizado mediante numerosos puentes salinos (Lyu et al., 1990)). Es interesante hacer notar que en todos



los casos, los sitios de substitución se encuentran en posiciones expuestas al disolvente y en las que la

*Tabla 4.A.4: Valores promedio de ASA para los distintos aminoácidos en posiciones centrales de tripéptidos Ala-X-Ala.*

Aminoácido	ASA <sub>ap, ex</sub>	ASA <sub>ap, num</sub>	ASA <sub>ap, opt</sub>
Ala	64.7	64.7	70.0
Arg	99.5	99.5	87.1
Asn	47.2	47.2	38.1
Asp	49.7	49.7	42.1
Cys	38.1	38.1	30.3
Gln	71.2	71.2	60.5
Glu	80.9	80.9	71.1
Gly	38.9	38.9	26.2
His	112.4	112.4	90.0
Ile	148.2	148.2	110.7
Leu	149.9	149.9	122.3
Lys	124.1	124.1	101.3
Met	122.2	122.2	104.6
Phe	194.0	194.0	186.8
Ser	54.3	54.3	55.1
Thr	88.3	88.3	79.5
Trp	201.9	201.9	184.5
Tyr	165.1	165.1	175.8
Val	132.4	132.4	88.7

Aminoácido	ASA <sub>pol, ex</sub>	ASA <sub>pol, num</sub>	ASA <sub>pol, opt</sub>
Ala	29.2	29.2	36.1
Arg	143.9	143.9	126.1
Asn	106.6	106.6	104.0
Asp	97.6	97.6	95.0
Cys	95.3	95.3	75.1
Gln	111.8	111.8	112.5
Glu	91.1	91.1	94.3
Gly	34.0	34.0	43.1
His	77.6	77.6	68.0
Ile	19.8	19.8	10.9
Leu	27.5	27.5	26.6
Lys	69.4	69.4	79.0
Met	70.1	70.1	64.0
Phe	27.0	27.0	36.5
Ser	59.2	59.2	81.9
Thr	51.8	51.8	41.1
Trp	56.8	56.8	52.3
Tyr	60.3	60.3	71.1
Val	18.7	18.7	17.8

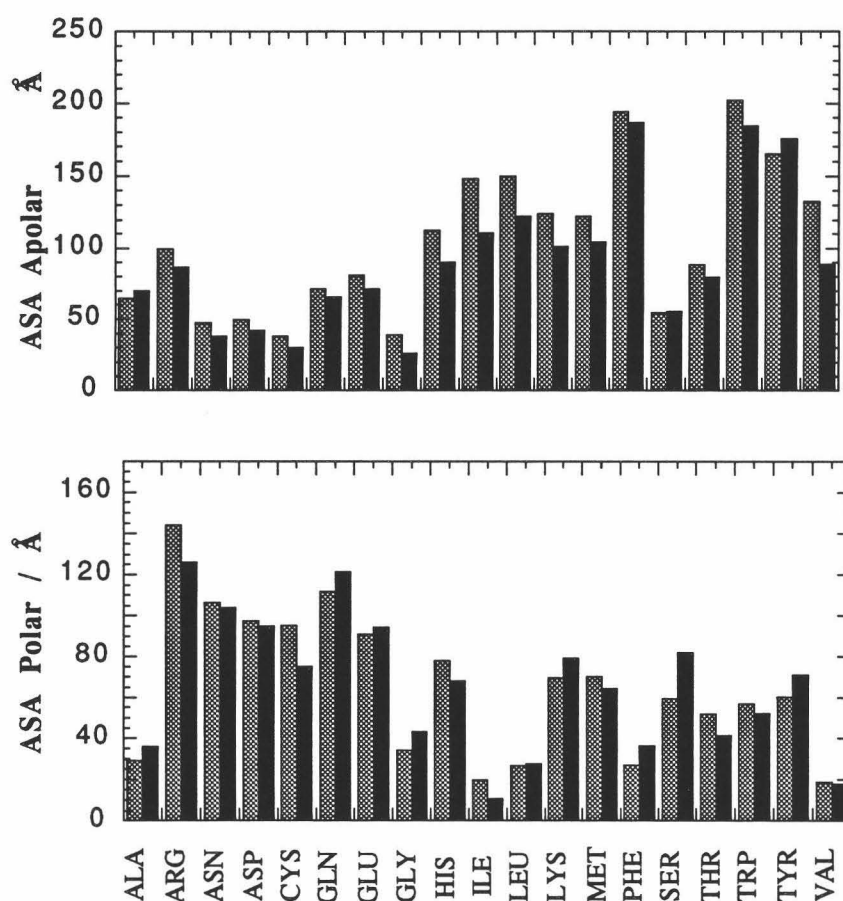
posibilidad de establecimiento de interacciones terciarias o específicas con otras cadenas laterales es muy pequeña. Esto implica que los efectos que las mutaciones pudieran tener en las interacciones internas características del estado nativo son muy reducidos, maximizándose, sin embargo, los efectos de la sustitución en el estado desplegado.

Estos datos experimentales no sólo representan sistemas de distinta naturaleza, sino que, además, han sido obtenidos mediante diversas técnicas y procedimientos de análisis a distintas temperaturas (para una descripción más detallada de cada sistema ver sección II.A.3.). De este modo, las propensiones helicoidales procedentes de las sustituciones llevadas a cabo en Lisozima T4 se han obtenido a 52 °C, las de Barnasa a 25 °C, los datos para la cremallera de leucinas a 20 °C y los del péptido monomérico a 4 °C. Los valores de  $\Delta\Delta G$  para cada uno de los aminoácidos en los distintos sistemas se muestran en la tabla II.A.1.

#### *IV.A.3.3.1.-Cálculo de propensiones helicoidales. Refinado de los valores de accesibilidad en el estado desplegado.*

Los cambios en energía libre asociados a cada mutación se calcularon de acuerdo con la parametrización estructural de la energética descrita en la sección III.2 y utilizando el mismo juego de coeficientes en todos los cálculos. Los valores de áreas accesibles polares y apolares para cada aminoácido en el estado nativo utilizados para llevar a cabo los cálculos de  $\Delta\Delta G$  son los correspondientes a los promedios de Boltzmann, obtenidos previamente y mostrados en la segunda columna de la tabla IV.A.3. Únicamente los valores de las áreas accesibles polares y apolares de cada aminoácido en el estado desplegado se variaron dentro del intervalo permitido por el análisis conformacional realizado con anterioridad, identificándose aquellos valores que minimizan las diferencias entre los valores calculados y los experimentales.

El mejor juego de valores de área accesible al disolvente para cada aminoácido en conformaciones desplegadas se muestra, junto con el valor medio de las accesibilidades correspondientes a todas las conformaciones permitidas estéricamente, en la figura IV.A.5. En general, los valores de las accesibilidades apolares son de menor magnitud que las correspondientes a la media aritmética de las áreas accesibles al disolvente, considerando todas las conformaciones permitidas. Este comportamiento refleja la tendencia general de las cadenas polipeptídicas a proteger sus grupos apolares del contacto con el medio acuoso y, por tanto, a exponer al disolvente un área apolar menor que la correspondiente a su conformación extendida. Para los grupos polares, la tendencia opuesta, aunque existente, no resulta tan significativa. Como promedio, los valores resultantes representan aproximadamente un 85% de los correspondientes a la media aritmética, en concordancia con los resultados obtenidos mediante análisis teóricos del comportamiento de diversos modelos de cadenas polipeptídicas desplegadas (Creamer & Rose, 1995; Creamer et al., 1995). Estos valores son, a su vez, similares a los descritos recientemente por Wimley y colaboradores (Wimley et al., 1996).



**Figura IV.A.5:** Áreas apolares (panel superior) y polares (panel inferior) accesibles al disolvente para 19 aminoácidos en la posición central de tripéptidos Ala-X-Ala. Las barras punteadas muestran los valores correspondientes a la media aritmética de la accesibilidad de cada aminoácido en las conformaciones estéricamente permitidas. Las barras negras representan los valores de accesibilidad optimizados, resultantes del ajuste de los datos de  $\Delta\Delta G$  experimentales.

En la tercera columna de la tabla IV.A.4 se muestra el juego de valores de área accesible al disolvente en el estado desplegado, que reproducen satisfactoriamente las propensiones helicoidales de los distintos aminoácidos en todos los sistemas experimentales considerados. Se recogen, asimismo, sus desviaciones estándar. La tabla IV.A.5 contiene las diferencias en  $\Delta\Delta G$  entre los valores experimentales, y los calculados utilizando este nuevo juego de datos para las áreas accesibles al disolvente en el estado desplegado para todos los aminoácidos en los distintos sistemas. En promedio, la diferencia absoluta entre ambos es de  $0.09 \text{ kcal}\cdot\text{mol}^{-1}$  para los datos de Lisozima T4,  $0.14 \text{ kcal}\cdot\text{mol}^{-1}$  para Barnasa,  $0.11 \text{ kcal}\cdot\text{mol}^{-1}$  para la cremallera de leucinas y  $0.08 \text{ kcal}\cdot\text{mol}^{-1}$  para el péptido monomérico.

Tabla 4.A.5: Valores de  $\Delta\Delta G$  experimentales y predichos para mutaciones de aminoácidos específicos<sup>a</sup>.

Aminoácido	$\Delta\Delta G_{\text{calc}} - \Delta\Delta G_{\text{exp}}$ (cal/mol)			
	Barnase	Coiled coil	T4-44	Péptido monomérico
Ala	- 94.76	22.01	- 31.87	- 77.45
Arg	- 85.40	- 37.43	112.74	-----
Asn	42.86	203.73	- 10.94	25.50
Asp	135.55	169.08	- 8.96	-----
Cys	163.20	- 6.50	- 29.30	-----
Gln	80.94	161.56	- 200.06	- 56.88
Glu	78.66	151.37	- 12.20	-----
Gly	0.00	0.00	0.00	0.00
His	69.66	110.29	- 235.30	-----
Ile	241.00	82.10	- 367.04	151.11
Leu	103.05	18.84	- 139.10	116.36
Lys	- 19.40	- 551.35	76.08	-----
Met	32.55	106.25	- 102.66	- 54.76
Phe	246.09	31.41	- 15.89	-----
Ser	- 140.58	- 42.68	82.30	- 151.05
Thr	140.09	94.61	- 6.73	- 190.33
Trp	400.69	- 10.59	33.59	-----
Tyr	277.83	163.91	- 206.55	-----
Val	432.63	294.08	- 31.33	- 4.71

<sup>a</sup> Los valores de  $\Delta\Delta G$  han sido evaluados a las temperaturas experimentales: 52° C para los datos de Lisozima T4 (Blaber et al., 1994), 25° C para los de Barnasa (Horovitz et al., 1992), 20° C para el coiled-coil (O'Neil & DeGrado, 1990) y 4° C para el péptido monomérico (Lyu et al., 1990).

El análisis de correlación entre los valores experimentales y los calculados en este trabajo para cada uno de los sistemas se presenta en la figura IV.A.6. Para todos los juegos de datos, las pendientes varían entre 0.8 y 1.27, siendo el valor medio  $1.03 \pm 0.2$ , lo que es indicativo de la ausencia de desviaciones sistemáticas en los cálculos. Los coeficientes de correlación varían entre 0.83 y 0.91, encontrándose en el mismo intervalo descrito por otros análisis (Muñoz & Serrano, 1995b).

#### IV.A.3.3.2.-Escala de propensión helicoidal. Comparación con otros trabajos.

Como se puede observar en las gráficas incluidas en la figura IV.A.6, la escala de propensiones helicoidales para los distintos aminoácidos, desarrollada haciendo uso de la parametrización estructural de la energética de plegamiento de proteínas, muestra una excelente concordancia con las cuatro escalas experimentales mencionadas en el apartado anterior.

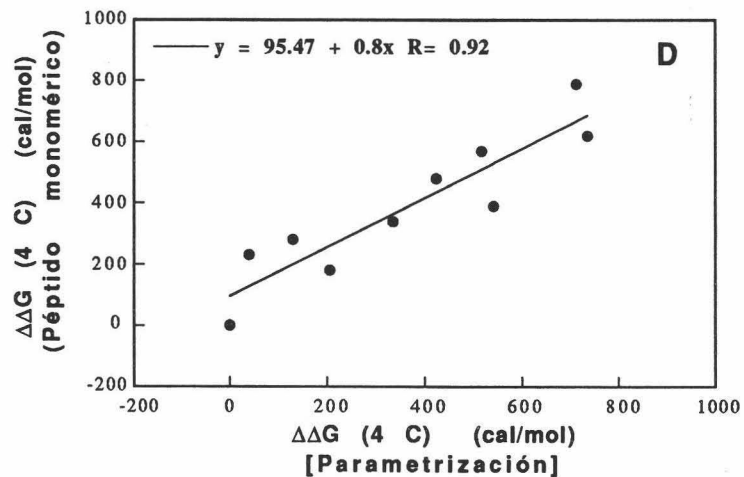
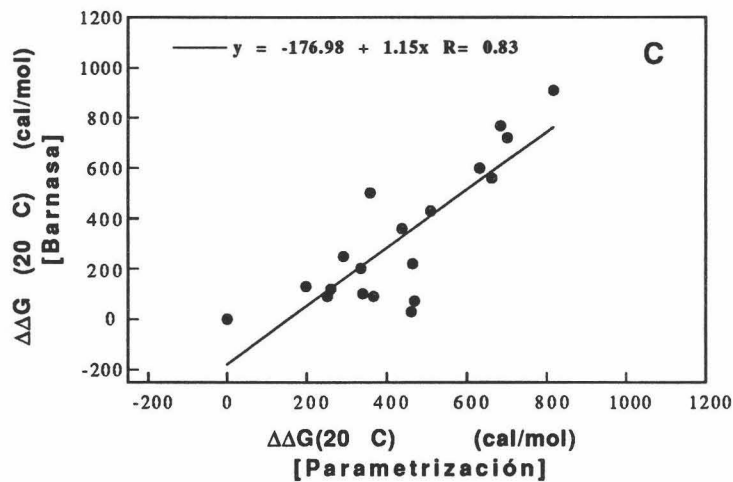
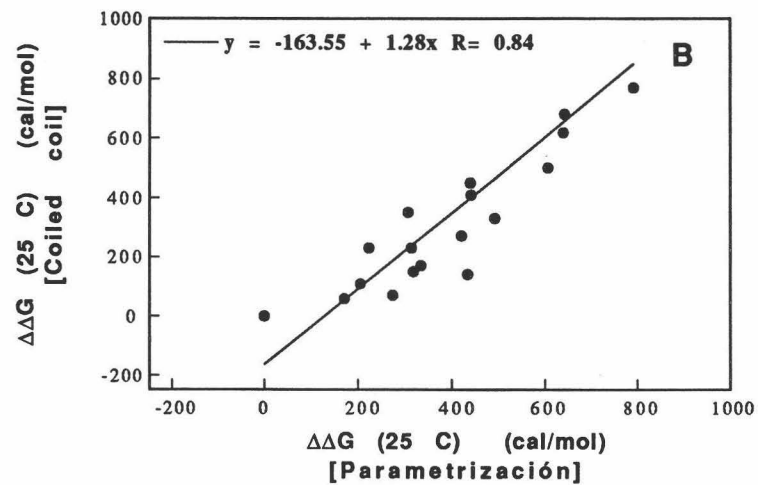
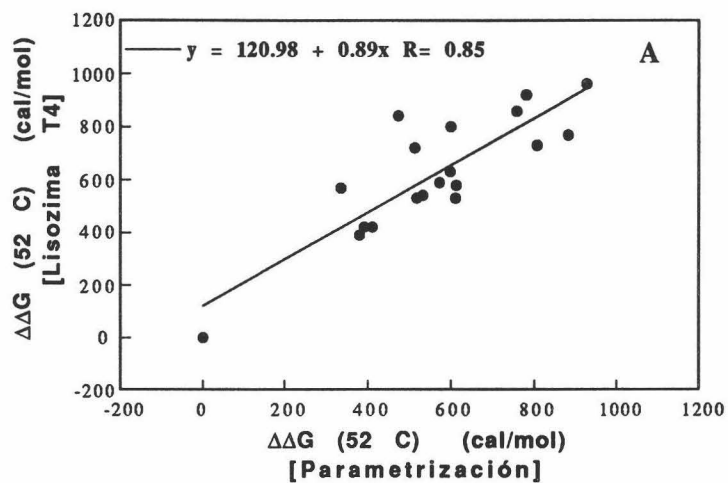


Figura IV.A.3.VI: Comparación entre los distintos juegos de datos experimentales y los valores de  $\Delta\Delta G$  calculados mediante la parametrización estructural de la energética utilizando los valores optimizados para las áreas accesibles al disolvente en el estado desplegado

Asimismo, se observa una muy buena correspondencia con otras escalas existentes en la literatura, tanto de carácter semiempírico como experimental, como se ilustra en la figura IV.A.7.

La correlación más notable es la observada entre los valores obtenidos mediante el algoritmo AGADIR, desarrollado por Muñoz y Serrano (Muñoz & Serrano, 1995b; Muñoz & Serrano, 1995c), y los calculados en este trabajo, que se representa en el panel A de la figura IV.A.8. Considerando la distinta naturaleza de los dos enfoques, la correspondencia encontrada entre los resultado que ambos proporcionan resulta especialmente interesante.

La parametrización estructural de la energética, como ya se ha discutido en el apartado III.2., es de naturaleza termodinámica y se ha desarrollado con el objeto de predecir la energía libre de Gibbs de péptidos y proteínas a partir de su estructura, y no con el propósito específico de generar una escala de propensiones helicoidales, por lo que no contiene consideraciones estructurales específicas para la formación de hélices alfa. Por el contrario, el algoritmo AGADIR está basado en las teorías de transición hélice-ovillo estadístico, y considera de forma explícita las interacciones específicas presentes en las hélices alfa en ausencia de interacciones terciarias. De este modo, los autores incluyen explícitamente en su tratamiento las distintas contribuciones procedentes de la formación de puentes de hidrógeno entre los grupos del esqueleto peptídico, las tendencias helicoidales intrínsecas de cada uno de los aminoácidos, las interacciones de "capping" en los extremos de la hélice, al igual que otras interacciones específicas, como las establecidas entre grupos cargados y el dipolo de la hélice o entre dos cadenas laterales. La correspondencia cuantitativa encontrada entre los dos procedimientos constituye una indicación de que la parametrización estructural de la energética captura de forma efectiva, aunque implícitamente, las principales interacciones implicadas en la formación de una hélice alfa.

El panel B de la figura IV.A.8 representa la correlación existente entre los valores de  $\Delta\Delta G$  calculados con nuestra parametrización estructural y los obtenidos mediante el análisis experimental, utilizando los correspondientes modelos termodinámico-estadísticos, de las transiciones hélice-ovillo estadístico en péptidos ricos en Ala (Rohl et al., 1996). Como se puede observar, la correlación entre ambos juegos de valores es buena, aunque en este caso, la pendiente de la recta de regresión es cercana a 2. Como se ha descrito en la sección II.A.4.2, los resultados experimentales obtenidos con este tipo de sistemas proporcionan sistemáticamente valores de  $\Delta\Delta G$  de mayor magnitud que los procedentes de otro tipo de estudios. Actualmente no se dispone de una explicación certera del origen de estas discrepancias, aunque algunos autores han sugerido las diferencias de volumen entre las cadenas laterales vecinas al sitio de sustitución como posibles responsables de las diferencias observadas (ver sección II.A.4.2.).

La gráfica C de la misma figura muestra la correlación entre los datos experimentales obtenidos mediante el análisis de los efectos que diversas sustituciones tienen en la estabilidad de la proteína Ribonucleasa T1 (Pace & Scholtz, 1997). En este caso, únicamente se han incluido en el análisis de correlación los valores correspondientes a los aminoácidos no cargados debido a que, de acuerdo con los autores, estos restos parecen estar implicados en interacciones específicas



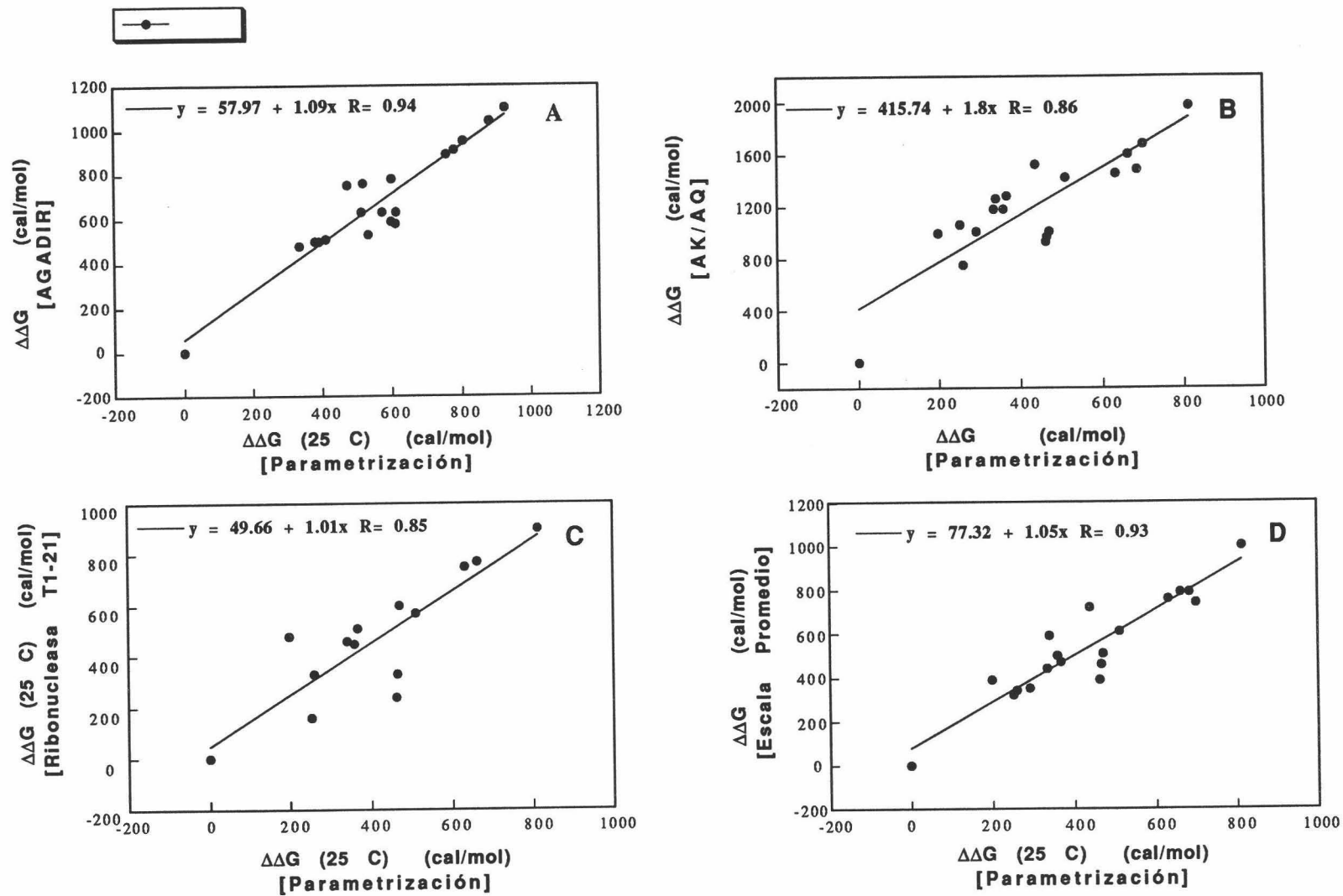


Figura IV.A.3.VII: Comparación entre varios juegos de datos de la literatura y los valores de  $\Delta\Delta G$  calculados mediante la parametrización estructural de la energética utilizando los valores optimizados para las áreas accesibles al disolvente en el estado desplegado.

con otros grupos de la proteína, perturbando, por tanto, los valores de  $\Delta\Delta G$  obtenidos experimentalmente.

Por último, la gráfica D representa la correlación existente entre las propensiones helicoidales calculadas en este trabajo y las procedentes de una escala general consenso (Pace & Scholtz, 1997), desarrollada recientemente con el fin de unificar todos los valores, tanto los obtenidos experimentalmente, como los procedentes de estudios estadísticos y semiempíricos, disponibles en la actualidad. Como puede observarse, y al igual que ocurre con los datos obtenidos con AGADIR, la correspondencia entre las dos escalas es excelente, obteniéndose una pendiente muy cercana a la unidad y un elevado coeficiente de correlación.

#### IV.A.3.4.-Análisis termodinámico-estructural de la formación de una hélice alfa.

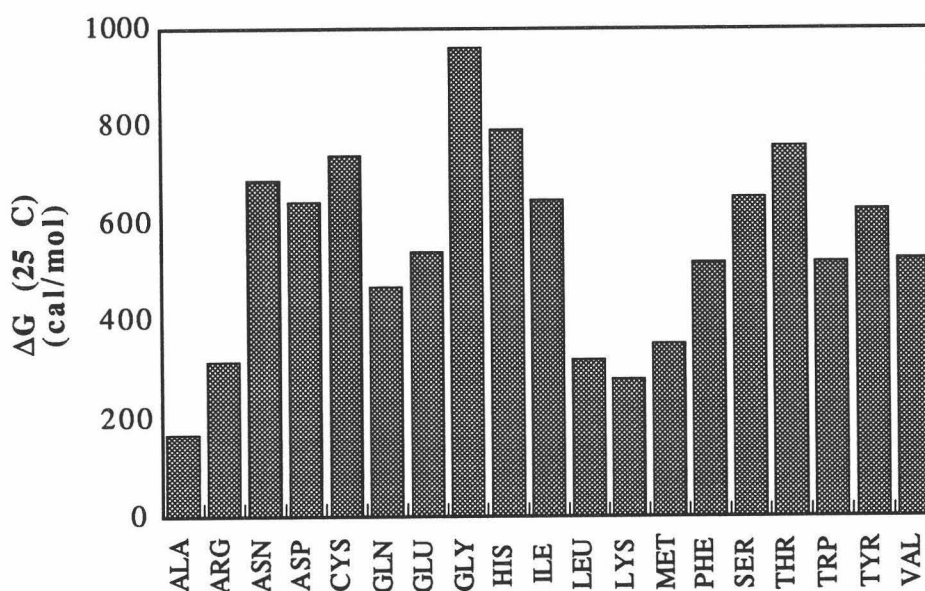
La escala de propensión helicoidal descrita hasta el momento refleja los efectos que cada aminoácido introducido en una posición helicoidal expuesta al disolvente tiene en la estabilidad del sistema con respecto a Gly. Sin embargo, el análisis termodinámico de la estabilidad de las hélices alfa requiere la evaluación de los valores absolutos de  $\Delta G$ , para lo cual es necesario estimar la energía libre correspondiente al mutante de referencia; el que contiene Gly en la posición de sustitución, en nuestro caso.

Baldwin y colaboradores (Baldwin, 1995) han obtenido un valor de  $\Delta G$  para Gly de  $1.6 \text{ kcal}\cdot\text{mol}^{-1}$  a  $0^\circ\text{C}$  mediante el análisis de los datos experimentales, obtenidos a partir del estudio de péptidos ricos en alanina. Por otro lado, el valor estimado por Lyu y colaboradores (Lyu et al., 1990) a partir del análisis de los resultados procedentes del estudio de un péptido monomérico estabilizado por puentes salinos a  $4^\circ\text{C}$  es de  $0.47 \text{ kcal}\cdot\text{mol}^{-1}$ . A estas mismas temperaturas, los valores calculados en este trabajo, mediante la aplicación de la parametrización estructural de la energética y considerando el juego de accesibilidades optimizadas para el estado desplegado, son  $0.76$  y  $0.80 \text{ kcal}\cdot\text{mol}^{-1}$ , respectivamente; valores que se encuentran dentro del intervalo determinado por los resultados experimentales.

La excelente correspondencia observada entre los valores de propensión helicoidal calculados en este trabajo y los obtenidos experimentalmente para cada aminoácido valida el empleo de la parametrización estructural de la energética para el cálculo de energías libres asociadas a procesos de desplegamiento de proteínas y unión de ligandos y permite la racionalización del origen de las diferencias en cuanto a tendencia helicoidal entre los distintos aminoácidos. Los valores de las diferentes magnitudes termodinámicas correspondientes al proceso de formación de una hélice alfa, calculados a  $25^\circ\text{C}$  para cada aminoácido, se muestran en la tabla IV.A.5. Como se puede observar, en general, la formación de una hélice alfa se encuentra favorecida entálpicamente y desfavorecida en términos entrópicos.

##### IV.A.3.4.1.-Energía libre de Gibbs.

Como se puede observar en la figura IV.A.9, en la que se muestran los valores de  $\Delta G$  calculados a 20 °C para cada aminoácido, las propensiones helicoidales de todos ellos se encuentran comprendidas en un intervalo de 0.8 kcal·mol<sup>-1</sup>, siendo Ala el aminoácido con mayor tendencia helicoidal y Gly el de carácter más fuertemente desestabilizador. La diferencia en energía libre observada entre ambos (0.8 kcal·mol<sup>-1</sup>) tiene su origen fundamentalmente en las diferencias en la entropía conformacional del esqueleto peptídico de ambos aminoácidos en el estado desplegado, que es del orden de 0.7 kcal·mol<sup>-1</sup> a 25 °C.



*Figura IV.A.3.VIII: Escala de propensión helicoidal a 25 °C calculada mediante la parametrización estructural de la energética, utilizando los valores optimizados para las áreas accesibles al disolvente en el estado desplegado.*

#### IV.A.3.4.2.-Entalpía.

En general, dado que los cambios en entalpía tienen su origen en las modificaciones de las interacciones intrínsecas y de la hidratación de los distintos grupos, provocados por el proceso de formación de la hélice, estos términos son sensibles al contexto en que se encuentra cada resto, por lo que deben calcularse para cada caso particular. Ambas contribuciones se encuentran incorporadas de forma implícita en la parametrización estructural de la energética (ver sección III.2.3.I), por lo que no se requiere su evaluación individual, lo que, como se ha discutido previamente en la sección III.2., evita tener que calcular los valores de entalpía como la diferencia

entre dos grandes cantidades de signos opuestos. La parametrización estructural predice un cambio de entalpía promedio para la formación de una hélice alfa expuesta al disolvente, y en ausencia de interacciones terciarias, de  $-0.7 \text{ kcal}\cdot(\text{mol}\cdot\text{res})^{-1}$  a  $25 \text{ }^\circ\text{C}$ , valor muy similar al derivado por Muñoz y Serrano (Muñoz & Serrano, 1995c).

Para los aminoácidos Ala y Gly, el cambio entálpico asociado a la formación de la hélice es del orden de  $-1.3 \text{ kcal}\cdot\text{mol}^{-1}$ , y tiene su origen en la formación de un enlace de hidrógeno entre los grupos del esqueleto peptídico, así como en la simultánea protección del contacto con el disolvente de una considerable porción de éste. El valor calculado para Gly y Ala a  $25 \text{ }^\circ\text{C}$  ( $-1.7 \text{ kcal}\cdot\text{mol}^{-1}$  y  $-1.2 \text{ kcal}\cdot\text{mol}^{-1}$ , respectivamente) es algo mayor que el correspondiente a los demás aminoácidos. Esto se debe fundamentalmente a la deshidratación parcial que las cadenas laterales sufren como consecuencia de la adopción de la conformación helicoidal. Esta deshidratación lleva asociado un término entálpico desfavorable, que compensa, en cierta manera, la entalpía favorable a la formación de la hélice procedente del establecimiento de los enlaces de hidrógeno.

La diferencia observada entre las entalpías obtenidas para Gly y Ala puede atribuirse a los efectos que la presencia del carbono beta en el resto de Ala tiene en la accesibilidad de su esqueleto peptídico. Esta diferencia en entalpía está prácticamente compensada en su totalidad por la diferencia en entropía de solvatación entre ambos ( $T\cdot\Delta\Delta S_{\text{solv}} = 0.65 \text{ kcal}\cdot(\text{K}\cdot\text{mol})^{-1}$ , lo que convierte a la diferencia en entropía conformacional del esqueleto peptídico de ambos aminoácidos (debida fundamentalmente a las restricciones conformacionales impuestas por el carbono beta del aminoácido Ala en los grados de libertad del esqueleto peptídico) en la principal responsable de la distintas propensiones helicoidales mostradas por estos dos aminoácidos.

Como ya se ha comentado, para los demás aminoácidos, la adopción de la conformación helicoidal se encuentra acompañada por una mayor pérdida de accesibilidad al disolvente, debido a la deshidratación parcial de las cadenas laterales. En estos casos existe una pequeña contribución adicional, de signo positivo, a la entalpía, que se encuentra acompañada de su correspondiente término entrópico compensatorio, y que se traduce en una disminución de la entalpía correspondiente a la formación de la hélice con respecto a Gly y Ala.

En la actualidad, únicamente se dispone de medidas calorimétricas directas de la entalpía de formación de una hélice alfa para el aminoácido Ala, para el que se han obtenido valores comprendidos entre  $-1.0$  y  $-1.3 \text{ kcal}\cdot\text{mol}^{-1}$  (Baldwin, 1995; Scholtz et al., 1991), muy cercanos a los predichos por los cálculos realizados en este trabajo.

#### IV.A.3.4.3.-Entropía.

En general, la pérdida de entropía conformacional del esqueleto peptídico y de las cadenas laterales como consecuencia del plegamiento contribuye desfavorablemente a la formación de la hélice. Dicha disminución en la entropía conformacional no es equivalente para todos los aminoácidos, de modo que existen diferencias significativas entre ellos. Como consecuencia, esta

magnitud contribuye notablemente a la determinación de las diferentes propensiones helicoidales de cada aminoácido (Creamer & Rose, 1992).

De este modo, como ya se ha comentado con anterioridad, la diferencia en propensión helicoidal existente entre Ala y Gly puede atribuirse casi completamente a la diferencia en entropía conformacional de sus respectivos esqueletos peptídicos en el estado desplegado. Para los demás aminoácidos, sin embargo, la disminución entrópica que conlleva el plegamiento de la cadena se encuentra parcialmente compensada por el incremento experimentado por la entropía del disolvente, a consecuencia de la deshidratación parcial de las cadenas laterales. A pesar de todo, sin embargo, el cambio total en entropía asociado a la formación de una hélice alfa es generalmente negativo y, por tanto, desfavorable al proceso de plegamiento.

Por tanto, la contribución entrópica a la propensión helicoidal de cada aminoácido es variable, y depende de la naturaleza de su cadena lateral. La diferencia entre Gly y Ala se debe fundamentalmente a la mayor entropía conformacional del esqueleto peptídico del aminoácido Gly, ya que la presencia del carbono beta en Ala induce un descenso significativo en la entropía del estado desplegado (del orden de 2.4 unidades entrópicas) (D'Aquino et al., 1996; Leach et al., 1966). Para los demás aminoácidos, el carbono beta de su cadena lateral es, asimismo, el responsable de las mayores restricciones conformacionales en el esqueleto peptídico. La presencia de grupos adicionales en la cadena lateral, con la excepción de las cadenas ramificadas, no conlleva una reducción significativa de los grados de libertad del esqueleto peptídico. En estos casos (es decir, aminoácidos con cadenas laterales no ramificadas), la disminución de la entropía conformacional de las propias cadenas laterales, como consecuencia de la adopción de la conformación helicoidal por parte del esqueleto peptídico, es la causa principal de la disminución en propensión helicoidal de estos aminoácidos con respecto a Ala.

Sin embargo, la pérdida de entropía de las cadenas laterales provocada por la formación de la hélice no es por sí sola suficiente para justificar por completo las diferentes tendencias helicoidales observadas entre los aminoácidos. Así, con la excepción de Ala y Gly, la protección del contacto con el disolvente de regiones apolares en las cadenas laterales proporciona, a su vez, una contribución entrópica favorable, que compensa parcialmente las contribuciones desfavorables procedentes de la entalpía de solvatación y de la entropía conformacional. Como consecuencia, las propensiones helicoidales observadas son el resultado de una compleja combinación de contribuciones entálpicas y entrópicas, que tienen su origen, tanto en cambios en la entropía conformacional, como en la solvatación de los distintos grupos de la cadena peptídica.

## IV.B.-APLICACIÓN DE LA PARAMETRIZACIÓN ESTRUCTURAL DE LA ENERGÉTICA DE PLEGAMIENTO DE PROTEÍNAS Y UNIÓN DE LIGANDOS AL DISEÑO MOLECULAR. DISEÑO DE LIGANDOS PEPTÍDICOS DE PROTEASAS ASPÁRTICAS.

### IV.B.1.-Consideraciones generales.

La aplicación de las diversas técnicas estructurales de diseño molecular a proteínas permitiría la racionalización de los procesos de diseño de ligandos (en numerosas ocasiones de naturaleza peptídica como hormonas, inhibidores enzimáticos, sustratos o efectores alostéricos) que presenten una afinidad predeterminada por un receptor proteico específico. Sin embargo, la carencia de una función energética adecuada, con capacidad para predecir con precisión sus afinidades de unión, ha constituido uno de los principales obstáculos para el desarrollo de algoritmos de diseño efectivos basados en consideraciones estructurales.

Como se ha demostrado en numerosos trabajos anteriores, la parametrización estructural de la energética de plegamiento de proteínas y unión de sus ligandos posee la capacidad para predecir la estabilidad global de proteínas, al igual que la existencia y características principales de intermedios termodinámicos de plegamiento (Freire, 1991; Murphy, 1992; Xie, 1993; Xie, 1994 ; Xie, 1994; Freire, 1995]. Asimismo, esta metodología se ha aplicado con éxito a la evaluación de las constantes de estabilidad de los restos individuales de la proteína (reflejadas en los factores de protección al intercambio hidrógeno/deuterio (Hilser & Freire, 1996a; Hilser & Freire, 1996b; Hilser et al., 1997) detectados mediante técnicas de resonancia magnética nuclear), y al cálculo de la energética de unión de diversos ligandos a proteínas (Gómez & Freire, 1995; Gómez et al., 1997; Murphy et al., 1993). Por último, y como se ha discutido en el capítulo anterior, la utilización de dicha parametrización permite estimar las propensiones helicoidales de los distintos aminoácidos con una precisión mayor de  $0,2 \text{ kcal}\cdot\text{mol}^{-1}$  (Luque et al., 1996).

Dado que esta parametrización ha alcanzado un estado de desarrollo tal, que posibilita la predicción precisa de la energética de plegamiento de proteínas, así como de las afinidades de unión de diversos ligandos, parece apropiado intentar la aplicación de sus principios al desarrollo de estrategias de diseño molecular. En los apartados siguientes se procederá a la descripción de la metodología de diseño desarrollada en este trabajo. Aunque su finalidad inmediata es el diseño de ligandos de naturaleza peptídica con afinidades predeterminadas por sus correspondientes enzimas, esta metodología es susceptible de ser extendida, en un futuro, al diseño de ligandos de diversa naturaleza.

Cualquier algoritmo de diseño molecular debe poseer necesariamente la capacidad de identificar las conformaciones que minimizan la energía libre de Gibbs, de modo que, tras cada modificación introducida en el sistema, sea posible determinar la conformación preferentemente poblada bajo ciertas condiciones experimentales, y sobre la que se ha de proceder a la estimación



de la afinidad de unión. Como consecuencia, el desarrollo de un algoritmo de minimización efectivo, así como su validación (que requería su aplicación a sistemas modelo para los que se disponga de datos energéticos obtenidos mediante procedimientos independientes) constituye una etapa preliminar indispensable. Por este motivo, el algoritmo de minimización desarrollado en este trabajo se ha aplicado inicialmente a la optimización conformacional de las cadenas laterales de los distintos aminoácidos situados en posiciones centrales de pequeñas hélices peptídicas, sistemas para los que se dispone de los perfiles energéticos obtenidos por varios autores mediante la aplicación de funciones potenciales estándar.

Una vez comprobado que la incorporación de la parametrización estructural de la energética en el algoritmo de minimización proporciona la resolución necesaria para localizar correctamente los mínimos energéticos en el espacio conformacional, es posible proceder con el desarrollo de las estrategias de diseño de inhibidores peptídicos de diversas enzimas. Los sistemas proteicos escogidos para la aplicación de esta metodología pertenecen al grupo de las proteasas aspárticas, que constituyen una gran familia de enzimas proteolíticas ampliamente distribuida entre vertebrados, hongos, plantas y retrovirus, algunos de cuyos miembros han adquirido extraordinaria relevancia médica. En concreto, nuestra atención se ha centrado en el estudio de la interacción de inhibidores y sustratos de diversa naturaleza con Endothiapepsina (una de las proteasas aspárticas más profundamente estudiadas, por lo que constituye un buen modelo para iniciar el proceso de diseño de inhibidores), y con la proteasa del virus V.I.H.-1. Esta enzima retrovírica, que recientemente ha adquirido una gran relevancia médica debido a su papel central en el proceso de maduración del virus del S.I.D.A, constituye uno de los principales objetivos en el desarrollo terapias efectivas contra la enfermedad.

Con anterioridad, se había llevado a cabo con éxito el análisis termodinámico-estructural de la energética de unión del inhibidor Pepstatina A a Endothiapepsina (Gómez & Freire, 1995), lo que permitió identificar los restos, tanto en la proteína como en el ligando, que contribuyen más notablemente a la determinación de la afinidad de unión. Asimismo, este tipo de análisis posibilitó la disección de las contribuciones de las distintas magnitudes termodinámicas a la energía libre de Gibbs correspondiente al proceso de unión del inhibidor. En este trabajo se presenta un análisis similar llevado a cabo con la proteasa del virus V.I.H.-1, de modo que se procede al estudio termodinámico-estructural de la formación de sus complejos con Acetil-pepstatina y diversos inhibidores de naturaleza no peptídica, algunos de los cuales están siendo empleados actualmente como agentes terapéuticos para el tratamiento contra el S.I.D.A.

Una vez comprobada la capacidad de la parametrización estructural de la energética para reproducir con precisión los resultados experimentales obtenidos con estos sistemas (con lo que queda asegurado que la metodología empleada posee la sensibilidad necesaria para captar el delicado balance de fuerzas que determinan las afinidades de unión de ligandos de distinta naturaleza a estas enzimas), se procedió a la aplicación de las estrategias de diseño molecular, que

incorporan la parametrización estructural como función potencial, al diseño de ligandos peptídicos para este grupo de enzimas.

En concreto, se ha llevado a cabo el análisis termodinámico de la unión de diversos mutantes del inhibidor universal Pepstatina A a Endothiapepsina. Los nuevos ligandos se generaron, bien mediante la substitución de las cadenas laterales en distintas posiciones del pentapéptido, o bien modificando su longitud. De este modo, ha sido posible establecer una escala de afinidades de unión para los posibles mutantes en una posición determinada en la secuencia, e identificar los restos de la proteasa y del inhibidor que contribuyen más notablemente a la energética de unión.

Así mismo, se ha abordado el proceso de modelado de los complejos de la proteasa del virus V.I.H.-1 con diversos sustratos peptídicos con objeto de elucidar las bases moleculares de la aparición de la resistencia a diversos fármacos. Para ello se ha recurrido al análisis de las características termodinámicas diferenciales de la unión de estos sustratos a la enzima con respecto a las interacciones establecidas con los distintos inhibidores no peptídicos, que potencian la aparición de variantes resistentes del virus.

## **IV.B.2.-Métodos**

**IV.B.2.1.-Descripción general del enfoque termodinámico-estructural del diseño molecular.**

### **IV.B.2.1.1.-Consideraciones generales.**

Dado que la afinidad de unión entre una proteína y un ligando cualesquiera está determinada por el cambio en la energía libre de Gibbs correspondiente al proceso de asociación entre ambos, el problema central que debe abordar cualquier metodología de diseño molecular basada en consideraciones estructurales es, por tanto, la predicción de su magnitud a partir de la estructuras tridimensionales, tanto del complejo como de la proteína y ligando libres. Sin embargo, hasta el momento, la termodinámica tradicional no ha desempeñado un papel relevante en el desarrollo de las técnicas de diseño molecular debido a la inexistencia de una base teórica que relacione, de forma precisa, las características estructurales del sistema con su energía libre de Gibbs. El desarrollo de ecuaciones paramétricas, como las que constituyen la base de la metodología empleada en este trabajo, que permitan la predicción precisa de las afinidades de unión a partir de consideraciones estructurales, sin necesidad de llegar al esclarecimiento absoluto del origen y naturaleza de cada una de las contribuciones, podría consituir una prometedora alternativa.

Cualquier proceso de diseño molecular implica la substitución o adición de grupos de diversa naturaleza química a la molécula del ligando con objeto de conseguir la afinidad de unión deseada. El efecto de estas substituciones se manifiesta de modo invariable en la magnitud de la

entalpía y entropía de unión. Sin embargo, como consecuencia del delicado balance establecido entre estos términos mutuamente compensatorios, los cambios que las modificaciones introducidas en el sistema originan en estos parámetros termodinámicos son generalmente de mayor magnitud que los efectos resultantes en la energía libre de Gibbs. Estos fenómenos de compensación entalpía/entropía constituyen el principal obstáculo en todo proceso de diseño debido a que cualquier modificación molecular que se traduzca en contribuciones entálpicas más favorables a la unión conlleva contribuciones entrópicas desfavorables. Asimismo, los cambios en entalpía y en entropía se encuentran, a su vez, compensados por varios términos de naturaleza opuesta, de tal modo que la entalpía correspondiente a la formación de interacciones internas es de signo contrario a la entalpía originada a consecuencia de los procesos de desolvatación, y la entropía de desolvatación se ve compensada por la entropía conformacional. Por tanto, el desarrollo de estrategias de diseño molecular efectivas requiere indiscutiblemente la comprensión de dichos fenómenos, con objeto de minimizar las consecuencias de sus efectos compensatorios.

De este modo, aunque la energía libre de Gibbs es claramente la magnitud de mayor interés, la determinación experimental, así como la predicción estructural, de las demás magnitudes termodinámicas permite el acceso a información de gran importancia para el diseño molecular. El estudio termodinámico de las interacciones entre una proteína y sus ligandos proporciona, no sólo el valor de la energía de Gibbs para el proceso de unión ( $\Delta G$ ), sino los cambios en entalpía ( $\Delta H$ ), entropía ( $\Delta S$ ) y capacidad calorífica ( $\Delta S$ ). Además de ser responsables de la dependencia con la temperatura de las constantes de afinidad, dichos cambios en entalpía, entropía y capacidad calorífica reflejan el balance existente entre las distintas fuerzas moleculares, que contribuyen a la determinación de la energía libre de Gibbs correspondiente al proceso de unión. De este modo, la comparación de los valores predichos por el análisis termodinámico-estructural con los obtenidos experimentalmente para las cuatro magnitudes termodinámicas ( $\Delta G$ ,  $\Delta H$ ,  $\Delta S$  y  $\Delta C_p$ ) permite determinar si los cálculos realizados captan correctamente la naturaleza y magnitud de las fuerzas implicadas en el proceso de asociación entre moléculas.

#### *IV.B.2.1.2.-Estrategia general para el diseño termodinámico-estructural de péptidos.*

Dada la naturaleza de la parametrización estructural de la energética, consistente en un conjunto de ecuaciones paramétricas, obtenidas a partir del análisis estadístico de las magnitudes termodinámicas y de las estructuras cristalográficas de alta resolución disponibles para proteínas globulares, la opción más natural para sus primeras aplicaciones al diseño molecular es, obviamente, el diseño de inhibidores y ligandos de naturaleza peptídica. Por lo tanto, el objetivo de la metodología expuesta a continuación consiste fundamentalmente en el desarrollo de un ligando peptídico que se una a la proteína bajo estudio con una afinidad determinada. Aunque posible, la aplicación de estos procedimientos al diseño de ligandos no peptídicos requeriría,

obviamente, el desarrollo de una parametrización adicional, específica para el tipo de compuestos que se deseen analizar.

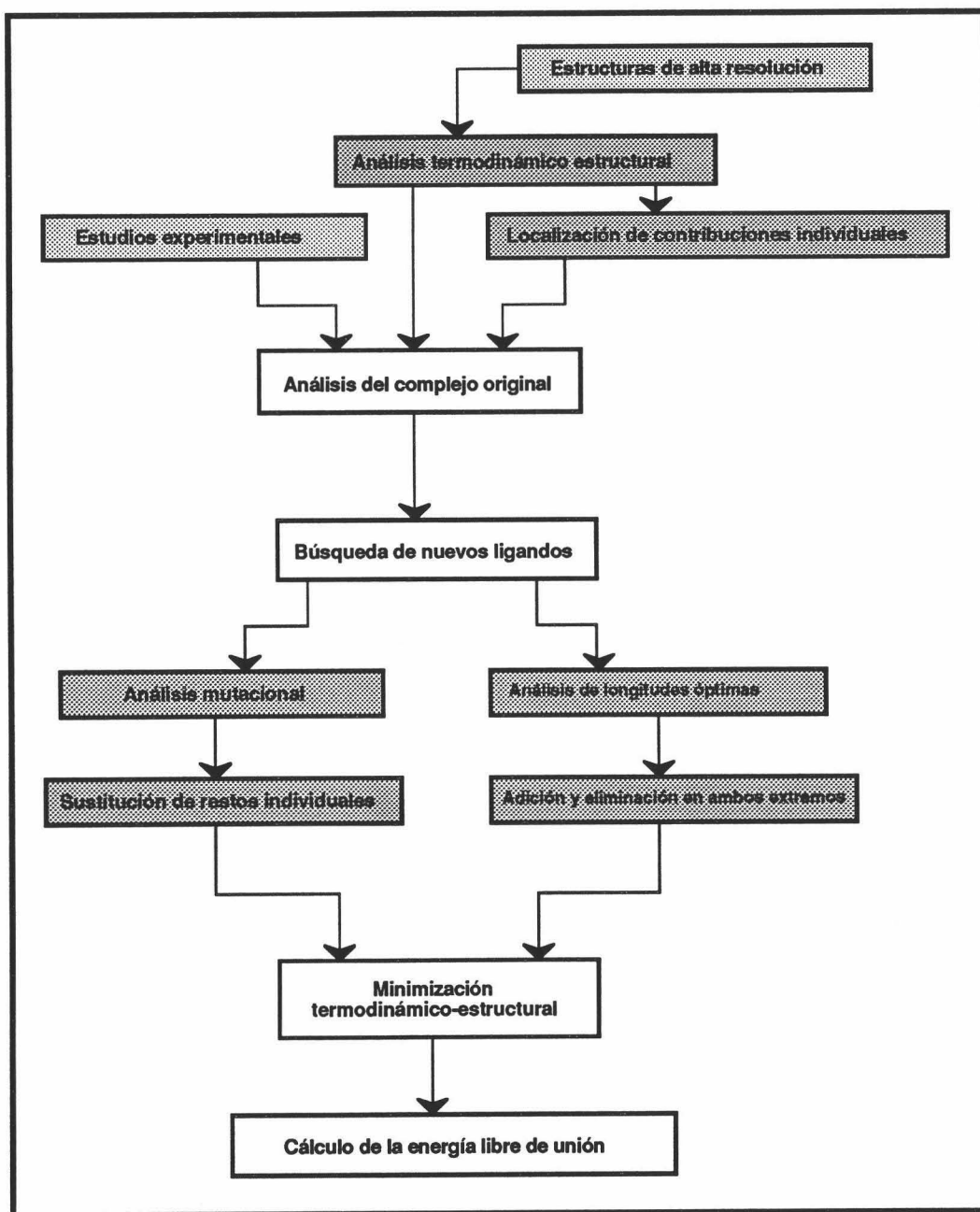
El procedimiento de diseño más sencillo implica el conocimiento de la estructura tridimensional de alta resolución de un complejo entre la proteína de interés y un ligando peptídico, de modo que, tomando dicha estructura como punto de partida, y mediante la modificación sistemática del péptido, se obtenga como resultado un nuevo ligando, caracterizado por una afinidad predeterminada por la enzima. Para llevar a cabo el cálculo de afinidades de unión de forma precisa se requiere, en principio, el conocimiento, tanto de la estructura de la proteína asociada con el ligando, como de su forma libre. En aquellos casos en que únicamente se dispone de la estructura correspondiente al complejo, es todavía posible evaluar las energías libres de Gibbs relativas ( $\Delta\Delta G$ ) para el proceso de unión de los péptidos modificados con respecto al complejo original con la misma precisión, siendo aún posible el desarrollo del proceso de diseño molecular. Esta situación se mantiene incluso en el caso de existencia de cambios conformacionales significativos entre la proteína libre y su forma complejada. En la figura IV.B.1 se muestra un esquema representativo de las principales etapas del proceso de diseño molecular empleado en este trabajo, y que se desarrollarán a continuación.

#### *a.-Análisis termodinámico-estructural del complejo original.*

Con anterioridad a la modificación de la estructura del ligando peptídico es necesario llevar a cabo el análisis termodinámico-estructural de la energética de unión del péptido original, por lo que el proceso de diseño termodinámico implica varias etapas preliminares. En primer lugar, es conveniente aplicar la parametrización estructural de la energética al cálculo de las magnitudes termodinámicas ( $\Delta G$ ,  $\Delta H$ ,  $\Delta S$  y  $\Delta C_p$ ) que cabría esperar para la formación del complejo entre el ligando original y la proteína. Es necesario, a continuación, estimar la precisión con que se han realizado las predicciones para el sistema bajo estudio mediante la comparación de los resultados de los cálculos con los obtenidos experimentalmente para las distintas magnitudes termodinámicas. Por último, debe llevarse a cabo un análisis de las contribuciones individuales de cada resto constituyente del ligando peptídico a la energética de unión. Este análisis preliminar proporcionará, por tanto, una visión completa de las interacciones establecidas entre el péptido y la proteína, así como del origen de las fuerzas que contribuyen más notablemente a la determinación de la energética de unión, y de las contribuciones relativas de cada aminoácido a la afinidad.

#### *b.-Modificación del péptido original.*

Una vez realizada la caracterización termodinámica del complejo original, se dispone de dos procedimientos complementarios para el diseño de la nueva versión del péptido, consistentes,



*Figura IV.B.1: Esquema general de las principales etapas del proceso de modelado molecular utilizado en este trabajo.*

bien en la sustitución de cadenas laterales en posiciones ya existentes en la cadena, o bien en la alteración de su longitud, mediante adiciones o eliminaciones de restos aminoacídicos. Dado que, en cualquiera de los casos, se adopta como punto de partida la estructura tridimensional de un complejo ya existente entre el péptido y la proteína, este procedimiento conlleva la suposición implícita de que la estructura global de la proteína en el complejo con los nuevos ligandos



permanece esencialmente invariable, de modo que los efectos de las modificaciones llevadas a cabo en el ligando se encuentran restringidos a sus alrededores más cercanos.

#### ***b.1.-Modificación de la secuencia: Análisis mutacional.***

La primera alternativa para conseguir modular la afinidad de unión del péptido a la proteína consistiría en la modificación de su secuencia aminoacídica. En el supuesto de que se pretenda conseguir la mayor afinidad posible sería necesario, por tanto, encontrar aquella secuencia que minimice el cambio en energía libre del proceso de unión. Este tipo de análisis implica la sustitución de los aminoácidos individuales en el péptido y el cálculo posterior de la energía libre de unión esperada para cada una de las mutaciones. Aunque, en principio, para péptidos de pequeña longitud sería posible llevar a cabo una búsqueda exhaustiva de la secuencia óptima, los requerimientos computacionales aumentan con extrema rapidez conforme se incrementa el número de restos que constituyen el ligando, de modo que este tipo de procedimientos resultan prohibitivos. La aplicación de un algoritmo de búsqueda, basado en un análisis termodinámico-estructural previo del complejo original, constituye una buena alternativa, dado que, en este caso, el proceso puede restringirse a un conjunto determinado de restos aminoacídicos. Por ejemplo, en el supuesto de que se pretenda maximizar la afinidad de unión, puede resultar conveniente centrar el análisis en aquellos restos que contribuyen en menor medida a la energía libre correspondiente al proceso de asociación.

#### ***b.2.-Modificación de la longitud.***

Una vez identificada la secuencia óptima, la siguiente etapa lógica para la modulación de la afinidad de unión del péptido implica la adición de aminoácidos en sus extremos amino o carboxilo terminal. Asimismo, puede considerarse como una posible alternativa la reducción de su secuencia mediante la eliminación de alguno de los restos, aunque esta operación tiene frecuentemente como resultado una disminución en la afinidad. De igual modo que en el caso anterior, este análisis implica el cálculo de la afinidad de unión para cada modificación (adición o eliminación) realizada sobre el péptido original.

### ***IV.B.2.1.3.-Minimización de conformaciones.***

#### ***a.-Descripción general.***

Como se ha indicado anteriormente, una de las tareas de mayor importancia en el proceso de diseño molecular consiste en el cálculo de las energías libres de unión correspondientes a las nuevas estructuras, obtenidas tras la modificación del ligando original, para lo que se requiere la identificación de las conformaciones adoptadas preferentemente por el nuevo ligando. Por tanto,



para los dos tipos de mutaciones descritas en el apartado anterior, la estrategia a seguir implica la minimización energética con carácter local de la estructura, asumiendo que la conformación global de la proteína unida al péptido modificado es similar a la presentada en el complejo original, de modo que los cambios provocados por las mutaciones se encuentran limitados a las regiones vecinas al sitio de unión.

En la figura IV.B.2 se muestra un esquema del procedimiento seguido para llevar a cabo la identificación de las conformaciones correspondientes a los mínimos energéticos. Las distintas mutaciones se llevan a cabo mediante la substitución de la cadena lateral original por la correspondiente al aminoácido deseado, manteniendo los átomos del esqueleto peptídico en su conformación original. A continuación se procede a la generación de diferentes conformaciones de la cadena lateral mediante la rotación sistemática de sus ángulos diedros, cubriendo un intervalo completo de  $360^\circ$ . En el caso de mutaciones que impliquen la adición de restos en los extremos del péptido el procedimiento es idéntico, salvo por la necesidad de incluir los ángulos correspondientes al esqueleto peptídico del aminoácido añadido al proceso de minimización energética.

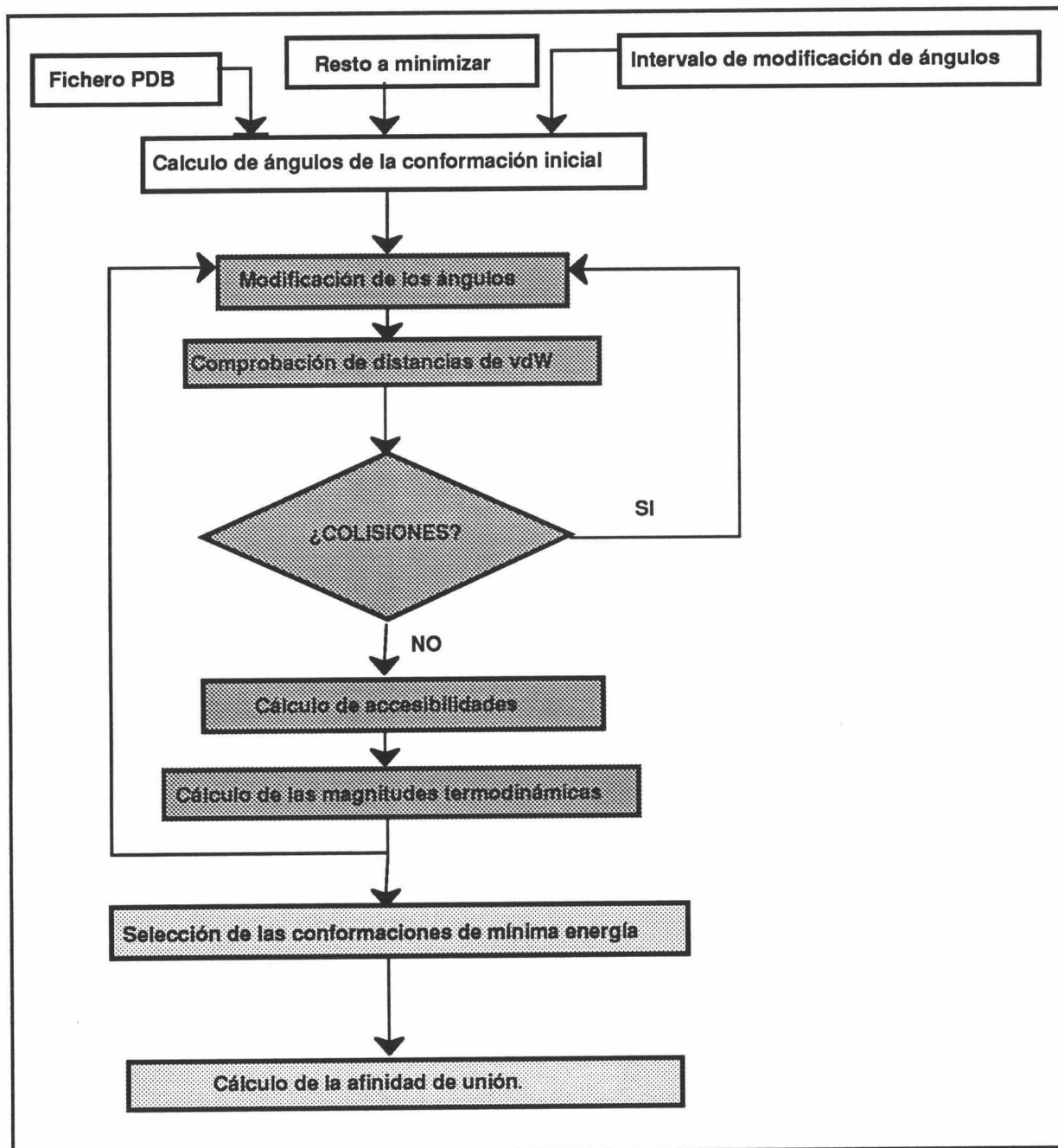
Para cada posible conformación, caracterizada por ciertos valores de los ángulos diedros, es necesario realizar las siguientes operaciones: en primer lugar comprobar que no existen colisiones entre los distintos átomos del complejo; a continuación, para aquellas conformaciones permitidas estéricamente, es necesario evaluar su energía libre con respecto a la conformación original mediante la aplicación de la parametrización estructural de la energética descrita en el apartado III.2. y, por último, seleccionar las conformaciones correspondientes a los mínimos energéticos. El proceso completo deberá llevarse a cabo para cada una de las mutaciones, con objeto de posibilitar la identificación de la secuencia peptídica que presenta la mayor afinidad de unión.

#### ***b.-Selección de conformaciones permitidas.***

Por razones estéricas, no todas las conformaciones generadas mediante la modificación sistemática de los ángulos diedros son posibles, lo que hace necesario analizar la posible presencia de solapamientos entre las esferas de van der Waals de los átomos del sistema. De este modo, toda conformación que posea al menos una pareja de átomos situados a una distancia menor que la suma de sus correspondientes radios de van der Waals es rechazada, procediéndose a una nueva modificación de los ángulos. Aquellas conformaciones que no presentan problemas estéricos son sometidas al análisis termodinámico-estructural, con objeto de estimar sus energías y probabilidades.

Como se ha comentado con anterioridad (ver sección IV.A.2.2.2), la amplitud y localización de las regiones permitidas estéricamente en el espacio conformacional es sensible a la elección de los radios de van der Waals asignados a cada tipo de átomos. En esta ocasión, el juego

de radios empleados es el denominado MMII, que ha sido obtenido por Iijima y colaboradores (Iijima et al., 1987). Estos radios, recogidos en la tabla IV.B.1, representan los valores que mejor



*Figura IV.B.2: Esquema general de las principales etapas del proceso de minimización utilizado en este trabajo.*

reproducen las estructuras tridimensionales de la base de datos estructurales de proteínas (Iijima et al., 1987).

**Tabla IV.B.1:** Radios de van der Waals utilizados para identificar las conformaciones con problemas estéricos.

(Å)					
C carbonílico	1.35	C alifático	1.35	O	1.20
N	1.27	S	1.02		

### c.-Función potencial.

Ha quedado claro que, para cada nuevo resto o cadena lateral introducidos en el péptido, es necesario llevar a cabo el análisis del conjunto de posibles conformaciones que éstos puedan adoptar y la evaluación de la energía y probabilidad de cada una de ellas. La probabilidad de una conformación peptídica individual, definida por un juego específico de coordenadas correspondientes al esqueleto peptídico y a las cadenas laterales, se encuentra determinada por su energía de Gibbs.

La función potencial empleada en nuestro caso,  $\Delta G_{ef}$ , está especificada por la entalpía procedente de las interacciones intra e intermoleculares, establecidas entre el péptido y la proteína, y por los correspondientes cambios en entalpías y entropías de solvatación, y es una función de los ángulos diedros de las cadenas laterales y del esqueleto peptídico. El potencial  $\Delta G_{ef}$  representa la energía libre de Gibbs o la función potencial de Gibbs de una única conformación, que, por definición, no incluye las contribuciones correspondientes a la entropía conformacional del péptido. Es, por tanto, importante evitar la confusión entre esta función potencial y la energía libre de Gibbs asociada al proceso de unión, que incluye las contribuciones de todas las posibles conformaciones del sistema. La figura IV.B.3 ilustra la situación para dos conformaciones hipotéticas de una cadena lateral cualquiera. Como se puede apreciar, dichas conformaciones no sólo presentan distintas interacciones intramoleculares, sino que también difieren en sus grados de hidratación, diferencias ambas que determinan el valor de la función energética  $\Delta G_{ef}$ .

La probabilidad de una conformación cualquiera se define de acuerdo con la ecuación:

$$P_i = \frac{e^{-\Delta G_{ef,i}/RT}}{\sum_j e^{-\Delta G_{ef,j}/RT}} \quad (IV.B.1)$$

donde el numerador es el exponente de Boltzman para la conformación  $i$ , y la sumatoria en el denominador representa la función de partición conformacional, definida como la suma de los pesos estadísticos de todas las posibles conformaciones.

La función potencial de Gibbs,  $\Delta G_{ef}$ , es la herramienta empleada para identificar la conformación más probable de una cadena lateral o un aminoácido completo, de modo que, para

cada conformación generada, el correspondiente valor de  $\Delta G_{ef}$  se calcula a partir de sus coordenadas tridimensionales, haciendo uso de la parametrización estructural de la energética descrita anteriormente (sección III.2.) y excluyendo del cálculo los términos correspondientes a la entropía conformacional. Para comodidad del lector, las principales ecuaciones y parámetros del formalismo se recogen en el apéndice A.2

#### **IV.B.2.2.-Minimización de conformaciones helicoidales.**

##### **IV.B.2.2.1.-Construcción de las hélices.**

Como se ha comentado en la sección introductoria, con anterioridad a la aplicación de la metodología descrita al diseño de ligandos peptídicos en sistemas reales, es imprescindible comprobar la habilidad de la función potencial empleada, es decir, de la parametrización estructural de la energética de plegamiento de proteínas y unión de ligandos, para localizar correctamente los mínimos energéticos. Para ello, se procedió a la aplicación del algoritmo de minimización descrito en el apartado anterior a la optimización conformacional de las cadenas laterales en posición central de hélices alfa expuestas al disolvente, sistemas para los que se dispone de los perfiles energéticos obtenidos con anterioridad por diversos autores (Janin et al., 1978; Lee et al., 1994) mediante la aplicación de funciones potenciales estándar.

Los cálculos se llevaron a cabo sobre hélices modelo, de nueve restos de longitud y secuencia  $(Ala)_4$ -X- $(Ala)_4$ , construidas empleando el programa InsightII (Biosym), de modo que los ángulos del esqueleto peptídico adopten los valores característicos de una hélice alfa ( $\phi = -40^\circ$  y  $\psi = -60^\circ$ ).

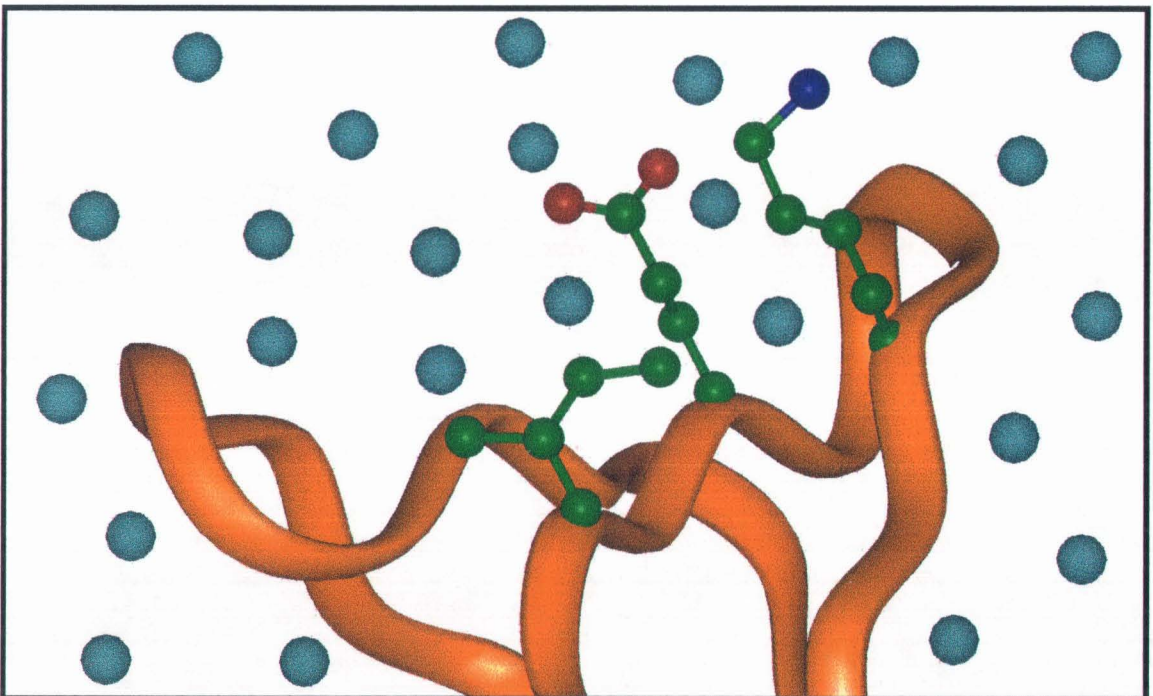
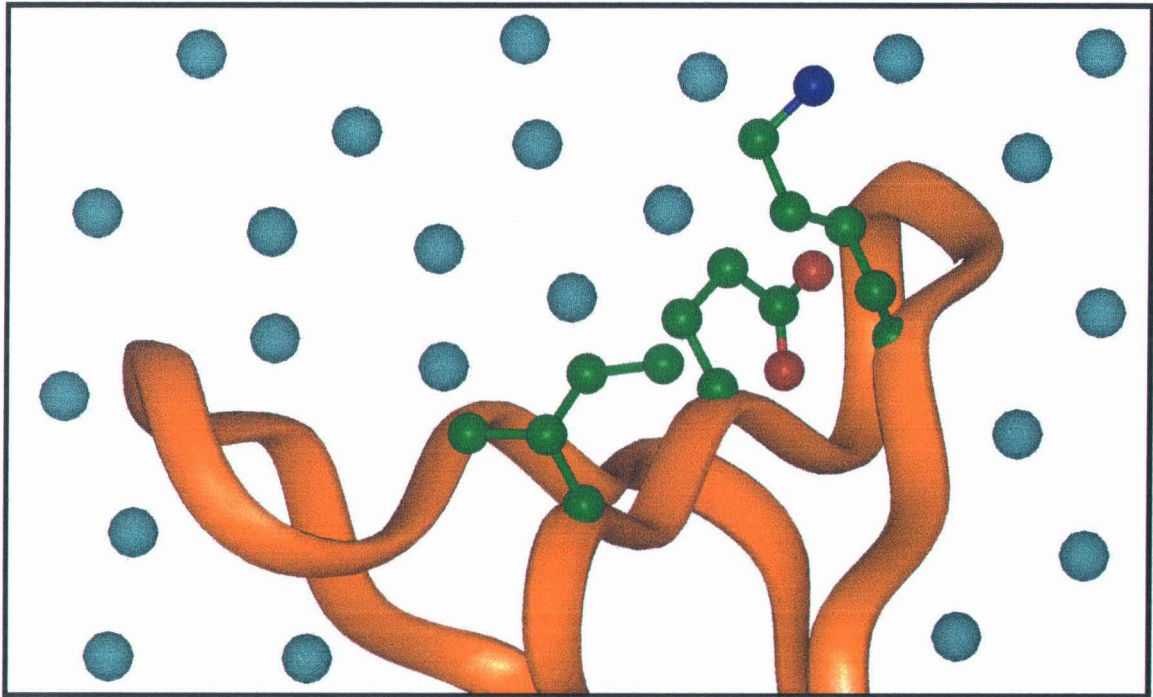
##### **IV.B.2.2.2.-Generación de conformaciones.**

Las distintas conformaciones de las cadenas laterales se generaron mediante la modificación sistemática de todos sus ángulos diedros ( $\chi_1$  para Cys, Ser, Thr y Val;  $\chi_1$  y  $\chi_2$  para Asn, Asp, His, Ile, Leu, Phe, Trp y Tyr;  $\chi_1$ ,  $\chi_2$  y  $\chi_3$  para Gln, Glu y Met;  $\chi_1$ ,  $\chi_2$ ,  $\chi_3$  y  $\chi_4$  para Lys; y  $\chi_1$ ,  $\chi_2$ ,  $\chi_3$ ,  $\chi_4$ , y  $\chi_5$  para Arg) en un intervalo de  $360^\circ$ . Para las cadenas laterales con un único ángulo diedro el valor de  $\chi_1$  se ha modificado cada grado, para cadenas con dos o tres diedros cada  $10^\circ$  y por último para cadenas con un número mayor de ángulos de torsión cada  $30^\circ$ .

#### **IV.B.2.3.-Estudio termodinámico-estructural de la energética de unión de inhibidores a la proteasa del virus V.I.H.-1.**

##### **IV.B.2.3.1.-Cálculo de la Energía libre de unión.**

*Figura IV.B.3: Ilustración de cómo dos conformaciones diferentes de la misma cadena lateral no sólo difieren entre sí en las interacciones internas en las que se encuentran implicadas , sino en su grado de hidratación, de modo que ambas contribuciones determinan su energía libre de Gibbs.*





*Figura 4.A.3.I: Ilustración del procedimiento seguido para el cálculo de los valores promedio de las áreas polares y apolares promedio en conformación helicoidal para el caso concreto del aminoácido Fenilalanina. En el panel superior izquierdo se muestra el perfil de energía en función de la conformación de la cadena lateral como curvas de energía espaciadas a intervalos de  $2 \text{ kcal}\cdot\text{mol}^{-1}$ . En el panel superior derecho se muestra el correspondiente perfil de probabilidad, donde las líneas de contorno corresponden a los valores  $10^{-5}$ ,  $10^{-4}$ ,  $10^{-3}$ ,  $3 \cdot 10^{-3}$ ,  $5 \cdot 10^{-3}$ ,  $7 \cdot 10^{-3}$ ,  $10^{-2}$ ,  $2 \cdot 10^{-2}$ ,  $3 \cdot 10^{-2}$ ,  $4 \cdot 10^{-2}$  y  $5 \cdot 10^{-2}$ . Los paneles inferiores representan los perfiles de accesibilidad polar y apolar, donde las regiones moradas corresponden a las regiones no permitidas estéricamente.*

Los cambios en energía libre asociados a la unión de los inhibidores se han calculado mediante la aplicación de la parametrización estructural de la energética de plegamiento de proteínas y unión de ligandos (D'Aquino et al., 1996; Gómez & Freire, 1995; Gómez et al., 1995; Hilser et al., 1996; Luque et al., 1996), desarrollada en el laboratorio del Dr. Freire y descrita en el apartado II.2. Las ecuaciones y parámetros específicos empleados en los cálculos se encuentran resumidos en el apéndice A.2.

El cálculo de las energías libres de unión mediante la aplicación de la parametrización estructural requiere el conocimiento de las estructuras de los complejos, así como de la proteína y ligando libres. En todos los casos, los cálculos se llevaron a cabo a partir de las estructuras cristalográficas publicadas con anterioridad y depositadas en el Banco de Datos de Proteínas. Las coordenadas del complejo de la proteasa del virus V.I.H.-1 con el inhibidor acetil-pepstatina se encuentran recogidas en el fichero PDB de código 5HVP, mientras que los códigos correspondientes a los complejos con los distintos inhibidores no peptídicos, considerados en este trabajo son: 1HVI para A-77003, 1HVJ para A-78791, 1HVK para A-76928, 1HVL para A-76889, 9HVP para A-74704, 1HVP para VX-478, 1SBG para SB-203386, 1HBV para SB-203238, 1HPS para SB-206343, 2UPJ para U-110313, 1GNO para U-89360, 1PRO para A-98881, y 1HIH para CGP-53820.

*a.-Situaciones especiales que deben ser consideradas.*

*a.1.-Presencia de aminoácidos y grupos poco comunes.*

En ocasiones, los ligandos de naturaleza peptídica contienen grupos funcionales específicos o aminoácidos poco comunes que deben considerarse de forma especial. En aquellos casos en los que el inhibidor contiene algún aminoácido poco frecuente, para el cual no se dispone de datos de entropía conformacional específicos, se procede a su asimilación a alguno de los veinte aminoácidos más comunes. Este es el caso de Norleucina, presente en algunos de los inhibidores peptídicos, y que posee el mismo número de enlaces y naturaleza muy parecida al aminoácido Metionina, o los grupos ABA (ácido amino butírico) que se emplean en ocasiones como sustitutos de Cys en la proteasa del virus V.I.H.-1. En el caso de los restos de estatina, la nomenclatura empleada se recoge en la figura II.B.3. Como se puede observar, el grupo hidroxilo de este resto se nombra como los correspondientes a Ser y Thr, debido a que, por su naturaleza, y al igual que sucede con estos aminoácidos, se espera que su contribución a la capacidad calorífica sea positiva (ver sección III.2.3.3.).

*a.2.-Estructuras incompletas o con conformaciones múltiples.*

Como se ha indicado anteriormente, los cálculos de afinidades de unión se realizan preferentemente, y siempre que sea posible, considerando las estructuras correspondientes al

complejo y a la proteína libre en disolución. Dado que se trata de dos estructuras elucidadas independientemente, es posible que existan discrepancias entre las conformaciones adoptadas por algunas de las cadenas laterales en la superficie molecular de ambas estructuras. Es importante, para la determinación precisa de las afinidades de unión, minimizar las variaciones de accesibilidad entre las dos estructuras que no se encuentren directamente relacionadas con la formación del complejo.

Entre las situaciones más frecuentes cabe destacar la presencia de dos conformaciones alternativas para alguna cadena lateral en la superficie molecular, en cuyo caso se procede al cálculo de los parámetros termodinámicos en ambos casos y al promediado de los resultados obtenidos. En aquellos casos en que las cadenas laterales no estén bien definidas y no se disponga de las coordenadas tridimensionales de algunos de sus átomos, y siempre que se trate de restos completamente expuestos y alejados del sitio de unión, se consideran como restos de Ala en ambas estructuras.

En otras ocasiones, como es el caso de la unión de inhibidores a la proteasa del virus V.I.H.-1, y debido a su naturaleza simétrica, los ligandos se encuentran en el complejo con dos orientaciones diferentes, en cuyo caso se procede, asimismo, a la evaluación de los parámetros termodinámicos correspondientes al proceso de unión para ambas estructuras.

#### *b.-Cálculo de accesibilidades.*

Los cambios en las superficies accesibles al disolvente asociadas a la unión de los inhibidores y sustratos se han calculado de acuerdo con el algoritmo de Lee y Richards (Lee & Richards, 1971), utilizando en todos los casos un radio de 1.4 Å para la molécula de disolvente y un espesor de corte transversal de 0.25 Å. Debido a la naturaleza del algoritmo empleado (descrito en el apartado III.2.2.3.1, los valores concretos para las superficies accesibles polares y apolares pueden verse modulados ligeramente con la orientación de la molécula con respecto al plano de corte. Como consecuencia, para evitar este tipo de interferencias, y con objeto de conseguir una mejor definición de las pequeñas diferencias en accesibilidad entre los distintos inhibidores y sus mutantes, se ha considerado la posibilidad de promediar los valores de accesibilidad entre múltiples orientaciones de las moléculas. En concreto, los cálculos presentados en este trabajo se han realizado considerando 64 orientaciones diferentes con respecto al plano de corte, generadas mediante la rotación de la molécula alrededor de los ejes X, Y o Z cada 90°. De este modo, para cada átomo, el valor de accesibilidad utilizado en todos los cálculos posteriores es el resultante del promedio numérico obtenido considerando todas las orientaciones moleculares.

En aquellos casos en que se requieren los valores de accesibilidad para las conformaciones desplegadas, se emplea el juego de valores optimizados para cada aminoácido (cuya obtención ha sido uno de los objetivos del trabajo y que se describe en la sección IV.A.2.2.2), de modo que el área accesible para el estado desplegado de cualquier cadena

polipeptídica se obtiene mediante el producto de la accesibilidad estimada para cada tipo de aminoácido por el número de veces que dicho resto aparece en la secuencia. El cálculo de las superficies accesibles correspondientes a los inhibidores no peptídicos libres en disolución se ha llevado a cabo sobre sus coordenadas, tal como aparecen en el fichero PDB del correspondiente complejo.

*c.-Cálculo de entropía conformacional para los inhibidores no peptídicos.*

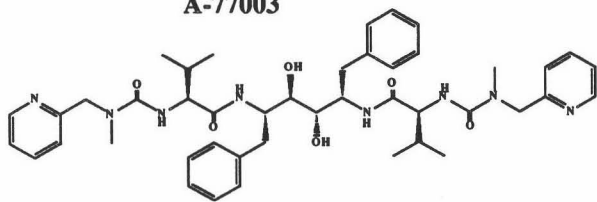
A pesar del carácter no peptídico de ciertos inhibidores de la proteasa del virus V.I.H.-1, es de esperar que las interacciones establecidas con la proteína sean de naturaleza similar a las que caracterizan, tanto el proceso de plegamiento de una proteína, como su interacción con ligandos de naturaleza peptídica. Por lo tanto, es razonable esperar que las ecuaciones que describen los cambios de las distintas magnitudes termodinámicas correspondientes a la unión de dichos inhibidores a la enzima en función de las áreas accesibles al disolvente sean aplicables a estos compuestos. Sin embargo, el cálculo de la entropía conformacional, determinada por el número de enlaces con libertad de giro, así como por las restricciones conformacionales impuestas por el resto de la molécula a cada una de las posibles rotaciones, depende de la naturaleza del compuesto bajo estudio.

Los cambios en entropía conformacional como consecuencia de la unión de ligandos o del plegamiento se evalúan generalmente mediante la consideración de tres términos correspondientes a la exposición y desplegamiento de las cadenas laterales y al desplegamiento del esqueleto. La magnitud de cada uno de ellos se ha estimado para los diferentes aminoácidos mediante el análisis teórico de la probabilidad de los diferentes conformeros en función de los ángulos diedros (D'Aquino et al., 1996; Lee et al., 1994). En el caso de los inhibidores considerados en este estudio, de naturaleza no peptídica, se ha desarrollado una parametrización especial con objeto de poder evaluar, en un primera aproximación, el cambio en los grados de libertad conformacionales de estos compuestos al pasar de su forma libre al complejo.

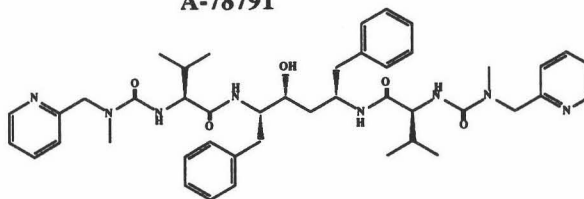
En la figura IV.B.4, donde se muestran las estructuras de los inhibidores considerados, se pueden observar las diferencias existentes entre ellos en cuanto a número de átomos y enlaces con libertad de giro. En general, la entropía conformacional de cada inhibidor en disolución será proporcional al número de enlaces con posibilidad de giro que éste posea, aunque, para un número determinado de estos enlaces, las restricciones conformacionales aumentarán con el número total de átomos de la molécula. Al igual que ocurre con los ligandos peptídicos, las posibilidades de rotación de estos compuestos cuando se encuentran formando parte del complejo son muy restringidas, considerándose en este caso con entropía conformacional nula.

Como primera aproximación, el cambio en entropía conformacional de los inhibidores no peptídicos como consecuencia de su unión a la enzima,  $\Delta S_{\text{no pept}}$ , se ha descrito como una función son muy restringidas, considerándose en este caso con entropía conformacional nula.

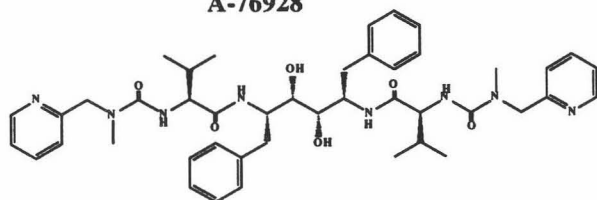
A-77003



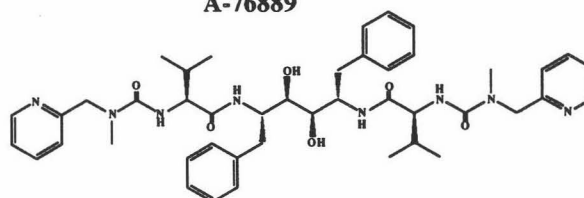
A-78791



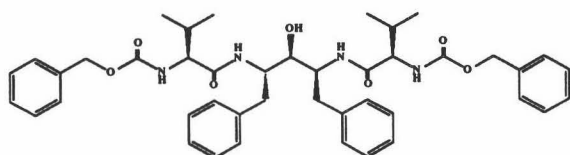
A-76928



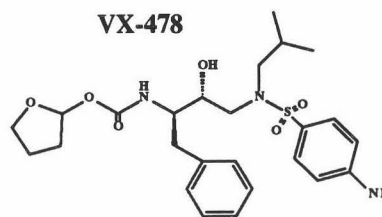
A-76889



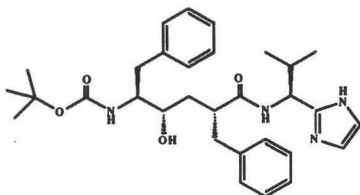
A-74704



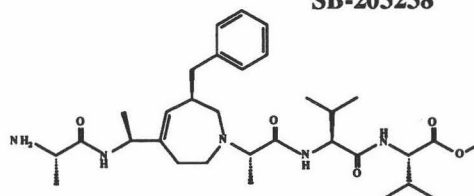
VX-478



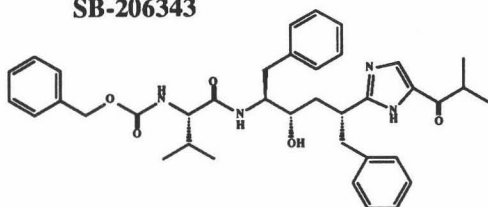
SB-203386



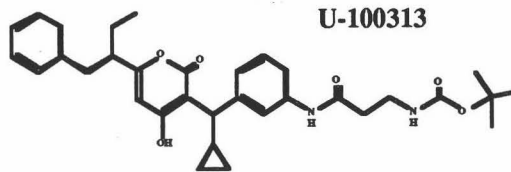
SB-203238



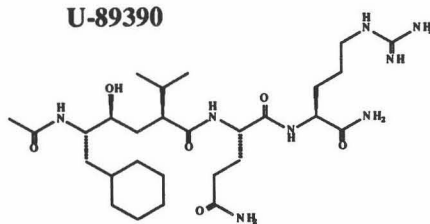
SB-206343



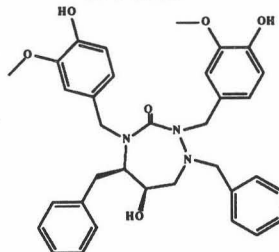
U-100313



U-89390



A-98881



CGP-53820

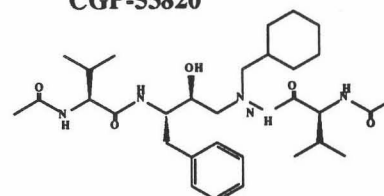


Figura IV.B.4: Estructura química de los inhibidores no peptídicos considerados en este trabajo.

Como primera aproximación, el cambio en entropía conformacional de los inhibidores no peptídicos como consecuencia de su unión a la enzima,  $\Delta S_{\text{no pept}}$ , se ha descrito como una función lineal del número de enlaces con capacidad de rotación ( $N_{\text{rot}}$ ) y del número total de átomos ( $N_{\text{átomos}}$ ) que los constituyen:

$$\Delta S_{\text{no pept}} = k_1 \cdot N_{\text{rot}} + k_2 \cdot N_{\text{átomos}} \quad (\text{IV.B.2})$$

El valor de los coeficientes  $k_1$  y  $k_2$  se obtuvo mediante el ajuste de los cálculos realizado con un juego más amplio de inhibidores no peptídicos que incluye, además de los empleados en este estudio, algunos inhibidores de renina (SC-46507, SC-46944, SC-47008, SC-47764, SC-47563, SC-47921), análogos de angiotensinógeno (Gómez et al., 1997). El valor del parámetro  $k_1$  obtenido del ajuste ( $-1.76 \text{ kcal} \cdot (\text{K} \cdot \text{mol})^{-1}$ ) es muy similar al estimado para la rotación de los enlaces carbono-carbono en parafinas de larga longitud (Schellman, 1955a; Schellman, 1955b), mientras que el coeficiente  $k_2$  ( $0.414 \text{ kcal} \cdot (\text{K} \cdot \text{mol})^{-1}$ ) refleja esencialmente las restricciones a la entropía conformacional del compuesto, cuando éste se encuentra en disolución, impuestas por los impedimentos estéricos procedentes del resto de la molécula.

#### d.-Contribuciones de los efectos de ionización.

Frecuentemente, el equilibrio de unión de los ligandos a una proteína se encuentra acoplado a la protonación o ionización obligatoria de uno o más grupos, tanto en la proteína como en el ligando. De ser éste el caso, la contribución adicional al cambio en energía libre con respecto a las condiciones en que el grupo se encuentra protonado puede calcularse en términos de los valores de  $\text{pK}_a$  en el estado inicial y final del proceso, de acuerdo con la ecuación:

$$\Delta G_{\text{ion}} = -R \cdot T \cdot \ln \left( \frac{10^{(\text{pK}_a - \text{pH})}}{1 + 10^{(\text{pK}_a - \text{pH})}} \right) \quad (\text{IV.B.3})$$

Del mismo modo, la constante de afinidad global puede calcularse como

$$K_b(\text{pH}) = K_{b, \text{int}} \cdot \frac{10^{(\text{pK}_a - \text{pH})}}{1 + 10^{(\text{pK}_a - \text{pH})}} \quad (\text{IV.B.4})$$

siendo  $K_{b, \text{int}}$  la constante de unión bajo condiciones en las que el grupo ionizable se encuentra protonado.

En nuestro caso, diversas evidencias experimentales indican que la unión de inhibidores y sustratos a la proteasa se encuentra acoplada a la protonación forzosa de uno de los grupos Asp 25 de la enzima, de forma que ambos presentan diferentes estados de ionización en el complejo.



Los  $pK_a$  de dichos restos aspárticos difieren significativamente en el estado complejoado ( $pK_a < 2.5$  y  $pK_a > 6.5$ ) (Smith et al., 1996; Wang et al., 1996). El ensayo de actividad estándar para la proteasa se realiza a un pH comprendido entre 4.7 y 5.5 (Abdel-Meguid et al., 1994; Erickson et al., 1990; Fassler et al., 1993; Hoog et al., 1995; Kim et al., 1995; Lin et al., 1995; Madhusoodan et al., 1994; Thaisrivongs et al., 1995), por lo que los datos experimentales considerados en este trabajo se han obtenido bajo condiciones en las que, al encontrarse uno de los grupos aspárticos protonado previamente a la unión de los substratos e inhibidores, la dependencia de la actividad con el pH es mínima y la contribución a la constante de afinidad del acoplamiento entre la unión de los ligandos y la protonación es pequeña. Para llevar a cabo los cálculos se ha considerado un  $pK_a$  igual a 3.5 para el grupo protonado como consecuencia de la unión (Hyland et al., 1991a; Hyland et al., 1991b).

#### *e.-Contribución de la entropía translacional.*

En el caso de la aplicación de la parametrización estructural de la energética al estudio de procesos de unión de ligandos o asociación entre macromoléculas es necesario incluir en el cálculo la contribución procedente de la reducción en los grados de libertad traslacionales de las moléculas implicadas como consecuencia de la unión. Debido a que, hasta el momento, no se dispone de cálculos precisos (Janin, 1995; Kauzmann, 1959; Murphy et al., 1994), el valor exacto de esta magnitud es aún motivo de controversia. Sin embargo en el proceso de desarrollo de la parametrización estructural empleada en este trabajo (Gómez & Freire, 1995; Murphy et al., 1994) se ha llegado a la conclusión de que el valor que mejor se ajusta a los resultados experimentales es el correspondiente a la entropía crítica definida por Kauzmann (Kauzmann, 1959). De este modo, para la formación de un complejo bimolecular el cambio en entropía translacional es del orden de  $-8 \text{ kcal}\cdot(\text{K}\cdot\text{mol})^{-1}$ , lo que se traduce en una contribución de aproximadamente  $2.4 \text{ kcal}\cdot\text{mol}^{-1}$  a la afinidad de unión a  $25^\circ \text{ C}$ .

#### **IV.B.2.3.2.-Localización estructural de las contribuciones a la energética de unión.**

La validación experimental del análisis termodinámico-estructural de la energética de unión de los distintos ligandos a sus correspondientes proteínas puede llevarse a cabo mediante la determinación de cuatro magnitudes termodinámicas accesibles experimentalmente como son la energía libre de Gibbs o afinidad de unión y los cambios en entalpía, entropía y capacidad calorífica. Mediante la comparación entre los valores calculados haciendo uso de la parametrización estructural de la energética y los obtenidos experimentalmente para estas cuatro magnitudes es posible evaluar su capacidad, no sólo para proporcionar los valores adecuados para la afinidad de unión, sino para captar el delicado balance existente entre las fuerzas que la determinan. Por lo tanto, una vez comprobado que éste es el caso, la construcción de un mapa

estructural que represente la contribución que cada resto individual aporta a la energética de unión se encuentra justificada.

Este mapa estructural está compuesto únicamente por aquellas interacciones que pueden asignarse a restos o átomos específicos en el complejo, de modo que las contribuciones globales como la entropía crática, que únicamente pueden definirse para el sistema en su conjunto, no se encuentran incluidas. Los mapas estructurales pueden definirse en términos de entalpías, entropía, capacidades caloríficas, energías libres de Gibbs, u otras contribuciones más específicas como entropías conformacionales y de solvatación.

Recientemente se han presentado los resultados del análisis termodinámico-estructural de la unión entre la proteasa aspártica Endothiapepsina y su inhibidor de alta afinidad Pepstatina A (Gómez & Freire, 1995), proceso para el que ha sido posible realizar la disección de las diferentes contribuciones entálpicas y entrópicas al cambio en energía libre total, así como estimar las contribuciones de cada resto a la energética global del proceso. De este modo, es posible identificar aquellas regiones de la proteína y del inhibidor que contribuyen más favorablemente a la energía libre de Gibbs e identificar posibles posiciones favorables que puedan facilitar el proceso de diseño.

#### IV.B.2.3.3.-Cálculo de las constantes de estabilidad por resto.

##### a.-Constantes de estabilidad de restos individuales.

Para un resto determinado, la constante de estabilidad aparente  $k_{p,j}$  se define como la relación entre la probabilidades de todos los estados en los que dicho resto se encuentra plegado  $P_{p,j}$  y aquéllas de los estados en que está desplegado  $P_{d,j}$ , de modo que:

$$k_j = \frac{\sum_{\text{estados con resto } j \text{ plegado}} P_i}{\sum_{\text{estados con resto } j \text{ desplegado}} P_i} = \frac{P_{p,j}}{P_{d,j}} \quad (\text{IV.B.5})$$

La constante de estabilidad por resto,  $k_{p,j}$ , es la magnitud que se mediría si fuera posible determinar experimentalmente la estabilidad de una proteína siguiendo el comportamiento individual de cada resto. Como consecuencia, los valores de las constantes de estabilidad individuales permiten estimar las diferencias en estabilidad estructural entre las distintas regiones de la proteína. En muchos casos, los factores de protección para el intercambio de hidrógeno/deuterio obtenidos mediante resonancia magnética nuclear posibilitan la determinación experimental de dichas constantes individuales de estabilidad.

El algoritmo COREX (Hilser & Freire, 1996a; Hilser & Freire, 1996b; Hilser et al., 1997), desarrollado recientemente en el laboratorio del Dr. Freire, permite, mediante la generación de una amplio conjunto de estados parcialmente plegados y la evaluación de sus energías libres

haciendo uso de la parametrización estructural de la energética, el cálculo de la distribución de probabilidad. Mediante este enfoque termodinámico-estadístico, una vez conocidas las probabilidades de cada estado, se dispone de toda la información necesaria para estimar, no sólo las ya mencionadas constantes de estabilidad por resto, sino otras magnitudes termodinámicas de naturaleza macroscópica, como la  $T_m$  y la entalpía asociadas a una transición calorimétrica.

#### ***b.-Aplicación al estudio de la proteasa del virus V.I.H.-1.***

Los valores de las constantes de estabilidad por resto para la proteasa del virus V.I.H.-1 se calcularon de acuerdo con el algoritmo COREX (Hilser & Freire, 1996a; Hilser & Freire, 1996b; Hilser et al., 1997), empleando la estructura de la proteína libre (correspondiente al fichero pdb de código 1HHP) o bien las estructuras de la proteína procedentes de los complejos tras haber eliminado las coordenadas correspondientes al inhibidor. De este modo, y de acuerdo al algoritmo COREX, se han generado 126 496 estados con diversos grados de plegamiento, desde el estado nativo hasta el completamente desplegado, utilizando una ventana de 16 restos que se va desplazando a lo largo de la secuencia de la proteína. La probabilidad relativa de cada uno de los estados generados, que contribuyen a la función de partición, se obtiene mediante el cálculo de su energía libre de Gibbs haciendo uso de la parametrización estructural de la energética de plegamiento, descrita anteriormente (Sección III.2).

### **IV.B.2.4.-Modelado de ligandos peptídicos para proteasas aspárticas.**

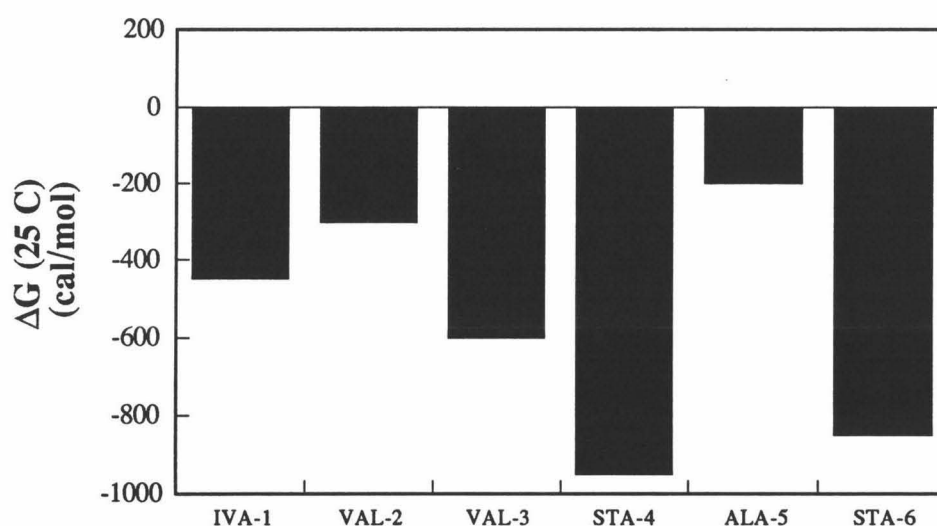
#### ***IV.B.2.4.1.-Diseño de inhibidores peptídicos para Endothiapepsina relacionados con Pepstatina A.***

##### ***a.-Identificación de posiciones de mutación.***

Como se ha comentado anteriormente, el pentapéptido Pepstatina A de secuencia Iva-Val-Val-Sta-Ala-Sta, donde Iva representa al ácido isovalérico, es un potente inhibidor natural de las proteasas aspárticas. Este compuesto contiene dos restos del aminoácido poco usual estatina (ácido 4(S)-amino-3(S)-hidroxi-6-metilheptanoico), de modo que el situado en posición central actúa como un análogo no hidrolizable de los dos restos adyacentes al enlace peptídico que sufre el proceso de hidrólisis en el sustrato. En la figura II.B.3 se muestra una representación esquemática de su composición química y estructura.

Con anterioridad, Gómez y Freire (Gómez & Freire, 1995) llevaron a cabo el análisis experimental y termodinámico estructural del proceso de unión entre este potente inhibidor y la proteasa fúngica Endothiapepsina, por lo que se dispone, no sólo de los valores de la energía libre de unión, sino de las demás magnitudes termodinámicas como entalpía, entropía y capacidad calorífica. Asimismo, se ha descrito la localización estructural de los restos individuales que

contribuyen más notablemente a la afinidad de unión, tanto en la enzima como en el inhibidor. En la figura IV.B.5 se muestran las contribuciones a la energía libre de cada uno de los aminoácidos que constituyen este inhibidor, información en la que nos hemos basado para seleccionar la posición en la que llevar a cabo el análisis mutacional, con el fin de optimizar la afinidad con que se une a la proteína. El resto de Ala en posición 5 es, con diferencia, el más débilmente estabilizador, por lo que constituye un buen candidato para llevar a cabo las diferentes sustituciones, ya que es de esperar que los efectos de cualquier mutación se maximicen en esta posición.



**Figura 4.B.5:** Contribuciones individuales de los restos de Pepstatina A a la afinidad de unión a Endothiapepsina.

#### **b.-Generación de los inhibidores mutantes.**

Una vez identificada la posición óptima en la que realizar las sustituciones, es necesario seleccionar las cadenas laterales que deberán introducirse en dicha posición. Dada la naturaleza hidrofóbica del sitio de unión y la importancia que estas contribuciones tienen en la determinación de la afinidad, de acuerdo con el análisis realizado por Gómez y Freire (Gómez & Freire, 1995), parece lógico proceder a la sustitución de la Ala en posición 5 por restos de creciente hidrofobicidad. De este modo, se han considerado doce posibles mutaciones en dicha posición: Cys, Gly, His, Ile, Leu, Met, Phe, Ser, Thr, Trp, Tyr y Val.

Así mismo, con objeto de mejorar la solubilidad del ligando y facilitar el análisis calorimétrico se decidió llevar a cabo la adición de un resto glutámico en el extremo carboxilo

terminal del péptido, aún a pesar de la predicción preliminar de un descenso en la afinidad de unión.

En todos los casos se emplearon las coordenadas del complejo original, disponible en el banco de datos de proteínas bajo el código 4ER2, como patrón para generar las estructuras de los complejos con los nuevos inhibidores.

#### *c.-Optimización energética de la conformación de los inhibidores.*

Como se ha comentado anteriormente, para cada una de estas mutaciones es necesario identificar la conformación más probable del péptido modificado cuando se encuentra unido a la enzima, proceso que se ha llevado a cabo de acuerdo con el algoritmo de minimización descrito en el apartado IV.B.1.1.3.

De este modo, se procedió a la modificación sistemática de todos los ángulos de torsión de las cadenas laterales entre en un intervalo de  $360^\circ$  ( $\chi_1$  para Cys, Ser, Thr y Val;  $\chi_1$  y  $\chi_2$  para Asn, Asp, His, Ile, Leu, Phe, Trp y Tyr;  $\chi_1$ ,  $\chi_2$  y  $\chi_3$  para Gln, Glu y Met;  $\chi_1$ ,  $\chi_2$ ,  $\chi_3$  y  $\chi_4$  para Lys; y  $\chi_1$ ,  $\chi_2$ ,  $\chi_3$ ,  $\chi_4$ , y  $\chi_5$  para Arg). Para las cadenas laterales con un único ángulo diedro el valor de  $\chi_1$  se ha modificado cada grado; para cadenas con dos o tres diedros cada  $10^\circ$ ; y, por último, para aquellas cadenas con un número mayor de ángulos de torsión, cada  $30^\circ$ . En aquéllos casos en que se requiere la minimización de las conformaciones del esqueleto se han empleado intervalos de  $10^\circ$  para la modificación de los ángulos diedros  $\phi$  y  $\psi$ .

#### *d.-Cálculo de las afinidades de unión.*

La afinidad de unión de los péptidos modificados por la proteína viene determinada por la energía libre de Gibbs que, como se ha descrito anteriormente (sección IV.B.1.1.3), puede obtenerse a partir de las estructuras del complejo, de la proteína y del péptido libres. En todos los casos, para el cálculo de las afinidades de unión, se han considerado las coordenadas correspondientes a la conformación que minimiza la función potencial de Gibbs en cada complejo. Para la proteína libre, se han utilizado las coordenadas espaciales correspondientes al fichero PDB de código 4APE. Para los péptidos libres se han empleado las accesibilidades al disolvente correspondientes a los valores optimizados descritos con anterioridad (sección IV.A.2.2.2) y recogidos en la tabla IV.A.4 (Luque et al., 1996).

La precisión con que se realizan los cálculos depende directamente de la precisión con que se determinan las superficies accesibles al disolvente, tanto del complejo como de la enzima y el inhibidor libres en disolución. Por lo tanto, siguiendo un procedimiento similar al descrito en la sección IV.B.2.3.1.b para el cálculo de las afinidades de unión de los inhibidores no peptídicos, se han considerado 64 orientaciones diferentes con respecto al plano de corte, generadas mediante la rotación de las estructuras alrededor de los ejes de coordenadas X, Y y Z cada  $90^\circ$ , de modo que

las áreas accesibles polares y apolares utilizadas en los cálculos de las afinidades de unión son las correspondientes al promedio entre todas las orientaciones.

#### IV.B.2.4.2.-Modelado de sustratos de la proteasa el virus V.I.H.-1.

El análisis termodinámico estructural del proceso de unión de los sustratos a la proteasa del virus V.I.H.-1 se ve dificultado por la imposibilidad de obtener estructuras cristalográficas de los complejos, debido a la rapidez con el sustrato se hidroliza una vez unido a la enzima. Por lo tanto, y con objeto de proceder al análisis de las interacciones que determinan la afinidad de unión de los sustratos, se procedió a la elaboración de un modelo estructural para sus complejos con la enzima. Las estructuras tridimensionales de los complejos entre los sustratos y la proteasa se han obtenido a partir de las existentes para distintos inhibidores peptídicos (que no son sino análogos no hidrolizables de los sustratos) mediante la aplicación del algoritmo de diseño descrito anteriormente.

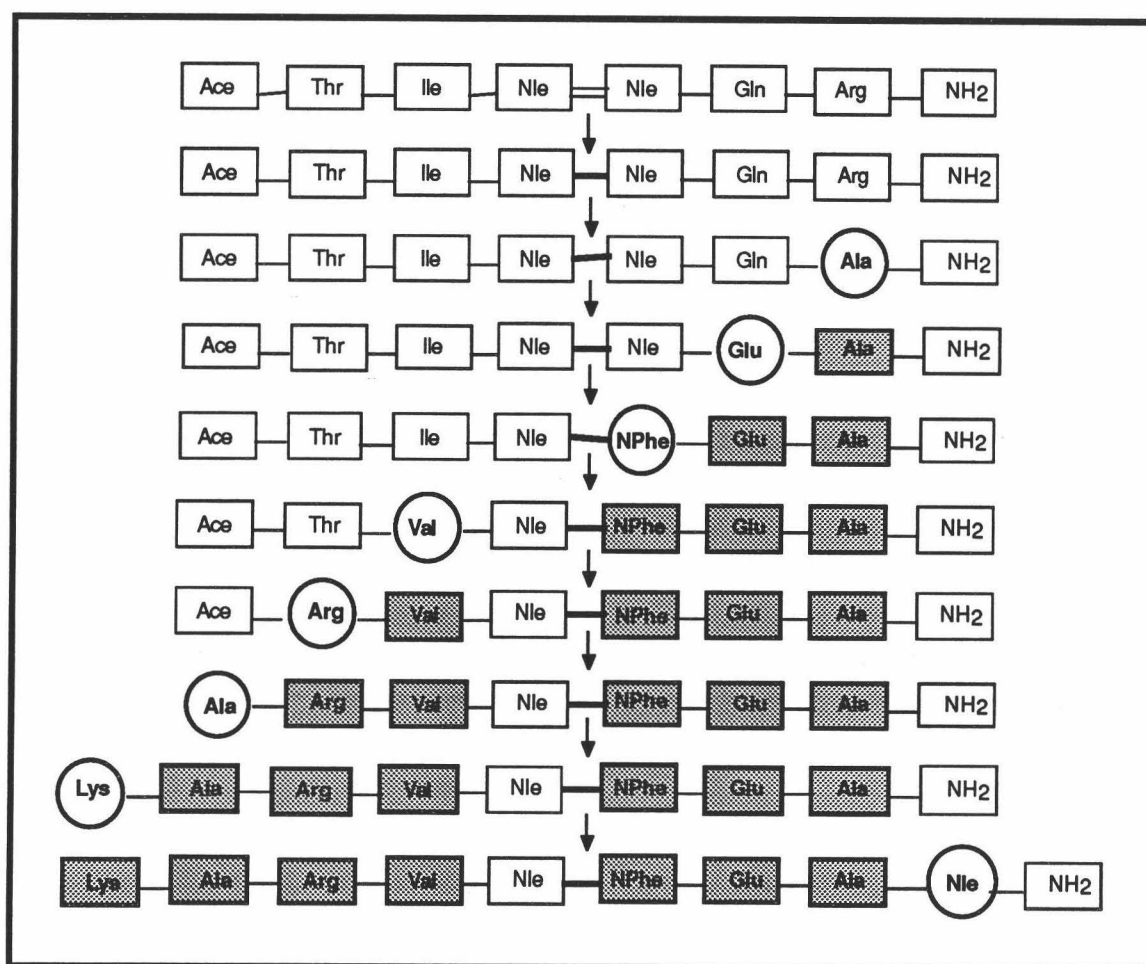
##### a.-Sustrato cromogénico.

En muchos casos la investigación de la cinética enzimática requiere la introducción en el sustrato de un grupo cromogénico, cuyas propiedades espectroscópicas se modifiquen como consecuencia del proceso de hidrólisis, y que, por tanto, permita seguir su curso mediante alguna de las técnicas espectroscópicas habituales. Uno de los sustratos más ampliamente utilizado para la realización de los ensayos enzimáticos con la proteasa del virus V.I.H.-1 es el nonapéptido cromogénico de secuencia Lys-Ala-Arg-Val-Nle-Nphe-Glu-Ala-Nle-NH<sub>2</sub>, donde nPhe y Nle representan a los aminoácidos p-nitro-fenilalanina y norleucina, respectivamente. La estructura tridimensional del complejo entre dicho sustrato y la proteasa se ha generado a partir de las coordenadas correspondientes al complejo con el inhibidor peptídico de secuencia Ace-Thr-Ile-Nle=Nle-Gln-Arg-NH<sub>2</sub> (donde = representa la existencia de un enlace peptídico reducido y Ace un grupo acetato), cuya estructura se ha resuelto a 2.3 Å de resolución y se encuentra depositada en el banco de datos de proteínas bajo el código 4HVP.

El inhibidor posee seis aminoácidos con un enlace peptídico reducido entre los restos de norleucina en posiciones 3 y 4, mientras que el sustrato cromogénico, por otro lado, consta de nueve restos, de los que únicamente se conserva el resto de Nle 3 del inhibidor. El diseño del sustrato se lleva a cabo utilizando el inhibidor peptídico como patrón, mediante la combinación adecuada de sustituciones y adiciones. Al igual que se ha descrito en la sección anterior con respecto a los mutantes de Pepstatina A, cada modificación debe ir acompañada por una etapa de minimización estructural, con el objeto de identificar, haciendo uso de la parametrización estructural de la energética descrita en el capítulo III.2, la conformación más probable adoptada por el nuevo péptido.



En la figura IV.B.6 se muestra un esquema del procedimiento de diseño seguido en este caso, en el que los círculos en negrilla representan los restos que se sustituyen en cada paso y los cuadros sombreados aquéllos que ya han sido modificados y optimizados en etapas anteriores. La primera etapa representa la modificación del enlace entre los restos de norleucina, que se encuentra reducido en el complejo original, utilizando el módulo Biopolymers del programa InsightII de Biosym. Las siguientes etapas, desde el tercer paso hasta el octavo, representan la sustitución de aminoácidos ya existentes en la cadena, llevadas a cabo utilizando el esqueleto peptídico del ligando original como patrón. En estos casos sólo se requiere la minimización posterior de sus cadenas laterales. Por último, las tres etapas siguientes consisten en la adición de un resto en los extremos del péptido, en cuyo caso, se procede a la minimización del aminoácido completo.



*Figura 4.B.6: Procedimiento seguido para el diseño del sustrato cromogénico a partir de la estructura correspondiente al complejo del inhibidor Ace-Thr-Ile-Nle=Nle-Gln-Arg-NH<sub>2</sub>, recogida en el fichero 4HVP.*

### *b.-Sustrato natural.*

El segundo de los sustratos modelados en este trabajo, de secuencia Val-Ser-Gln-Asn-Tyr-Pro-Ile-Val-Gln, representa uno de los ocho sitios fundamentales de corte en los precursores poliproteicos gag y gag-pol (Dunn et al., 1994), representados en la tabla II.B.1. La estructura del complejo de este nonapéptido con la proteasa del virus V.I.H.-1 se ha obtenido utilizando como punto de partida dos estructuras distintas, correspondientes a complejos de la enzima con diferentes inhibidores peptídicos, con el fin de comparar las estructuras finales obtenidas en ambos casos.

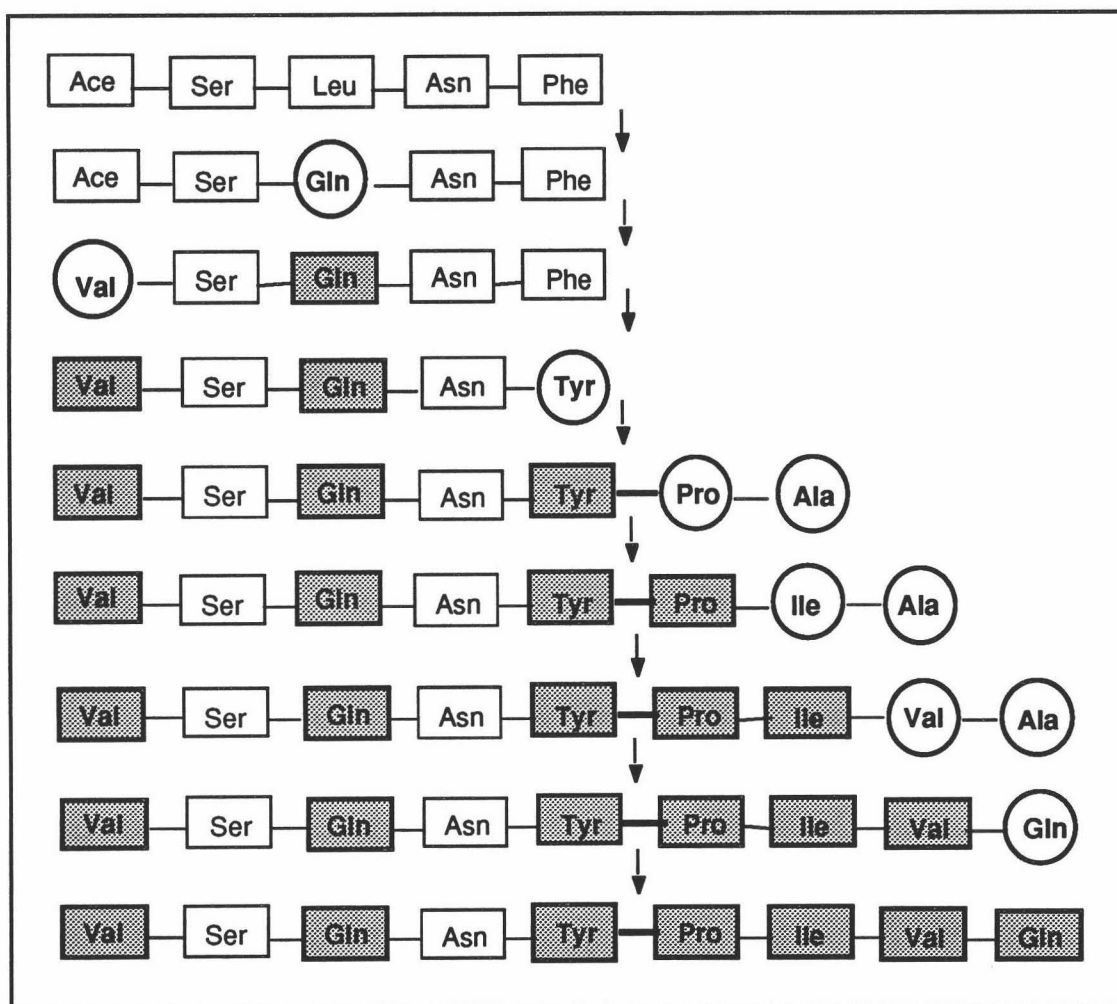
En primer lugar se procedió al modelado de dicho sustrato nonapeptídico a partir del complejo con el inhibidor Ace-Thr-Ile-Nle=Nle-Gln-Arg-NH<sub>2</sub> (caracterizado por la presencia de un enlace reducido entre los dos restos de Norleucina), correspondiente al fichero PDB de código 4HVP, y que ha sido utilizado asimismo para el modelado del sustrato cromogénico. El procedimiento empleado es muy similar al descrito para el caso anterior.

La segunda estructura empleada como patrón es la correspondiente al complejo entre el inhibidor peptídico JG-365, de secuencia Ace-Ser-Leu-Asn-Phe=Pro-Ile-Val-oMe, en el que el enlace peptídico hidrolizable (representado como =) se ha sustituido por un análogo hidroxietil-amina, resistente a la hidrólisis. La estructura de dicho complejo se encuentra depositada en el banco de datos de proteínas bajo el código 7HVP.

El procedimiento llevado a cabo en este caso es ligeramente diferente al anterior ya que, dado que el análogo no hidrolizable posee un átomo adicional con respecto a un enlace peptídico normal, no es posible la utilización del esqueleto completo como guía. Por lo tanto, se decidió tomar como patrón inicial la primera parte del ligando constituido por los restos Ace-Ser-Leu-Asn-Phe y, tras realizar las sustituciones necesarias en dicha porción del péptido (siguiendo un proceso similar al descrito anteriormente), se procedió a la adición de cada uno de los restos siguientes hasta completar la secuencia.

Las adiciones llevadas a cabo hasta el momento constituyeron los extremos del péptido, por lo que no fue necesario considerar la posible influencia que la presencia de restos posteriores en la cadena pudiese tener sobre el resto minimizado. En este caso, sin embargo, dado que la cadena peptídica debe continuar elongándose, se consideró necesario realizar el proceso de optimización conformacional en presencia del resto inmediatamente posterior. Debido a los elevados requerimientos computacionales que la variación sistemática de todos los ángulos de los esqueletos y cadenas laterales de dos restos aminoacídicos suponen, se ha utilizado en todos los casos un resto de Ala en el extremo de la cadena. La figura IV.B.7 muestra esquemáticamente el procedimiento de optimización seguido en este caso.

### *c.-Optimización de la conformación de los restos mutados.*



*Figura 4.B.7: Procedimiento seguido para el diseño, a partir de la estructura del inhibidor JG-365, del sustrato nonapeptídico correspondiente a uno de los sitios de corte en los precursores poliproteicos.*

Como se ha comentado anteriormente, para cada una de estas mutaciones es necesario identificar la conformación que el nuevo péptido adoptará con mayor probabilidad en el complejo, proceso que se ha llevado a cabo de acuerdo con el algoritmo de minimización descrito en el apartado IV.b.1.1.1.

De este modo, se procedió a la modificación sistemática de todos los ángulos de torsión de las cadenas laterales en un intervalo de  $360^\circ$  ( $\chi_1$  para Cys, Ser, Thr y Val;  $\chi_1$  y  $\chi_2$  para Asn, Asp, His, Ile, Leu, Phe, Trp y Tyr;  $\chi_1$ ,  $\chi_2$  y  $\chi_3$  para Gln, Glu y Met;  $\chi_1$ ,  $\chi_2$ ,  $\chi_3$  y  $\chi_4$  para Lys; y  $\chi_1$ ,  $\chi_2$ ,  $\chi_3$ ,  $\chi_4$ , y  $\chi_5$  para Arg). Para las cadenas laterales con un único ángulo diedro el valor de  $\chi_1$  se ha modificado cada grado, para cadenas con dos o tres diedros cada  $10^\circ$  y por último para cadenas con un número mayor de ángulos de torsión cada  $30^\circ$ . En aquéllos casos en que se requiere la

minimización de las conformaciones del esqueleto se han empleado intervalos de  $10^\circ$  para la modificación de los ángulos diedros  $\phi$  y  $\psi$ .

#### *d-Cálculo de las afinidades de unión.*

La afinidad de unión de los nuevos péptidos por la proteína, determinada por la energía libre de Gibbs, se calcula a partir de las estructuras del complejo, de la proteína libre y del péptido libre. En todos los casos, los cálculos se han realizado sobre las coordenadas correspondientes a la conformación que minimiza la función potencial de Gibbs para cada uno de los sustratos. Asimismo, se han utilizado las coordenadas para la proteína libre, recogidas en el fichero PDB de código 1HHP, modificado previamente de acuerdo con los cánones descritos en la sección IV.B.2.2.1.a. Para los sustratos peptídicos libres en disolución se han empleado las accesibilidades al disolvente correspondientes a los valores optimizados con anterioridad (Luque et al., 1996) como se ha descrito en la sección IV.A.2.2.2) (Luque et al., 1996).

De igual forma, y siguiendo un proceso idéntico al llevado a cabo en ocasiones anteriores, los cálculos de accesibilidad se han realizado considerando 64 orientaciones diferentes de cada una de las moléculas implicadas con respecto al plano de corte, generadas mediante la rotación de las estructuras alrededor de los ejes de coordenadas X, Y y Z cada  $90^\circ$ , de modo que para cada átomo las áreas accesibles polares y apolares consideradas en los cálculos son las correspondientes al promedio entre todas las orientaciones.

### **IV.B.3.-Resultados y discusión.**

#### **IV.B.3.1.-Minimización de conformaciones helicoidales.**

Los perfiles energéticos en función de los ángulos diedros de la cadena lateral para cada uno de los aminoácidos naturales en la posición central de una hélice alfa se han obtenido mediante la aplicación del algoritmo de minimización estructural descrito en la sección IV.B.2.1.2. Como se ha comentado anteriormente, la utilización de la parametrización estructural permite la obtención de los valores en función de los ángulos diedros, no sólo de la función potencial de Gibbs, sino de las restantes magnitudes termodinámicas de interés. En la figura IV.B.8 se muestran los perfiles de energía potencial de Gibbs obtenidos para cada uno de los aminoácidos, donde se representa el valor de energía de cada conformación relativa a la correspondiente a la conformación original. Se puede observar que, en efecto, la parametrización estructural de la energética posee la resolución y precisión necesarias para permitir la correcta identificación de los mínimos energéticos.

*Figura 4.B.8.A: Perfiles de  $\Delta G_{ef}$  en función de los ángulos  $\chi_1$  y  $\chi_2$  para los aminoácidos Arg, Asn, Asp y Gln. (Página 197)*

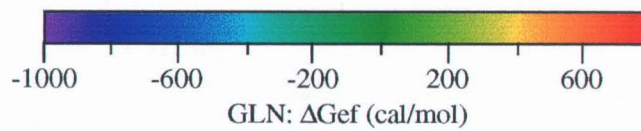
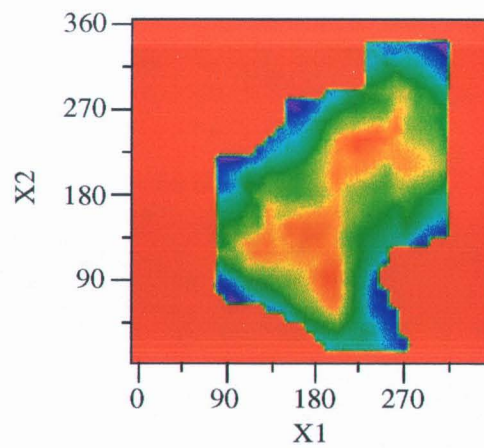
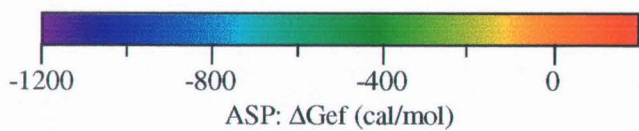
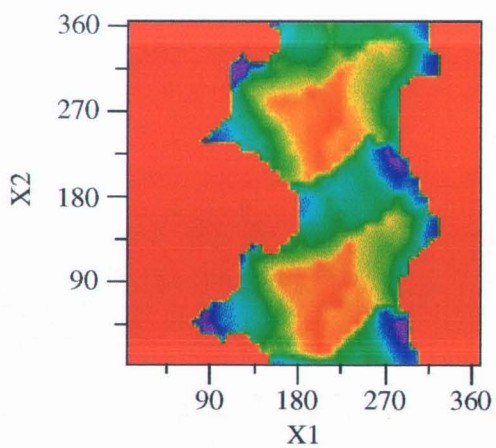
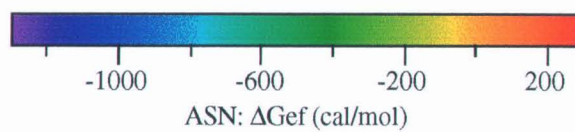
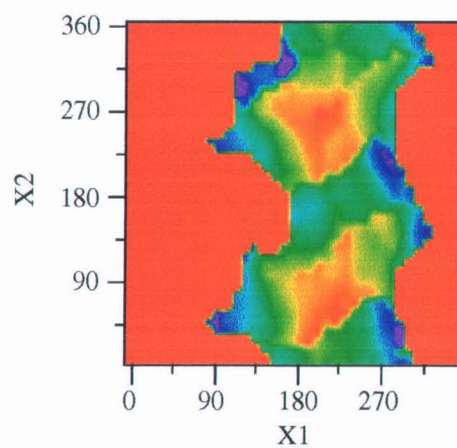
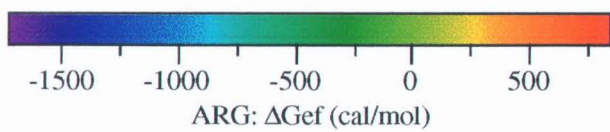
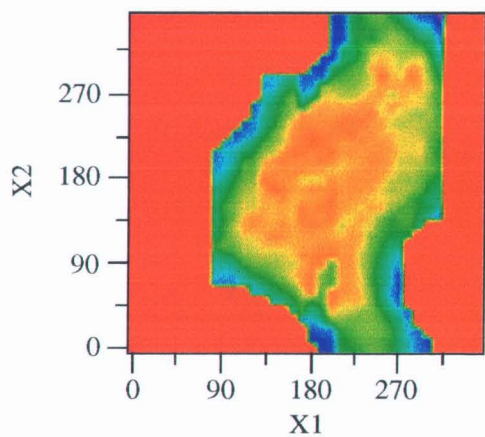
*Figura 4.B.8.B: Perfiles de  $\Delta G_{ef}$  en función de los ángulos  $\chi_1$  y  $\chi_2$  para los aminoácidos Glu, His, Ile, y Leu. (Página 199)*

*Figura 4.B.8.C: Perfiles de  $\Delta G_{ef}$  en función de los ángulos  $\chi_1$  y  $\chi_2$  para los aminoácidos Lys, Met, Phe, y Trp. (Página 201)*

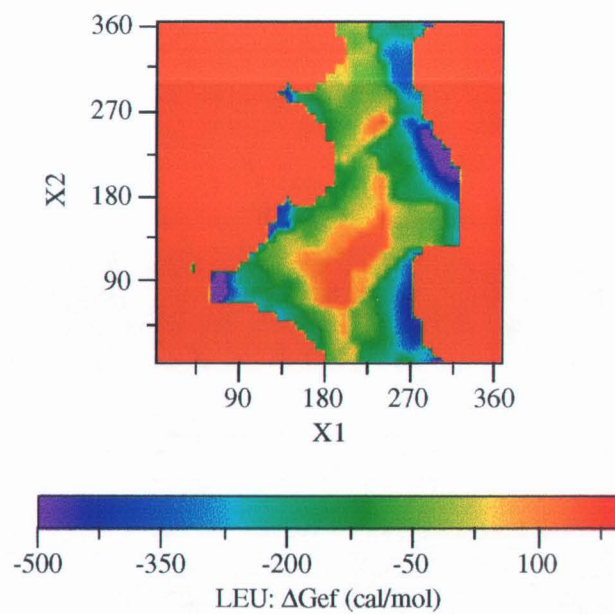
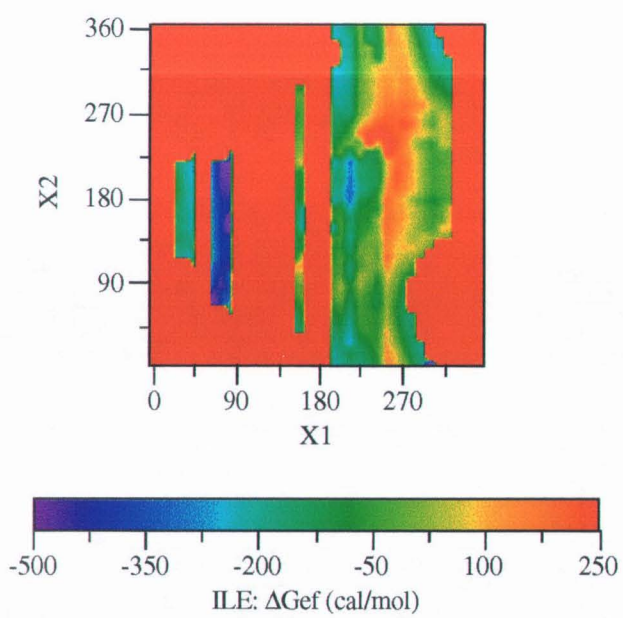
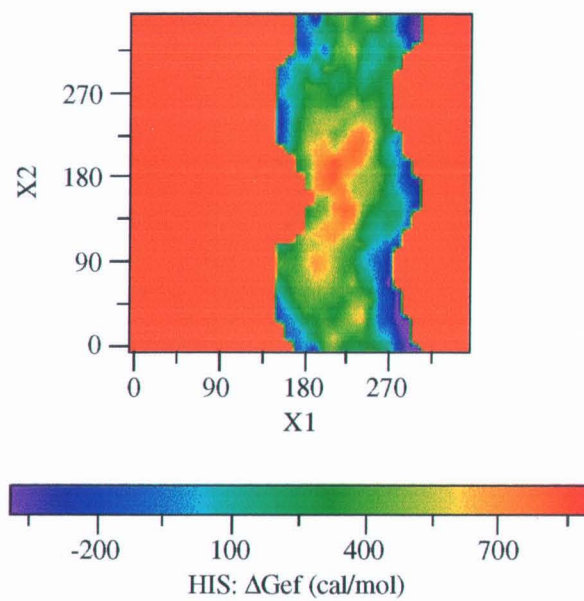
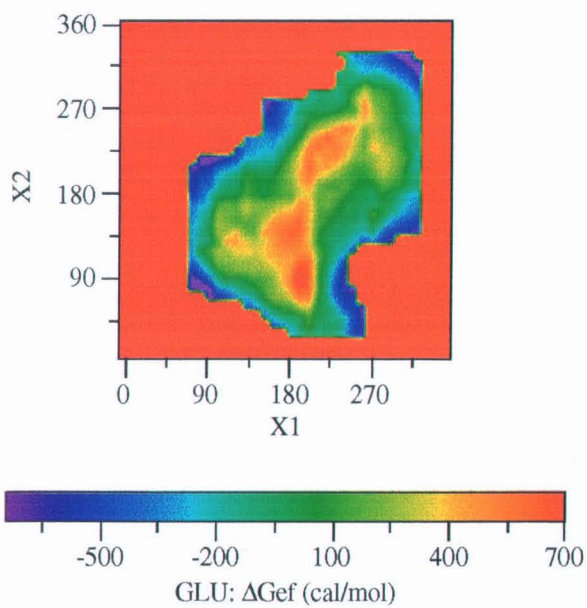
*Figura 4.B.8.D: Perfiles de  $\Delta G_{ef}$  en función de los ángulos  $\chi_1$  y  $\chi_2$  para el aminoácido Tyr. (Página 203)*

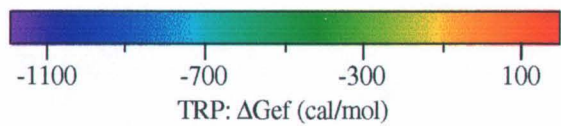
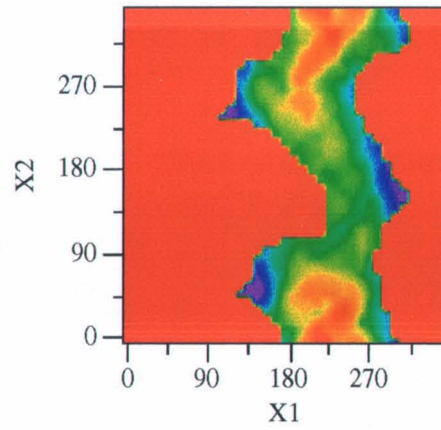
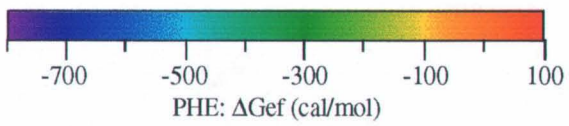
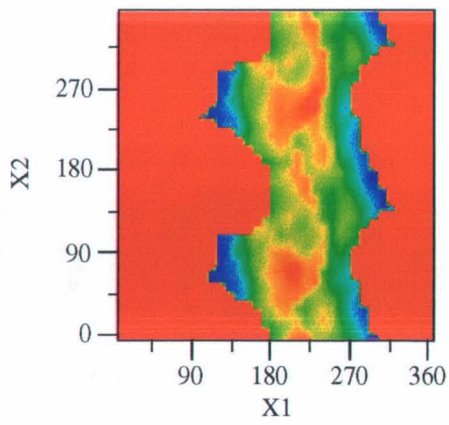
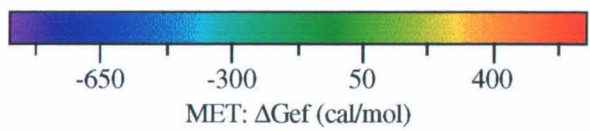
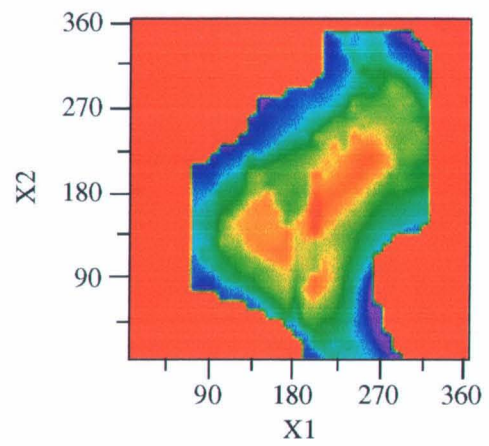
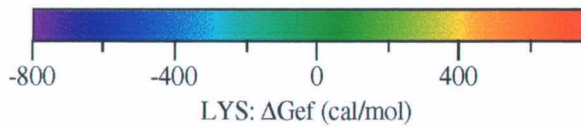
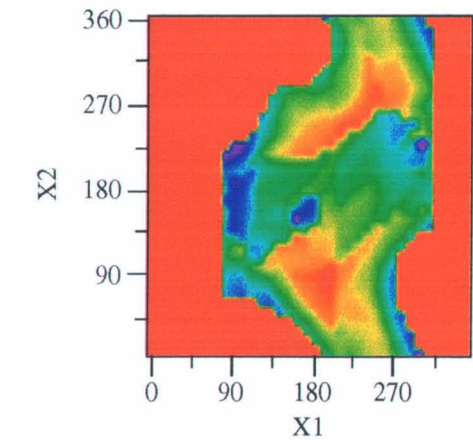
*Se muestran únicamente los restos con dos o más ángulos diedros en su cadena lateral. En los perfiles correspondientes a aminoácidos con más de dos ángulos, los posteriores a  $c_2$  se han mantenido constantes a  $180^\circ$ . El código de colores es el mostrado en las barras bajo cada figura, de modo que las zonas en rojo representan las regiones del espacio conformacional no permitidas estéricamente para cada cadenas lateral*

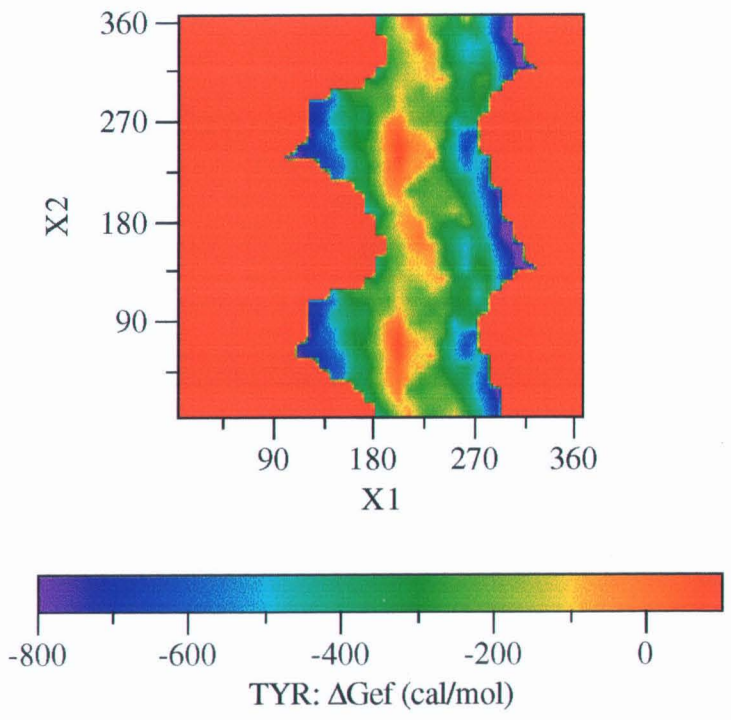












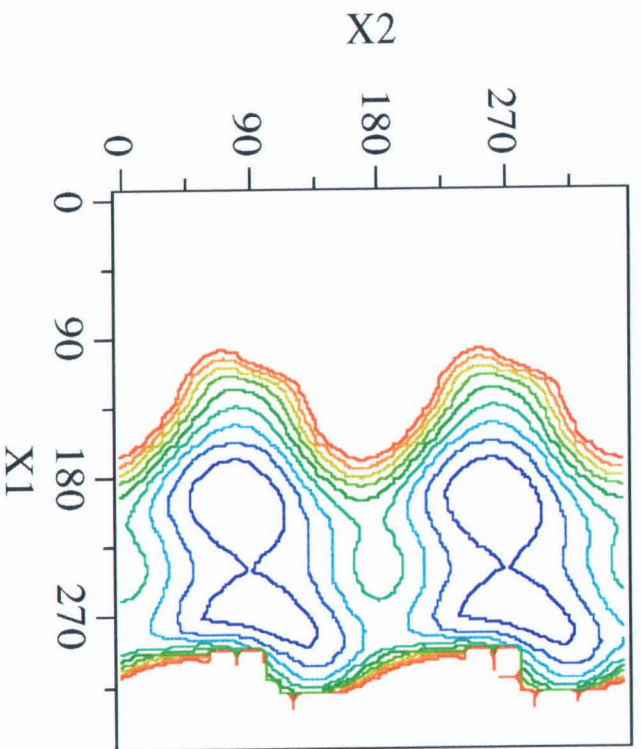
La mayoría de los perfiles energéticos disponibles en la literatura se han obtenido mediante la utilización de funciones potenciales estándar, que han sido parametrizadas a partir de datos experimentales obtenidos en vacío, por lo que únicamente proporcionan información acerca de la entalpía del sistema en función de su conformación. Con el objeto de permitir la comparación entre los resultados obtenidos utilizando este tipo de funciones y los proporcionados por la parametrización estructural es necesario, por tanto, recurrir al análisis de los perfiles entálpicos, aún a pesar de que los valores de entalpía proporcionados por la parametrización estructural contienen las contribuciones debidas los fenómenos de hidratación, no siempre considerados en los perfiles existentes en la literatura.

En la figura IV.B.9 se muestran los perfiles de entalpía en función de los ángulos diedros  $\chi_1$  y  $\chi_2$  correspondientes al caso concreto del aminoácido Phe, junto con la reproducción de uno de los perfiles descritos por otros autores en la literatura (Janin et al., 1978; Lee et al., 1994) para este mismo aminoácido. El panel de la derecha corresponde a los resultados obtenidos en este trabajo y que, al igual que los mostrados anteriormente para  $\Delta G_{\text{Gef}}$ , representan los valores de entalpía de cada conformación relativos a la de partida. El panel de la izquierda representa los valores de energía obtenidos por otros autores (Lee et al., 1994) utilizando funciones potenciales estándar. En el caso concreto del perfil mostrado en esta figura, los valores de energía se obtuvieron mediante el empleo de los parámetros y funciones potenciales propuestos por Weiner y colaboradores (Weiner et al., 1984) y por Jorgensen y Tirado-Rives (Jorgensen & Swenson, 1985), incluyendo las interacciones electrostáticas entre cargas parciales, que se calcularon asignando un valor de 78 a la constante dieléctrica del disolvente (Lee et al., 1994). Como puede observarse, la localización de los mínimos y submínimos en los perfiles obtenidos en este trabajo se encuentra perfectamente definida y en concordancia con los resultados obtenidos por otros autores

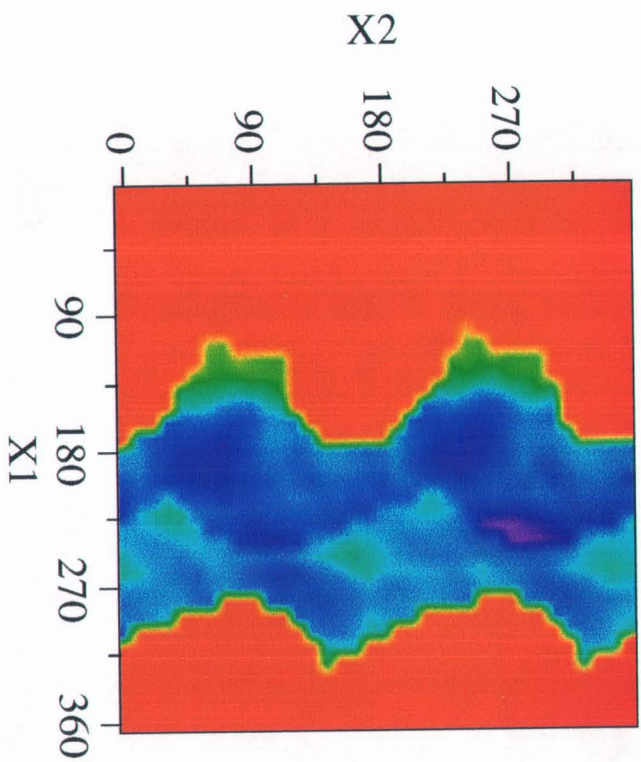
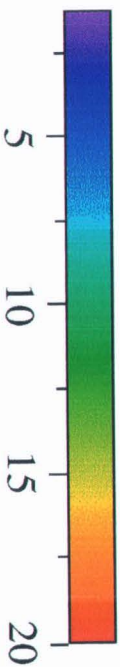
El análisis de los perfiles correspondientes a las distintas magnitudes termodinámicas para cada uno de los aminoácidos permite extraer ciertas conclusiones sobre el papel de las distintas interacciones en la determinación de la conformación preferentemente adoptada por la cadena lateral bajo unas condiciones determinadas. Así, en el caso de aquellas cadenas laterales que carecen de grupos polares y poseen, por tanto, un marcado carácter hidrofóbico, puede observarse cómo en ausencia de interacciones estabilizadoras con otras cadenas laterales (como es el caso de estas hélices monoméricas de polialanina) las conformaciones preferidas entálpicamente son aquellas en las que la cadena lateral se encuentra extendida y alejada del resto de la hélice; de modo que se minimizan los posibles efectos estéricos desfavorables como consecuencia de la presencia de los otros grupos de la molécula y se optimizan sus interacciones de van der Waals con el disolvente. Por el contrario, debido a la naturaleza hidrofóbica de estas cadenas laterales, dichas conformaciones se encuentran muy desfavorecidas entrópicamente, por lo que, como resultado del fenómeno de compensación entalpía/entropía, llevan asociadas los valores máximos en energía de Gibbs, por lo que deben encontrarse muy poco pobladas en disolución acuosa.

*Figura 4.B.9: Perfiles de entalpía en función de los ángulos diedros  $\chi_1$  y  $\chi_2$  correspondientes al aminoácido Fenilalanina. La figura de la izquierda muestra los valores obtenidos por Lee y colaboradores (Lee et al., 1994) y la de la derecha los valores de entalpía calculados en este trabajo haciendo uso de la parametrización estructural de la energética. Los colores corresponden a los códigos mostrados en las barras bajo cada figura. Las regiones del espacio conformacional no permitidas estéricamente se representan en color rojo en la figura de la derecha.*





$\Delta H$  [Lee et al.] (kcal/mol)



$\Delta H$  (25 C) (cal/mol)





El análisis de estos efectos en el caso de las cadenas laterales de carácter polar resulta más complejo debido a la coexistencia de grupos hidrofóbicos y polares en la cadena, lo que complica la interpretación de los efectos compensatorios entre las distintas contribuciones. Sin embargo, en cualquier caso, es importante resaltar la importancia que tiene la consideración de los efectos de hidratación en este tipo de estudios, ya que, como se ha discutido, puede resultar esencial en la determinación de la probabilidad de una conformación determinada. Este es uno de los motivos fundamentales por los que el empleo de formalismos como la parametrización estructural de la energética, que incluye implícitamente los efectos asociados con los fenómenos de hidratación, puede proporcionar una valiosa alternativa para el desarrollo de técnicas efectivas de diseño molecular.

#### **IV.B.3.2.-Estudio termodinámico-estructural de la unión de inhibidores a la proteasa del virus V.I.H.-1**

##### *IV.B.3.2.1.-Consideraciones generales.*

Como se ha discutido en la sección II.B.4, la proteasa del virus V.I.H.-1 se ha convertido en uno de los principales objetivos en la lucha contra la infección debido a la importancia que su función tiene para el proceso de maduración del virus. Los tratamientos clínicos que incorporan algunos de los inhibidores de esta proteína, recientemente desarrollados, han producido resultados muy prometedores. En particular, las terapias combinadas, consistentes en la utilización simultánea de varios inhibidores de la proteasa, o bien en su administración conjunta con inhibidores nucleosídicos de la retrotranscriptasa, son de especial interés. Sin embargo, la aparición de cepas víricas, que muestran resistencia a muchos de estos fármacos ha consituido un revés importante para el desarrollo de tratamientos terapéuticos de la enfermedad (Condra et al., 1995; Ho et al., 1994; Kaplan et al., 1994; Roberts, 1995; Tisdale, 1996). La pérdida de sensibilidad hacia los inhibidores tiene su origen en el desarrollo de nuevas cepas del virus que codifican moléculas de proteasa con mutaciones específicas, que tienen como resultado la disminución de la afinidad de la enzima por los inhibidores, pero que, sin embargo, mantienen tanto su afinidad por los sustratos, como su actividad catalítica a niveles suficientes para la supervivencia del virus.

Los orígenes de la resistencia aún permanecen oscuros, y se han observado, tanto mutaciones localizadas en las cercanías del sitio de unión, como otras situadas en regiones muy alejadas del centro catalítico y de efectos desconocidos. Además, los distintos inhibidores producen patrones mutacionales diferentes, de modo que, a pesar de que todos ellos han sido diseñados con el mismo centro activo como objetivo, se han observado patrones de resistencia múltiple bastante variados. El desarrollo de una nueva generación de inhibidores de la proteasa, que resuelva de modo efectivo el problema de la resistencia, requiere necesariamente la

comprensión de las interacciones implicadas en la determinación de la afinidad de la enzima, tanto por los inhibidores, como por sus sustratos.

En este contexto, el siguiente apartado contiene una descripción de los resultados obtenidos de la aplicación de la metodología descrita en las secciones anteriores al estudio de la asociación de 13 inhibidores de la proteasa del virus V.I.H.-1 (A-77003, A-78791, A-76928, A-74704, A-76889, VX-478, SB-203386, SB-203238, U-100313, U-89360, A-98881 y CGP-53820), para cuyos complejos con la enzima se dispone de estructuras tridimensionales de alta resolución (Abdel-Meguid et al., 1994; Erickson et al., 1990; Fassler et al., 1993; Hoog et al., 1995; Kim et al., 1995; Lin et al., 1995; Madhusoodan et al., 1994; Thaisrivongs et al., 1995; Thompson et al., 1994), así como de sus constantes de inhibición.

La secuenciación de las proteínas mutantes, aisladas de pacientes bajo tratamiento con este tipo de inhibidores, ha permitido la identificación y localización de las mutaciones características de las cepas resistentes al virus, pero no ha proporcionado una descripción molecular del origen de dicha resistencia. En general, las mutaciones observadas en la proteasa del virus V.I.H.-1 pueden ejercer su efecto en la energética de unión mediante una interacción directa con el inhibidor (como es el caso de las mutaciones localizadas en las cercanías del sitio de unión). Alternativamente, una mutación puede modular la afinidad de la enzima por un inhibidor determinado mediante efectos cooperativos, como ocurre en aquellos casos en los que el aminoácido mutado no está en contacto directo con el inhibidor, pero modifica algunas de las interacciones existentes entre los restos de la enzima, de forma que altera el patrón de interacciones entre ésta y el inhibidor.

Así, por ejemplo, algunas mutaciones se encuentran localizadas en la región de los "flaps" o en la base de éstos, de modo que ocupan posiciones muy alejadas del sitio de unión (L63P ó A71V). Como se discutido anteriormente (ver sección II.B.4.2), la región del "flap" se encuentra esencialmente desordenada en la enzima libre, por lo que, si la mutación induce un cambio substancial en la barrera energética que debe ser superada para adoptar la conformación competente para la unión del inhibidor, la afinidad de unión se verá afectada, incluso en el caso de que la mutación se encuentre alejada del sitio de unión. Se ha postulado que ciertas mutaciones como L63P o A71V podrían producir una disminución en los grados de libertad de esta región, haciendo ciertas conformaciones inaccesibles o energéticamente desfavorables. Es también posible que las variaciones en afinidad aparezcan como consecuencia de una combinación de los dos efectos anteriores.

El análisis termodinámico-estructural de la unión de los distintos inhibidores realizado en este trabajo proporciona una descripción detallada de la energética de unión de inhibidores de distinta naturaleza, así como de la localización de los restos de la proteína y del ligando determinantes de la afinidad de unión. Con objeto de profundizar en el estudio de los determinantes del fenómeno de la resistencia, se ha abordado el estudio de los efectos que una de las mutaciones más frecuentemente observadas, V82A, tiene sobre las distintas interacciones

establecidas con el inhibidor no peptídico A-77003 y, por lo tanto, sobre su afinidad (Baldwin et al., 1995b). De este modo, se ha procedido al análisis de la influencia de esta mutación sobre las contribuciones de los restos implicados en el proceso de unión del inhibidor a la enzima.

De forma complementaria, el estudio de las características termodinámicas del proceso de unión de los sustratos de estas enzimas permitiría la identificación de los rasgos diferenciales de los dos tipos de ligandos, que posibilitan la aparición del fenómeno de resistencia. Como un primer paso, el análisis termodinámico-estructural de inhibidores peptídicos, diseñados como análogos de estos sustratos, aunque con el enlace peptídico hidrolizable sustituido, puede proporcionar información interesante acerca de las características termodinámicas del proceso de unión de péptidos a estas proteasas. En este sentido, se ha realizado el estudio del proceso de unión a la proteasa del virus V.I.H.-1 del inhibidor universal de las proteasas aspárticas, Acetil-pepstatina, que nos permite comparar los parámetros termodinámicos correspondientes a este polipéptido, con los que caracterizan la unión de los inhibidores no peptídicos, así como con los correspondientes a la unión del mismo inhibidor con Endothiapepsina, llevados a cabo previamente por Gómez y colaboradores (Gómez & Freire, 1995).

#### *IV.B.3.2.2.-Inhibidores no peptídicos.*

##### *a.-Predicción de afinidades de unión.*

Como se ha mencionado anteriormente, para llevar a cabo los cálculos debe disponerse de la estructura tridimensional de los complejos entre la proteína y los inhibidores así como de las estructuras de uno y otro libres en disolución. Dada su pequeña longitud y su relativa rigidez, se considera que los inhibidores no peptídicos libres en disolución adoptan la misma conformación que presentan en el complejo, de modo que, para realizar los cálculos de las afinidades de unión, se utilizan sus coordenadas en el fichero PDB del complejo, pero en ausencia de las coordenadas correspondientes a la proteína. Con objeto de poder realizar comparaciones posteriores con los resultados obtenidos de los cálculos con los mutantes, para los que no se dispone de la estructura libre en disolución, en los cálculos se ha considerado, tanto la estructura tridimensional de la proteína libre (fichero PDB de código 1HHP), como la estructura de la proteína en el complejo en ausencia de las coordenadas correspondientes al inhibidor.

Los resultados de la aplicación de la parametrización estructural de la energética al cálculo de las afinidades de unión de los distintos inhibidores no peptídicos de la proteasa del virus V.I.H.-1 se recogen en la tabla IV.B.2. En la figura IV.B.10 se muestran los valores de afinidad de unión calculados (barras negras) en comparación con los resultados experimentales (barras sombreadas) para los 13 inhibidores. El análisis estadístico de los datos revela que las energías libres de unión se calculan con una desviación estándar de  $\pm 1$  kcal·mol<sup>-1</sup>, correspondiente a una incertidumbre relativa de  $\pm 10\%$ , y a un error estándar de 0.3 kcal·mol<sup>-1</sup>. El análisis de regresión lineal entre los valores de  $\Delta G$  calculados y los experimentales origina una pendiente de 0.982 con

un coeficiente de correlación de 0.85, de tal forma que no se observan desviaciones sistemáticas entre los datos. Por lo tanto, los cálculos estructurales parecen poseer la precisión necesaria como para permitir el examen y análisis de las diferentes contribuciones a la energética de unión.

*b.-Diseción de la contribución de las distintas magnitudes termodinámicas a la afinidad de unión.*

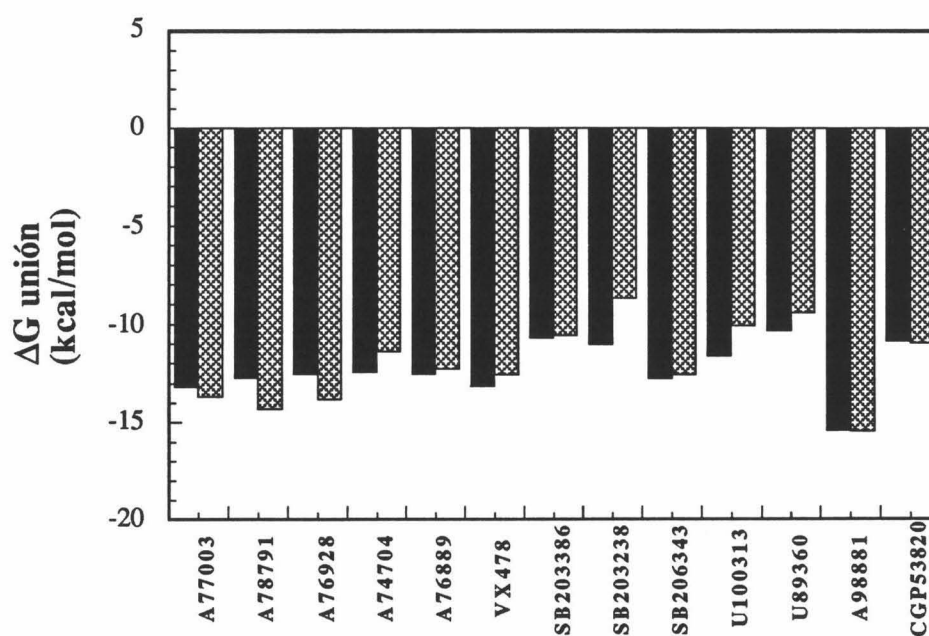
De acuerdo con los valores obtenidos para las distintas magnitudes termodinámicas recogidos en la tabla IV.B.2, el proceso de unión de cada uno de los trece inhibidores se encuentra dominado por el efecto hidrofóbico. A consecuencia de la formación del complejo, una porción considerable, tanto de los inhibidores, como de los restos de la proteína localizados en las cercanías del sitio de unión, quedan protegidos del contacto con el disolvente, de modo que disminuye significativamente la superficie apolar expuesta al medio. De hecho, la fracción de área apolar que se oculta al disolvente tras la unión de los inhibidores es, en promedio,  $0.737 \pm 0.02$ , valor considerablemente mayor que el área promedio protegida del disolvente en una proteína globular típica como consecuencia del proceso de plegamiento (entre 0.55 y 0.60). Por tanto, no resulta sorprendente que las mayores contribuciones a la energía libre de Gibbs de unión procedan de los cambios entrópicos favorables, que se originan como consecuencia de la deshidratación de las superficies apolares.

Por otra parte, las contribuciones entálpicas son desfavorables a temperatura ambiente, dado que se encuentran dominadas, asimismo, por la contribución positiva correspondiente a la deshidratación de las superficies hidrofóbicas. Conviene hacer notar que los valores de entalpía mostrados en la tabla IV.B.2 incluyen únicamente contribuciones genéricas (no contienen efectos de protonación u otros efectos relacionados con el tampón de trabajo, que deben considerarse independientemente y de forma explícita), por lo que no pueden ser comparados directamente con los valores experimentales (ver sección II.2.3.I). La entalpía genérica está compuesta por dos contribuciones de carácter opuesto: un componente favorable, que tiene su origen en la formación de interacciones de van der Waals, enlaces de hidrógeno y otras interacciones entre el inhibidor y la proteasa; y otro desfavorable procedente de la desolvatación de las superficies que entran en contacto. Debido al carácter marcadamente hidrofóbico de los inhibidores considerados, el término dominante es el de hidratación, lo que se traduce en una contribución entálpica desfavorable a la energía libre de unión. Ésta, sin embargo, no es una situación universal, como ha quedado demostrado mediante el estudio de la unión de ciertos inhibidores peptídicos de las proteasas aspárticas, que presentan un valores de entalpía favorables al proceso de unión bajo ciertas condiciones experimentales (Gómez & Freire, 1995).

La magnitud de los cambios de capacidad calorífica es similar a los determinados para otros inhibidores de estas proteasas (Gómez & Freire, 1995), y se encuentran, a su vez, dominados por los cambios en solvatación de los grupos polares y apolares. La existencia de

**Tabla IV.B.2: Parámetros termodinámicos resultantes del análisis termodinámico-estructural de la unión de los distintos inhibidores no peptídicos a la proteasa del virus HIV-1**

Inhibidor	$\Delta C_p$ (cal/Kmol)	$\Delta H_{gen}$ (cal/mol)	$\Delta S_{solv}$ (cal/Kmol)	$\Delta S_{conf}$ (cal/Kmol)	$\Delta G_{gen}$ (cal/mol)	$\Delta G_{otros}$ (cal/mol)	$\Delta G_{total}$ (cal/mol)	$\Delta G_{exp}$ (cal/mol)	$\Delta \Delta G$ (cal/mol)
A-77003	-397	11 733	115.2	-23.7	-15 552	2385	-13 167	-13 700	532
A-78791	-400	12 183	115.8	-24.2	-15 127	2385	-12 742	-14 300	1557
A-76928	-392	11 853	113.4	-23.6	-14 932	2385	-12 547	-13 797	1249
A-74704	-379	11 254	110.1	-20.8	-15 378	2953	-12 425	-11 387	-1037
A-76889	-387	11 303	112.7	-23.7	-15 229	2680	-12 549	-12 278	-271
VX-478	-320	8 641	93.9	-11.1	-16 046	2903	-13 143	-12 580	-563
SB-203386	-343	10 410	96.2	-16.8	-13 263	2555	-10 688	-10 565	-123
SB-203238	-320	8 641	94.0	-18.1	-13 959	2918	-11 041	-8 685	-2356
SB-206343	-332	8 109	98.9	-19.8	-15 481	2724	-12 757	-12 580	-177
U-100313	-317	8 471	93.4	-16.6	-14 416	2807	-11 608	-10 077	-1531
U-89360	-236	1 255	76.5	-28.0	-13 211	2877	-10 334	-9 441	-893
A-98881	-293	7 512	86.6	-1.0	-18 018	2615	-15403	-15 417	14
CGP-53820	-294	6 294	88.7	-22.3	-13 504	2643	-10 861	-10 976	115



**Figura 4.B.10: Energías libres de unión de los distintos inhibidores no peptídicos a la proteasa del virus V.I.H.-1. Las barras sombreadas representan los valores obtenidos experimentalmente y las barras negras los valores calculados en este trabajo, de acuerdo con el procedimiento descrito en el texto.**



contribuciones significativas procedentes de otras interacciones es poco probable. (Gómez et al., 1995).

Como se puede observar en la tabla IV.B.2. los términos correspondientes a los efectos estructurales y de solvatación, recogidos en el valor de  $\Delta G_{gen}$ , aportan las contribuciones más significativas a la energía libre de Gibbs correspondiente al proceso de unión de los distintos inhibidores. Todos ellos poseen un marcado carácter hidrofóbico y carecen de grupos polares, por lo que las interacciones de naturaleza electrostática no contribuyen significativamente a la determinación de la afinidad de unión. La única excepción proviene del cambio en el ambiente que rodea a los grupos Asp25, Asp29 y Asp30 de ambas cadenas, expuestos al disolvente en la proteína libre y que, a consecuencia de la unión del inhibidor, pasan a estar inmersos en un medio altamente hidrofóbico. Esta contribución, reflejada en la ecuación III.51 del apartado III.2.4, y que asciende hasta  $0.7 \text{ kcal}\cdot\text{mol}^{-1}$ , dependiendo del inhibidor, tiene su origen en la transferencia de un grupo cargado desde el medio acuoso hasta otro de menor constante dieléctrica (de acuerdo con los resultados obtenidos por García-Moreno y colaboradores (García-Moreno et al., 1997), la constante dieléctrica en el interior de una proteína no es inferior a 15, oscilando en las distintas regiones de la proteína entre 15 y 78.5 en función de su accesibilidad al disolvente).

*c.-Localización estructural de los restos de la proteasa que contribuyen más notablemente a la afinidad de unión.*

Como se ha discutido con anterioridad (Gómez & Freire, 1995), el análisis termodinámico estructural del proceso de unión de ligandos a proteínas ofrece la posibilidad de estimar las contribuciones de cada resto a la constante de unión. En el caso de la unión de los trece inhibidores peptídicos estudiados en este trabajo, se ha observado que las contribuciones a la afinidad de unión proceden esencialmente del mismo conjunto de restos de la enzima, aunque la magnitud de dichas contribuciones varía para cada uno de los casos. Este comportamiento refleja el hecho de que todos estos compuestos han sido diseñados teniendo como objetivo la misma región en la molécula.

La tabla IV.B.3 recoge aquellos aminoácidos de la enzima que contribuyen con más de  $0.1 \text{ kcal}\cdot\text{mol}^{-1}$  (al menos para uno de los inhibidores) a la energía libre total correspondiente al proceso de unión. Los valores mostrados en la tabla no incluyen la contribución correspondiente al inhibidor (que aproximadamente asciende a un 55% de la energía libre de Gibbs total) ni la entropía translacional que, siendo una propiedad del sistema en su conjunto, no puede ser adscrita a un resto en particular. La posición de estos aminoácidos en la proteasa se representa en el panel superior de la figura IV.B.11, donde los restos cuyas contribuciones se encuentran entre  $-0.7$  y  $-0.9 \text{ kcal}\cdot\text{mol}^{-1}$  se muestran en color rojo, los que contribuyen con valores de energía libre entre  $-0.5$  y  $-0.7 \text{ kcal}\cdot\text{mol}^{-1}$  en naranja, aquéllos entre  $-0.3$  y  $-0.5 \text{ kcal}\cdot\text{mol}^{-1}$  en amarillo y, por último, los que aportan entre  $-0.1$  y  $-0.3 \text{ kcal}\cdot\text{mol}^{-1}$ , en verde.



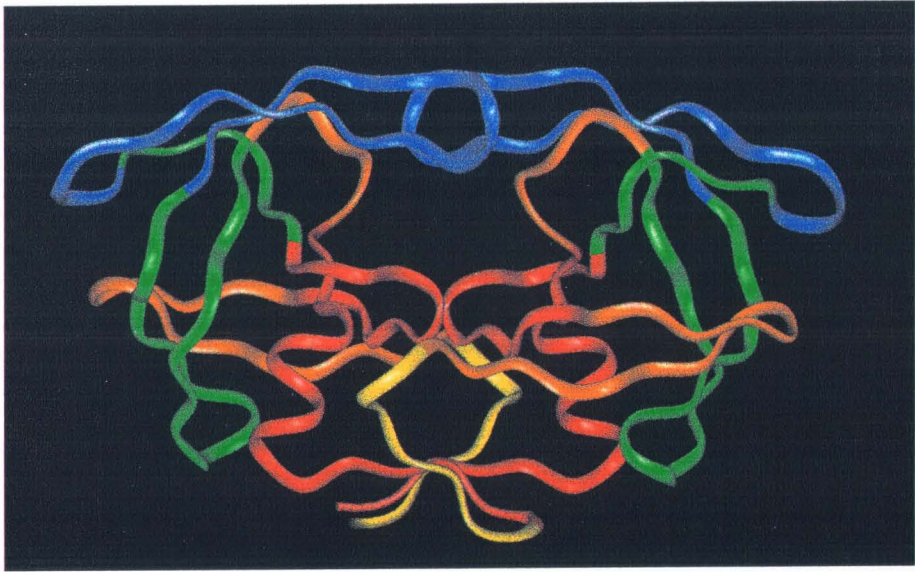
Desde un punto de vista energético, el sitio de unión se encuentra constituido por aminoácidos pertenecientes a cuatro regiones no contiguas en la secuencia: algunos restos situados en la región cercana a los grupos Aspárticos catalíticos (Asp 25, Gly 27, Ala 28, Asp 29 y Asp30), los pertenecientes a la región denominada “flap”, de elevada flexibilidad en la proteína libre y que se cierra sobre el ligando tras su unión a la enzima (Met 46, Ile 47, Gly 48, Gly 49 e Ile 50), otros situados en la hebra de una de las lámina beta comprendida entre los restos 80 y 86 (Pro 81, Val 82 e Ile 84), y el resto de Arg en posición 8, que contribuye significativamente a la energética de unión. En todos los casos, la región correspondiente a los restos aspárticos catalíticos aporta las mayores contribuciones a la energética de unión. Debido a la estructura parcialmente simétrica de los inhibidores, ambas cadenas de la proteasa contribuyen, asimismo, de forma más o menos simétrica a la energía libre de Gibbs correspondiente al proceso de unión.

*d.-Análisis de la estabilidad estructural de los restos de la proteasa que contribuyen más notablemente a la afinidad de unión.*

**Tabla IV.B.3: Localización de los restos de la proteasa del virus HIV-1 que contribuyen a la energía libre de unión de los distintos inhibidores.**

Resto	$\Delta\Delta G(\text{cal/mol})$												
	A77003	A78791	A76928	A74704	A76889	VX478	SB203386	SB203238	SB206343	U100313	U89360	A98881	CGP53820
Arg-8-A	-351	-358	-397	-602	-385	-132	-311	-468	-228	-106	-140	-363	-425
Asp-25-A	-344	-351	-342	-146	-337	-283	-320	-283	-248	-313	-346	-290	-312
Gly-27-A	-778	-747	-579	-629	-596	-538	-810	-436	-775	-437	-703	-542	-789
Ala-28-A	-189	-178	-207	-192	-209	-146	-165	-207	-159	-181	-135	-195	-177
Asp-29-A	-435	-447	-407	-702	-413	-312	-197	-455	-422	-308	-570	-331	-336
Asp-30-A	-198	-248	-192	-184	-232	-110	-158	-137	-220	-382	-437	-187	-202
Met-46-A	0	0	0	-290	0	0	0	0	0	0	-17	0	0
Ile-47-A	-47	-39	-38	-146	-30	-84	-46	-78	-87	-107	-97	-77	-55
Gly-48-A	-965	-962	-906	-935	-933	-719	-592	-616	-1081	-1021	-1186	-821	-904
Gly-49-A	-351	-378	-265	-173	-241	-125	-155	-128	-140	-192	-282	-107	-253
Ile-50-A	-271	-294	-253	-185	-272	-239	-260	-167	-271	-210	-151	-235	-231
Phe-53-A	0	0	0	-91	0	0	0	0	-65	-69	-161	0	0
Pro-81-A	-121	-129	-152	-166	-160	-78	-131	-181	-123	-42	-97	-98	-148
Val-82-A	-351	-357	-347	-191	-359	-119	-200	-179	-165	-95	-211	-159	-120
Ile-84-A	-86	-64	-82	-67	-83	-99	-60	-102	-90	-71	-95	-73	-51
Arg-8-B	-505	-513	-376	-230	-462	-245	-214	-283	-764	-624	-418	-352	-486
Asp-25-B	-323	-345	-369	-315	-366	-273	-225	-275	-334	-384	-372	-337	-367
Gly-27-B	-648	-652	-590	-501	-561	-594	-722	-444	-832	-353	-755	-451	-729
Ala-28-B	-184	-189	-196	-237	-195	-146	-203	-243	-183	-198	-167	-180	-183
Asp-29-B	-473	-478	-377	-432	-404	-275	-360	-523	-589	-420	-24	-218	-294
Asp-30-B	-167	-164	-156	-88	-149	-273	-326	-427	-313	-267	0	-270	-188
Ile-47-B	-56	-53	-60	-75	-59	-73	-68	-110	-60	-91	-32	-58	-66
Gly-48-B	-968	-954	-906	-1413	-931	-507	-966	-1089	-846	-701	-433	-591	-958
Gly-49-B	-156	-189	-191	-201	-185	-150	-166	-216	-192	-170	-174	-144	-150
Ile-50-B	-319	-256	-263	-125	-261	-215	-193	-204	-362	-215	-216	-212	-305
Pro-81-B	-247	-233	-154	-133	-179	-153	-126	-76	-219	-362	-348	-122	-160
Val-82-B	-398	-388	-352	-151	-393	-142	-179	-134	-156	-215	-121	-135	-94

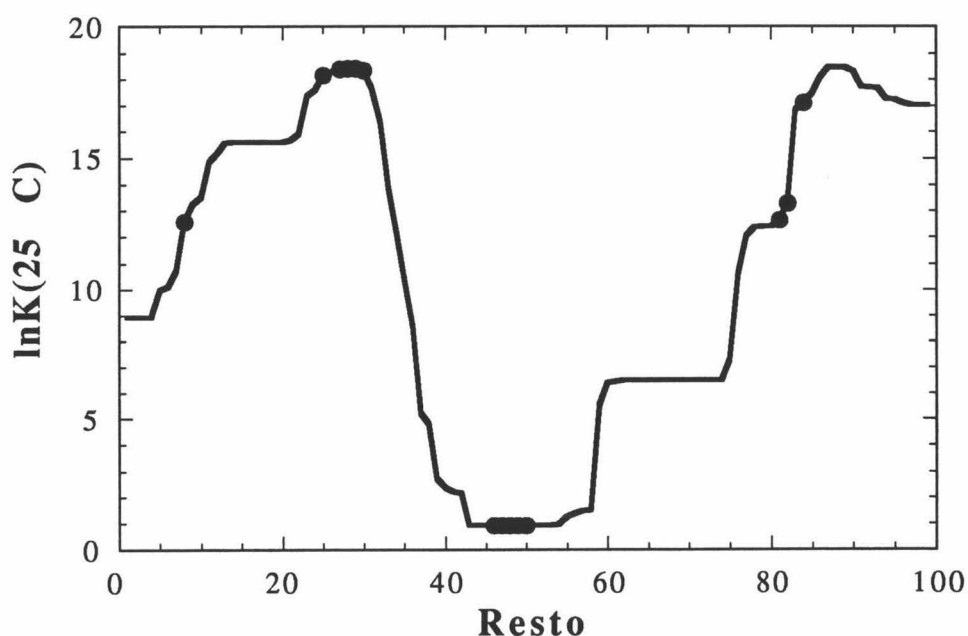
*Figura IV.B.11: En el panel superior se muestran las distintas regiones de la proteasa del virus V.I.H.-1 coloreadas de acuerdo a las constantes de estabilidad de cada aminoácido, de modo que las coloreadas en rojo están caracterizadas por una mayor estabilidad mientras que las azules son las más flexibles. En el panel inferior se representa la localización en la estructura tridimensional de los restos de la proteasa que contribuyen con más de 1 kcal/mol a la afinidad de unión de al menos uno de los inhibidores no peptídicos. Los restos que contribuyen entre -0.7 y -0.9 kcal/mol se muestran en color rojo, los que aportan entre -0.5 y -0.7 kcal/mol en naranja, los que contribuyen entre -0.3 y -0.5 en amarillo y, por último, los que aportan entre -0.1 y -0.3 en verde.*



A ribbon diagram of a protein structure, likely a beta-barrel, shown against a black background. The protein is represented by a white ribbon structure. The central cavity is filled with a complex arrangement of atoms, represented by small spheres in various colors (green, yellow, orange, red, blue). The protein structure is highly detailed, showing the intricate folding of the beta-strands.

Como se ha discutido anteriormente en la sección IV.B.1.2.3, las constantes de estabilidad aparentes por resto aminoácido permiten clasificar las distintas regiones de la proteína de acuerdo con su estabilidad estructural. De este modo, los restos con alta probabilidad de adoptar conformaciones nativas estarán caracterizados por valores elevados de sus constantes de estabilidad, mientras que aquéllos con mayor preferencia por los estados no estructurados presentarán constantes de estabilidad pequeñas (Hilser & Freire, 1996a; Hilser & Freire, 1996b; Hilser et al., 1997).

En la figura IV.B.12 se muestran los valores de las constantes de estabilidad calculadas mediante la aplicación del algoritmo COREX (Hilser & Freire, 1996a; Hilser & Freire, 1996b; Hilser et al., 1997) y de la parametrización estructural de la energética para la proteasa del virus V.I.H.-1. Estos mismos valores se muestran sobre la estructura tridimensional de la enzima en el panel superior de la figura IV.B.11 donde las regiones de la proteasa están coloreadas de acuerdo con las constantes de estabilidad de cada aminoácido, de modo que las representadas en rojo corresponden a las zonas más estables de la proteína y las azules a las caracterizadas por una mayor flexibilidad, siguiendo el orden rojo, naranja, amarillo, verde y azul.



*Figura IV.B.12: Constantes de estabilidad individuales para los distintos restos de la proteasa del virus V.I.H.-1 calculadas con el algoritmo COREX. Los puntos representan los restos que contribuyen más significativamente a la energía libre de Gibbs de unión.*

Como se puede observar en ambas figuras, son dos las regiones para las que se predice la mayor estabilidad estructural: la región que comprende los restos entre las posiciones 13 y 32, por un lado, y la correspondiente a los aminoácidos entre las posiciones 82 y 92, por otro. Aunque ambos segmentos se encuentran muy distantes en la secuencia, en la estructura tridimensional ocupan posiciones cercanas entre sí, definiendo una porción significativa del corazón hidrofóbico de la molécula. Asimismo se predice que una parte de estas regiones (en concreto los restos 23 a 29 en el extremo amino-terminal y los restos 86-99 en el carboxilo-terminal), junto con los nueve primeros restos de la secuencia, contribuyen significativamente a la interfase de dimerización. Según las predicciones, por tanto, estas regiones de la proteína se encuentran plegadas y correctamente estructuradas en la mayoría de los estados conformacionales accesibles a la proteasa bajo condiciones nativas. La tríada correspondiente al centro activo de la enzima, constituida por los restos Asp 25, Thr 26 y Gly 27 de ambas cadenas, se encuentra localizada en la zona más estable de la molécula como se muestra en la figura IV.B.12. Esta conclusión concuerda con los resultados obtenidos del análisis cristalográfico de la estructura, según los cuales dicha región se caracteriza por su elevada rigidez, consecuencia de la existencia de una densa red de enlaces de hidrógeno entre los restos que la constituyen (Wlodawer & Erickson, 1993). El segmento correspondiente a los restos 82-92 comprende la única y bien definida hélice de la estructura, caracterizada asimismo por una elevada estabilidad.

Por el contrario, la región comprendida entre los restos 40 y 60, que corresponde a la zona flexible de la estructura denominada "flap", está caracterizada por constantes de estabilidad extremadamente bajas, por lo que, según las predicciones, debe encontrarse desestructurada en condiciones nativas. Sin embargo, en los complejos esta región se estabiliza notablemente como consecuencia de las interacciones establecidas con los inhibidores. Los resultados obtenidos al realizar los cálculos con la estructura cristalográfica de la proteína libre (1hhp) o con las estructuras de la proteína en los complejos en ausencia de los inhibidores son muy similares, lo que confirma la idea de que, en los complejos, la región correspondiente al flap se encuentra estabilizada por las interacciones con los inhibidores y sustratos, y no por posibles contactos con otras regiones de la proteína.

En la figura IV.B.12 se han representado asimismo los restos que contribuyen más significativamente a la energía libre de Gibbs para el proceso de unión, mostrados asimismo sobre la estructura tridimensional en el panel superior de la figura IV.B.11. Resulta evidente, tras el análisis de los datos, que el sitio de unión está constituido por restos pertenecientes a las regiones más y menos estables de la molécula. Esta doble naturaleza determina una de las características fundamentales de la unión de los distintos ligandos a la proteasa, de modo que, esencialmente, una parte del sitio de unión se encuentra preformado con anterioridad a la unión del ligando, mientras que otra se constituye posteriormente, como consecuencia de la presencia de éste.

Así, la región más estable (que contiene los restos Asp 25, Gly 27, Ala 28, Asp 29, Asp 30, Pro81, Val82 e Ile 84) se encuentra fija en una conformación competente para la unión con los



ligandos con anterioridad a su unión. Por el contrario, la región del “flap”, constituida por los restos Met46, Ile 47, Gly 48, Gly 49 e Ile 50, está en su mayoría desestructurada en la proteína libre, y se ve forzada a adoptar una única conformación rígida como consecuencia de las interacciones establecidas con los sustratos e inhibidores en el proceso de unión. Este es el motivo por el que la sustitución de restos de la proteasa, que no se encuentran en contacto directo con el inhibidor, pero que pueden afectar, tanto estructural, como energéticamente a la capacidad del “flap” para adoptar la conformación correcta para la unión de los ligandos, pueden influir notablemente en la energética global del proceso de unión.

*e.-Bases moleculares de la resistencia de la proteasa frente a distintos fármacos.*

El análisis de la energética de unión de los distintos inhibidores, descrita en los apartados anteriores, proporciona algunos indicios sobre la posible naturaleza de los efectos producidos por las mutaciones de la proteasa del virus V.I.H.-1 observadas *in vivo*, y responsables del desarrollo de resistencia a múltiples inhibidores.

En algunos casos se han determinado experimentalmente las constantes de inhibición de los distintos fármacos con algunos de los mutantes resistentes (Kaplan et al., 1994) y se dispone, asimismo, de la estructura cristalográfica de algunos complejos (1GNM, (Hong et al., 1996); 1HPO, (Skulnick et al., 1997); 1HVS, (Baldwin et al., 1995b); 1ODX, (Kervinen et al., 1996); 1TCX, (Hoog et al., 1996)) Este es el caso del mutante resistente V82A, cuya estructura cristalográfica con el complejo A-77003 se encuentra disponible en el banco de datos estructurales de proteínas bajo el código 1HVS (Baldwin et al., 1995b).

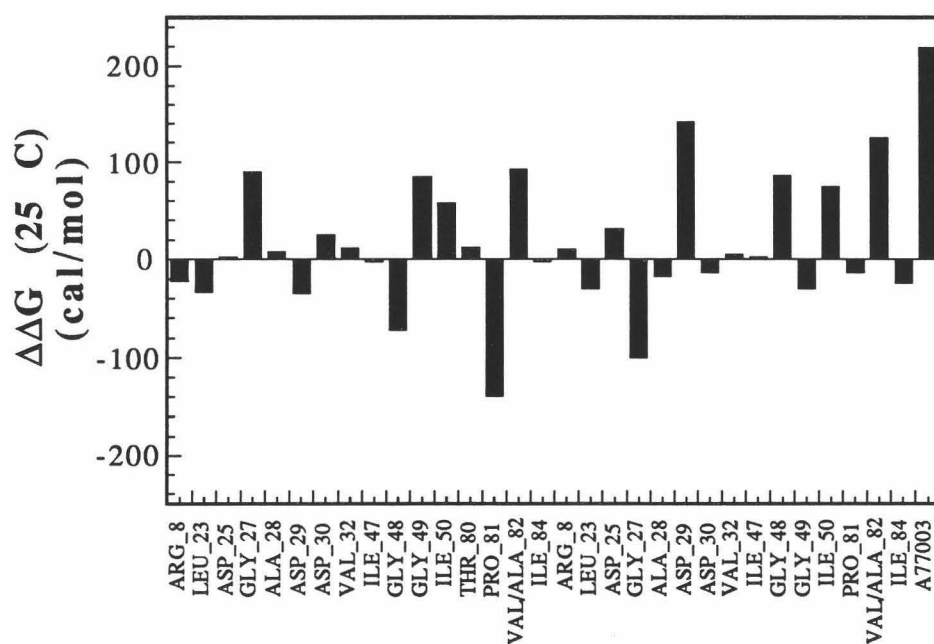
Con objeto de analizar la energética de unión del inhibidor a la proteína mutante e intentar elucidar los efectos que dicha mutación produce en las distintas magnitudes termodinámicas y, en particular, en la afinidad de unión, se llevó a cabo un proceso similar al descrito en los apartados anteriores, haciendo uso de esta estructura (Baldwin et al., 1995b). La energía libre de unión obtenida ( $-12.4 \text{ kcal}\cdot\text{mol}^{-1}$ ) representa una reducción de 3.7 veces en la afinidad de unión con respecto a la proteína silvestre (caracterizada por una energía libre de unión de  $-13.1 \text{ kcal}\cdot\text{mol}^{-1}$ ). Estos valores son similares a los obtenidos experimentalmente, y que revelan una afinidad cuatro veces menor del inhibidor por la proteína mutante (Baldwin et al., 1995b).

Aunque la reducción observada en la constante de inhibición es importante, la información de mayor interés reside en el mecanismo mediante el cual una simple mutación afecta tan significativamente a la afinidad de unión. Para ello se ha recurrido a la comparación de las contribuciones por resto correspondientes a la proteína silvestre y al mutante, mostrados en la figura IV.B.13, donde se representan las diferencias de  $\Delta G$  entre los restos que contribuyen con más de  $0.1 \text{ kcal}\cdot\text{mol}^{-1}$  a la energía libre de unión.

Los resultados indican que el efecto de la mutación no puede asignarse a una única posición, sino que las diferencias en energía libre de Gibbs se encuentran redistribuidas a lo largo



de todo el sitio de unión. Esta conclusión concuerda con las obtenidas por Baldwin y colaboradores (Baldwin et al., 1995b) a partir del análisis cristalográfico de la estructura del complejo con el mutante, según el cual, el efecto global de la mutación V82A no puede racionalizarse mediante la eliminación de un simple grupo metilo en cada una de las cadenas, sino que se debe a la reorganización de los grupos del esqueleto peptídico inducida por la presencia del nuevo aminoácido. Los resultados mostrados en la figura IV.B.13 son, por tanto, un claro ejemplo de que, incluso para la sustitución de un simple grupo metilo, la energía libre de Gibbs del proceso de unión no puede racionalizarse mediante la simple adición de las contribuciones de cada grupo, sino que éstas dependen del contexto global en que se encuentra cada uno de ellos. Asimismo, queda claro que la interpretación de los efectos que las diversas mutaciones tienen en la afinidad de unión de un inhibidor determinado requiere el desarrollo de



**Figura IV.B.13:** Diferencias observadas entre los restos de la proteasa que contribuyen más significativamente a la afinidad de unión ( $>0.1 \text{ kcal}\cdot\text{mol}^{-1}$ ). Los valores de  $\Delta\Delta G$  mostrados representan las contribuciones de cada resto en el complejo con la proteína mutante con respecto a la proteína silvestre.

un análisis global, incluso en los casos en que dicha mutación se encuentra localizada en el sitio de unión.

#### IV.B.3.2.3.-Inhibidores peptídicos: Acetil-pepstatina.

##### a.-Análisis de la energética del proceso de unión.

La estructura cristalográfica del complejo entre acetil-pepstatina y la proteasa del virus V.I.H.-1, resuelta a 2.0 Å de resolución, se encuentra disponible en el banco de datos de proteínas bajo el código 5HVP (Fitzgerald et al., 1990). Dado que, en dicha estructura, el inhibidor se observa en dos orientaciones diferentes, los cálculos termodinámico-estructurales se han aplicado a ambas conformaciones por separado. En todos los casos se ha utilizado la estructura correspondiente a la enzima libre en disolución, recogida en el fichero PDB de código 1HHP (Spinelli et al., 1991).

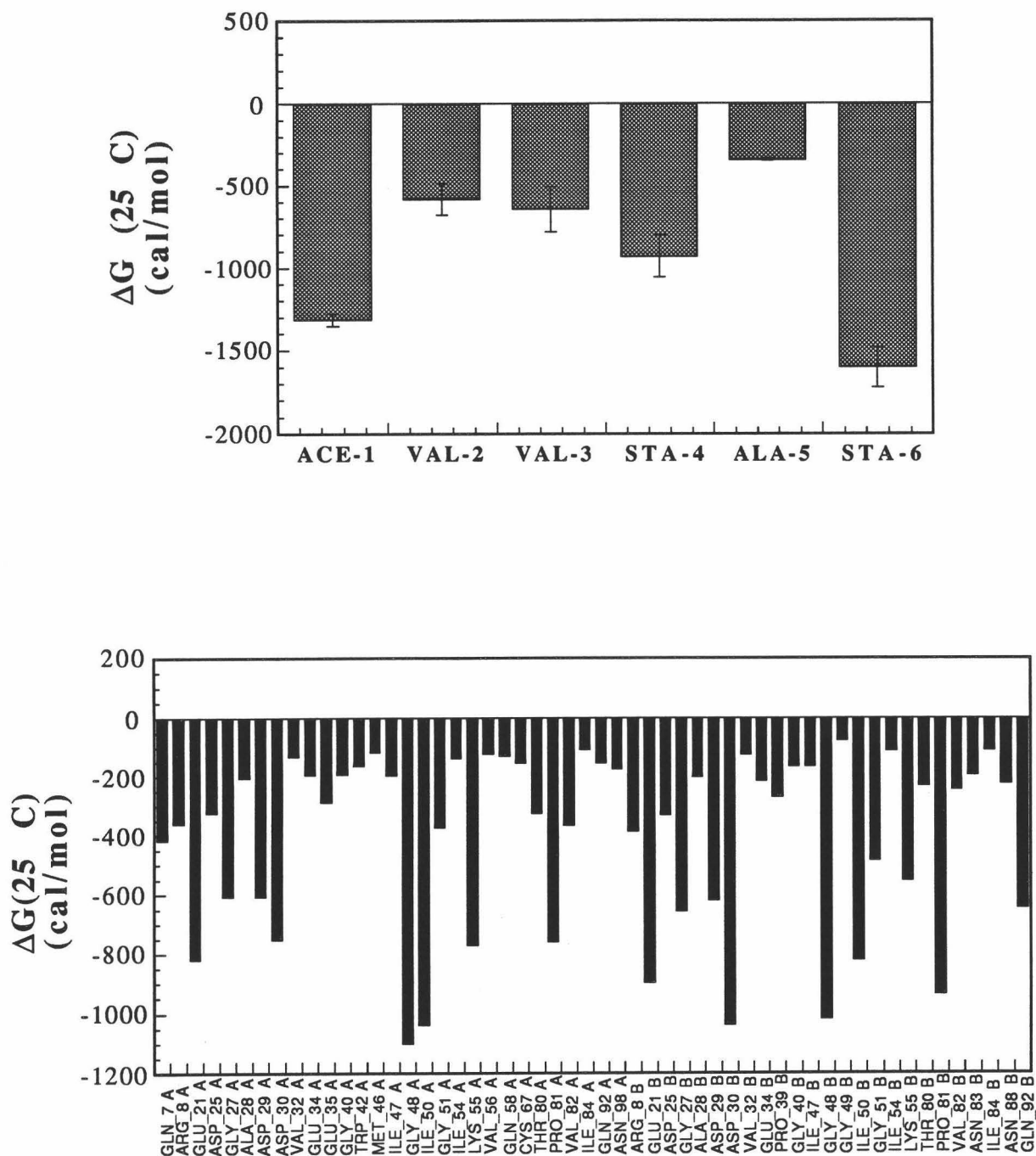
Como se ha discutido anteriormente (ver Sección II.B.2.3.), los grupos aspárticos catalíticos de la proteasa del virus V.I.H.-1 (al igual que ocurre con Endothiapepsina y otras proteasas aspárticas) presentan, aún a pesar de su disposición simétrica, diferentes tendencias a la ionización, que se reflejan en las diferencias entre sus pKa en el complejo (pKa1 < 2.5 y pKa2 > 6.5). En estos casos, en que la unión del ligando implica la protonación forzosa de uno de los grupos de la enzima y, como consecuencia, un intercambio de protones con el tampón, la entalpía aparente de unión obtenida experimentalmente ( $\Delta H_{\text{unión}}$ ) es el resultado de la combinación de una contribución intrínseca, característica del proceso de unión e independiente del tampón utilizado, y un término proporcional a la entalpía de ionización de dicho tampón. De este modo:

$$\Delta H_{\text{unión}} = \Delta H_{\text{intr}} + N_{\text{H}^+} \cdot \Delta H_{\text{ion}} \quad (\text{IV.B.6})$$

donde la constante de proporcionalidad,  $N_{\text{H}^+}$ , es el número de protones intercambiados con el medio y  $\Delta H_{\text{ion}}$  la entalpía de ionización del tampón.

Los cálculos se han realizado de acuerdo con la parametrización estructural ya descrita, considerando una temperatura de 25 °C y un pH de 5.5 y asumiendo el acoplamiento de la unión del ligando con la protonación de un grupo aspártico caracterizado por un pKa de 3.5 (Hyland et al., 1991a; Hyland et al., 1991b). En la tabla IV.B.4 se muestran los resultados obtenidos del análisis termodinámico estructural de la unión de este inhibidor a la proteasa para cada una de las dos conformaciones encontradas en el complejo. Asimismo, se muestran los resultados, obtenidos mediante calorimetría isotérmica de titulación bajo condiciones similares (25 °C, pH 5.5, 20mN NaOAc, 0.2 mM EDTA), procedentes del estudio experimental llevado a cabo paralelamente en el laboratorio del Dr. Freire.

Como se puede observar, los resultados obtenidos de los cálculos reproducen razonablemente bien los valores experimentales de las distintas magnitudes termodinámicas, indicando que la parametrización estructural de la energética captura de forma efectiva la magnitud de las fuerzas que determinan la afinidad de unión y el delicado balance establecido entre ellas para este sistema. De acuerdo con estos resultados, la fuerza dominante en el proceso de unión es el aumento de entropía originado por la deshidratación de las superficies, de naturaleza predominantemente hidrofóbica, como consecuencia de la asociación entre la proteína



**Figura IV.B.14:** Contribuciones a la energía de Gibbs genérica del proceso de unión de los restos de Acetil-pepstatina (panel superior) y de la proteasa (panel inferior). Únicamente se han representado los restos de la enzima que contribuyen con más de  $0.1 \text{ kcal}\cdot\text{mol}^{-1}$  a la afinidad de unión.

y el inhibidor. A esta interacción favorable se oponen el cambio negativo en entropía conformacional, correspondiente tanto al inhibidor peptídico como a las cadenas laterales de la proteasa localizadas en el sitio de unión, la pérdida de entropía translacional, y el cambio entálpico positivo.

*Tabla IV.B.4: Análisis termodinámico-estructural de la unión de Acetil-pepstatina a la proteasa del virus V.I.H.-1*

		Acetil pepstatina (Conf 1)	Acetil pepstatina (Conf 2)	Experimental
$\Delta C_p$	cal·(K·mol) <sup>-1</sup>	-435	-419	-452
$\Delta H$	kcal·mol <sup>-1</sup>	11.6	8.5	7.0
$\Delta S_{conf}$	cal·(K·mol) <sup>-1</sup>	-35.9	-38.0	-----
$\Delta S_{solv}$	cal·(K·mol) <sup>-1</sup>	124.9	122.7	-----
$\Delta S_{otros}$	cal·(K·mol) <sup>-1</sup>	-25.2	-25.0	-----
$\Delta S_{total}$	cal·(K·mol) <sup>-1</sup>	63.8	59.7	50.0
$\Delta G$	kcal·mol <sup>-1</sup>	-7.4	-9.3	-8.0

*b.-Identificación de los restos que aportan las principales contribuciones a la energética de unión.*

En el panel superior de la figura IV.B.14 se representan las contribuciones genéricas de cada uno de los restos del inhibidor Acetil-pepstatina a la energía libre de unión a la proteasa, donde las barras representan el promedio entre las dos conformaciones presentes en la estructura cristalográfica y las barras de error la correspondiente desviación estándar. Como puede observarse, al igual que ocurre con el complejo entre pepstatina A y Endothiapepsina (Gómez & Freire, 1995), todos los restos del inhibidor contribuyen de forma favorable y significativa a la energía libre de unión, lo cual es indicativo de la fortaleza de las interacciones establecidas con la enzima. Es de destacar, particularmente, la contribución de los restos de estatina en las posiciones central y C-terminal, que contribuyen con más de -1 kcal·mol<sup>-1</sup> (-9.3 y -1.6 kcal·mol<sup>-1</sup>, respectivamente) a la energía libre de unión, en concordancia con el papel esencial que se les ha atribuido para la unión de este inhibidor a las distintas proteasas aspárticas.

En el panel inferior se muestran las contribuciones individuales a la energía libre de Gibbs de unión correspondientes a los restos de la enzima. Como se puede observar, la situación es muy similar a la descrita en secciones anteriores (sección IV.B.2.2.1) para la unión de los distintos inhibidores no peptídicos. De este modo, desde un punto de vista energético, es posible identificar cuatro regiones alejadas en la secuencia que contribuyen notablemente a la energética de unión: la región correspondiente a las proximidades de los aspárticos catalíticos (Asp 25, Gly 27, Ala 28,

Asp 29 y Asp 30), la región flexible correspondiente al “flap” (Met 46, Ile 47, Gly 48, Gly 49 e Ile 50), la hebra beta que comprende los restos 80-86 (Pro 81, Val82 e Ile 84) y por último, el resto de Arginina en posición 8. La correspondencia entre estas regiones y las descritas con anterioridad (sección IV.B.2.2.1) para la unión de inhibidores no peptídicos es casi total, de lo cual se puede deducir que el mecanismo de estabilización de los complejos entre la enzima y ambos tipos de inhibidores es muy similar.

#### **IV.B.3.-Modelado de ligandos peptídicos para proteasas aspárticas.**

##### *IV.B.3.3.1.-Diseño de inhibidores peptídicos para Endothiapepsina.*

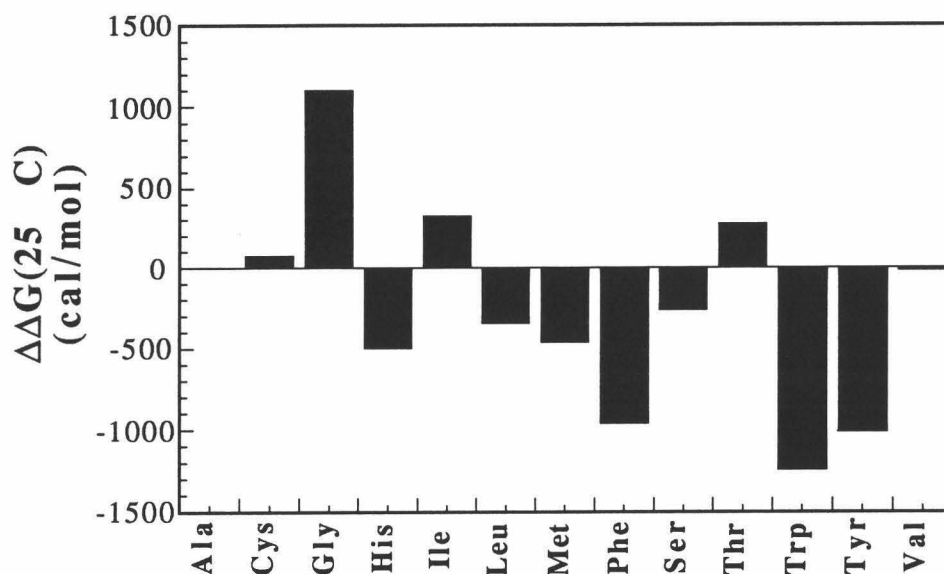
###### *a.-Análisis mutacional en la posición 5 de Pepstatina A.*

De acuerdo con el razonamiento expuesto anteriormente (sección IV.B.1.3.1), se procedió a la sustitución en posición 5 de doce aminoácidos (Cys, Gly, His, Ile, Leu, Met, Phe, Ser, Thr, Trp, Tyr y Val) con cadenas laterales de diversos tamaños y grados de hidrofobicidad. En cada uno de los casos, una vez generada la mutación, se procedió a la optimización energética de la conformación adoptada por la nueva cadena lateral en el complejo. Como ejemplo, en la figura IV.B.15 se muestra el perfil energético para el caso del aminoácido Phe, en el que se puede observar, en comparación con el perfil obtenido para el mismo aminoácido en una hélice expuesta al disolvente (Figura IV.B.9), una fuerte restricción del espacio conformacional accesible a la cadena lateral en el complejo.

Una vez identificadas las conformaciones correspondientes a los mínimos energéticos para cada uno de los mutantes, se procedió a la estimación de su afinidad de unión mediante la aplicación de la parametrización estructural de la energética descrita en la sección II.2. En la figura IV.B.16 se muestran las diferencias en afinidad obtenidas para los nuevos péptidos con los diferentes aminoácidos substituidos en posición 5, con respecto al complejo original.

Los datos reflejados en la gráfica muestran que, según los cálculos realizados, la substitución de Ala por cualquiera de los aminoácidos aromáticos (Phe, Trp y Tyr) conduce a los mayores incrementos en la afinidad de unión. Las constantes de afinidad predichas para los péptidos que contienen estos tres aminoácidos se encuentran comprendidas en un intervalo de unas  $0.2 \text{ kcal}\cdot\text{mol}^{-1}$ . Dado que se encuentran dentro del intervalo de incertidumbre con que se realizan las predicciones, las diferencias observadas entre las tres cadenas laterales aromáticas no pueden considerarse estadísticamente significativas.

Con objeto de comprobar los resultados de las predicciones, se procedió a la síntesis del derivado A5F de Pepstatina y a la determinación experimental de su afinidad de unión a Endothiapepsina, llevada a cabo paralelamente en el laboratorio del Dr. Freire (Luque et al., 1997). Los resultados del análisis experimental se recogen en la tabla IV.B.5, donde se puede observar que, de acuerdo con las predicciones, el péptido modificado se une a la enzima con



**Figura IV.B.16:** Diferencias de afinidad entre los distintos mutantes en posición 5 de Pepstatina A. Se representan las variaciones observadas en energía libre con respecto al inhibidor original (Ala 5).

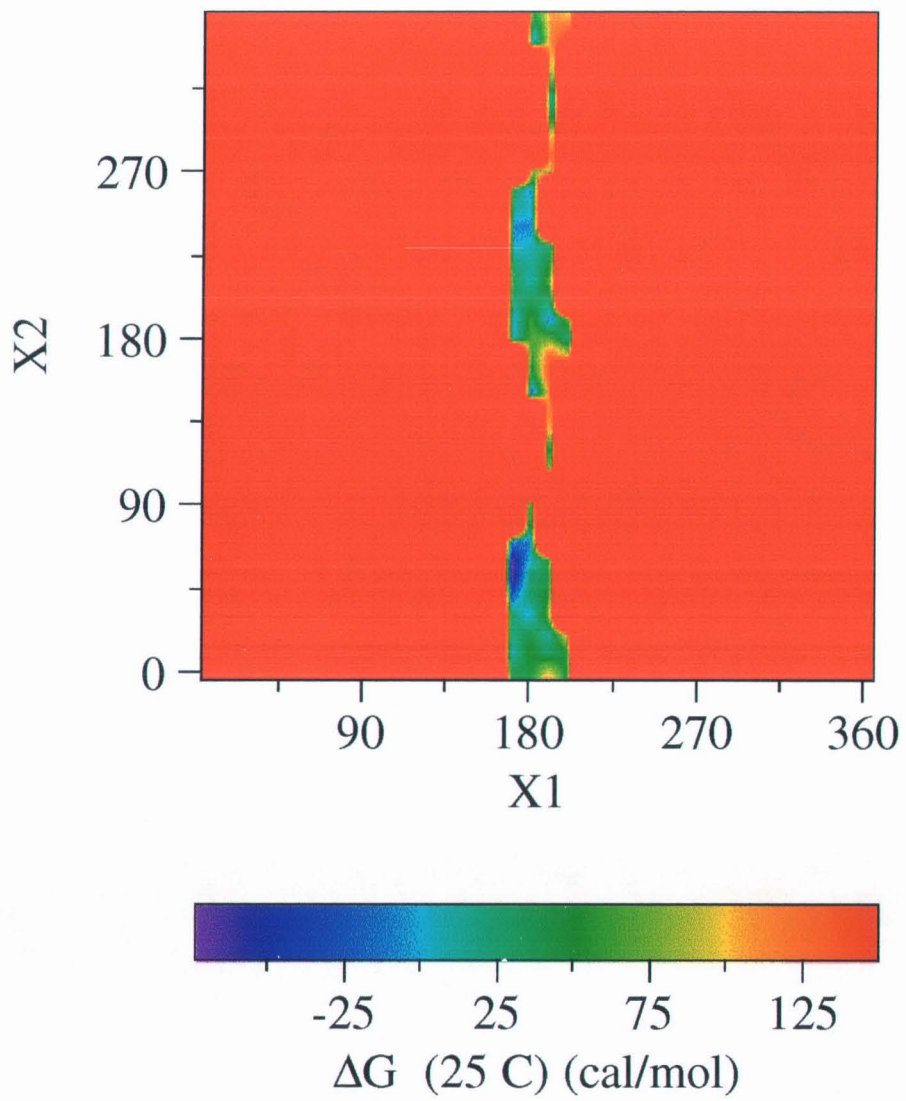
mayor fortaleza que el inhibidor original. La constante de afinidad predicha aplicando la parametrización estructural de la energética es  $7.4 \times 10^9 \text{ M}^{-1}$  es muy similar a la determinada experimentalmente, ( $5.3 \times 10^9 \text{ M}^{-1}$ ), de modo que la diferencia entre ambas ( $0.2 \text{ kcal}\cdot\text{mol}^{-1}$ ) se encuentra dentro del intervalo de error experimental. Desgraciadamente, el mayor carácter hidrofóbico del mutante A5F disminuye notablemente su solubilidad, por lo que, en este caso, no ha sido posible llevar a cabo la medida directa de la entalpía y capacidad calorífica del proceso de unión mediante técnicas calorimétricas.

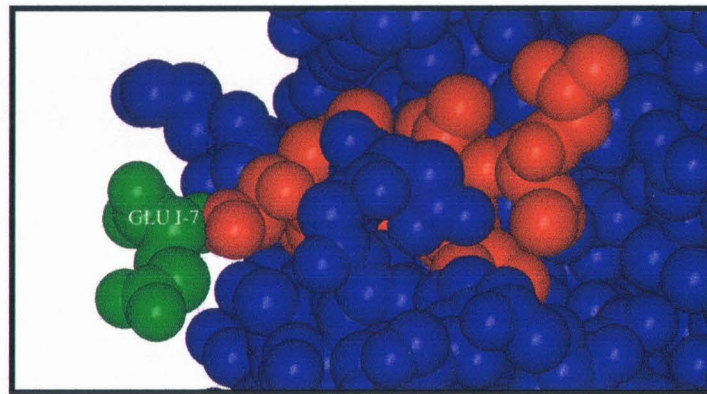
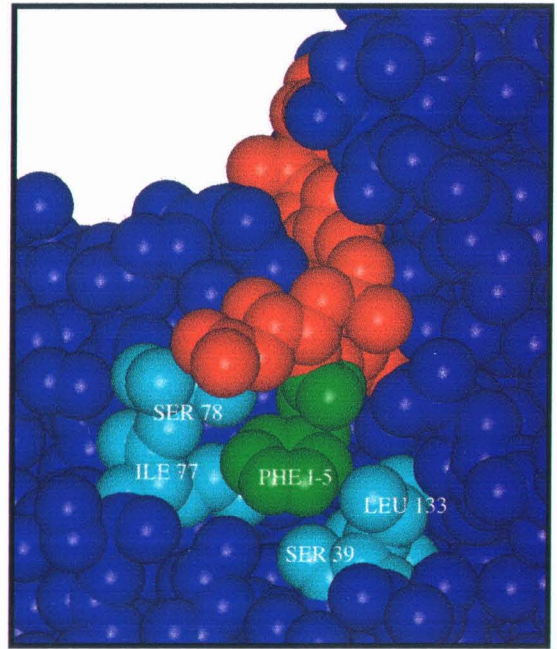
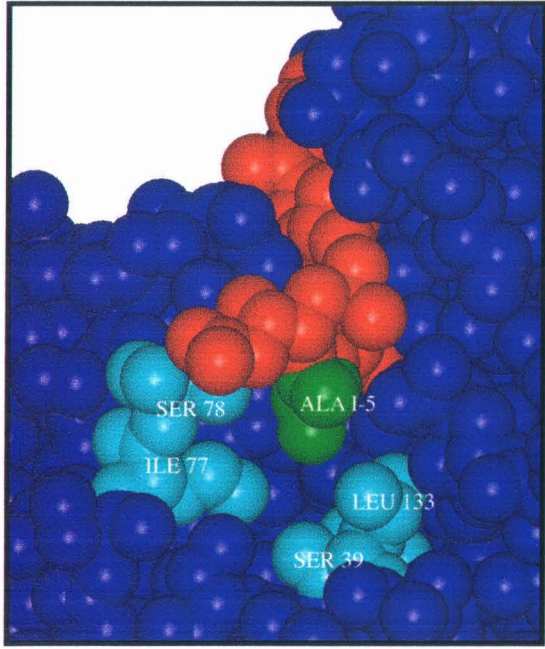
**Tabla IV.B.5:** Afinidades de unión de distintos mutantes de Pepstatina A a Endothiapepsina

Inhibidor	$\Delta G(25 \text{ }^\circ\text{C})_{\text{cal}}$ (kcal/mol)	$\Delta G(25 \text{ }^\circ\text{C})_{\text{exp}}$ (kcal/mol)	$K_{\text{b, calc}}$ ( $\text{M}^{-1}$ )	$K_{\text{b, exp}}$ ( $\text{M}^{-1}$ )
Pepstatina A:				
Iva-Val-Val-Sta-Ala-Sta	-12.5	-12.7	$1.5 \times 10^9$	$2.3 \times 10^9$
Iva-Val-Val-Sta-Phe-Sta	-13.5	-13.3	$7.4 \times 10^9$	$5.3 \times 10^9$
Iva-Val-Val-Sta-Phe-Sta-Glu	-11.8	-11.3	$4.5 \times 10^8$	$2.1 \times 10^8$



*Figura IV.B.15: Perfil de energía libre en función de la conformación adoptada por la cadena lateral del resto de Fenilalanina en la posición 5 del inhibidor Pepstatina A, en el complejo con Endothiapepsina. Los valores de energía libres se muestran en distintos colores de acuerdo con la escala mostrada bajo la figura, mientras que las zonas representadas en color rojo corresponden a las regiones no permitidas del espacio conformacional.*





*b.-Adición de un resto Glutámico en el extremo C-terminal.*

Con objeto de mejorar la solubilidad de péptido y facilitar el análisis calorimétrico, que permitiría la comparación entre las magnitudes termodinámicas predichas por la metodología desarrollada en este trabajo y los resultados experimentales, se decidió añadir un resto glutámico en el extremo carboxilo terminal del péptido, aún a pesar de la predicción inicial de una disminución en la afinidad de unión.

El análisis calorimétrico llevado a cabo en el laboratorio del Dr. Freire con este nuevo péptido (Luque et al., 1997) indica que el proceso de unión del nuevo inhibidor a la enzima es de naturaleza exotérmica a 16 °C, y se caracteriza por un cambio de entalpía ( $\Delta H$ ) de  $-4.6 \pm 0.1$  kcal·mol<sup>-1</sup>. El cambio de capacidad calorífica, obtenido de la dependencia con la temperatura de los cambios en entalpía, es igual a  $-260 \pm 20$  cal·(K·mol)<sup>-1</sup>. El valor predicho por el análisis termodinámico estructural es  $-220$  cal·(K·mol)<sup>-1</sup> para el cambio en capacidad calorífica. Para el cambio en entalpía genérica, que corresponde exclusivamente las contribuciones estructurales y de hidratación y que no contiene los efectos de protonación, el valor procedente de los cálculos es  $-6.8$  kcal·mol<sup>-1</sup>. Estos valores son del mismo orden que los determinados bajo las mismas condiciones para pepstatina A ( $\Delta H = -4.1$  kcal·mol<sup>-1</sup> y  $\Delta C_p = -310$  cal·(K·mol)<sup>-1</sup>, indicando que el origen de la diferencia entre las afinidades de unión entre el péptido original y el mutante E7 es de naturaleza fundamentalmente entrópica.

Los valores de las afinidades de unión predichos por la parametrización estructural de la energética, así como los determinados experimentalmente para pepstatina A y ambos mutantes. se recogen en la tabla IV.B.5. De acuerdo con las predicciones, el mutante A5F presenta una mayor afinidad que la pepstatina original, mientras que la adición del resto glutámico en posición 7 (mutante E7) conlleva la disminución de la afinidad. La correspondencia entre los valores de  $\Delta G$  predichos y experimentales es muy buena, siendo la diferencia promedio entre ambos de  $0.23 \pm 0.06$  kcal·mol<sup>-1</sup>, lo cual indica que la parametrización estructural de la energética posee la sensibilidad y resolución necesarias para su aplicación al diseño de ligandos peptídicos.

*c.-Localización estructural de las contribuciones a las afinidades de unión.*

*c1.-Pepstatina A.*

*Figura IV.B.17: Localización en la estructura tridimensional del complejo entre Pepstatina A y Endothiapepsina de los restos del inhibidor en posición 5, donde se han llevado a cabo las mutaciones. La estructura correspondientes al inhibidor original (Ala5) se muestra en el panel superior izquierdo, el mutante A5F en el superior derecho y el E7 en el inferior. Los átomos de la proteína se muestran en color azul oscuro, de modo que los restos de color más claro son los que se ven más afectados por la introducción del resto de Fenilalanina. El inhibidor se muestra en color rojo y la posición de mutación en verde.*

Como puede observarse en el panel superior izquierdo de la figura IV.B.17, en el complejo entre la enzima y pepstatina A, el resto de alanina en posición 5 del inhibidor se encuentra localizado en una cavidad hidrofóbica, estableciendo escasos contactos de van der Waals con la enzima, por lo que su presencia no conlleva la protección del contacto con el disolvente de una porción significativa de la superficie molecular en esta región.

En el panel superior de la figura IV.B.18 se representan las contribuciones de los restos de la enzima y del inhibidor a la energía libre de Gibbs genérica correspondiente al proceso de unión entre pepstatina A y Endothiapepsina. Con objeto de simplificar la representación, se muestran únicamente aquellos aminoácidos que contribuyen con más de  $0.1 \text{ kcal}\cdot\text{mol}^{-1}$  a la afinidad de unión. Se han incluido asimismo, los restos que, como se describirá a continuación, se ven más afectados por las sustituciones llevadas a cabo en la posición 5 del ligando (Leu 133, Ser78, Ile 77, Ser39).

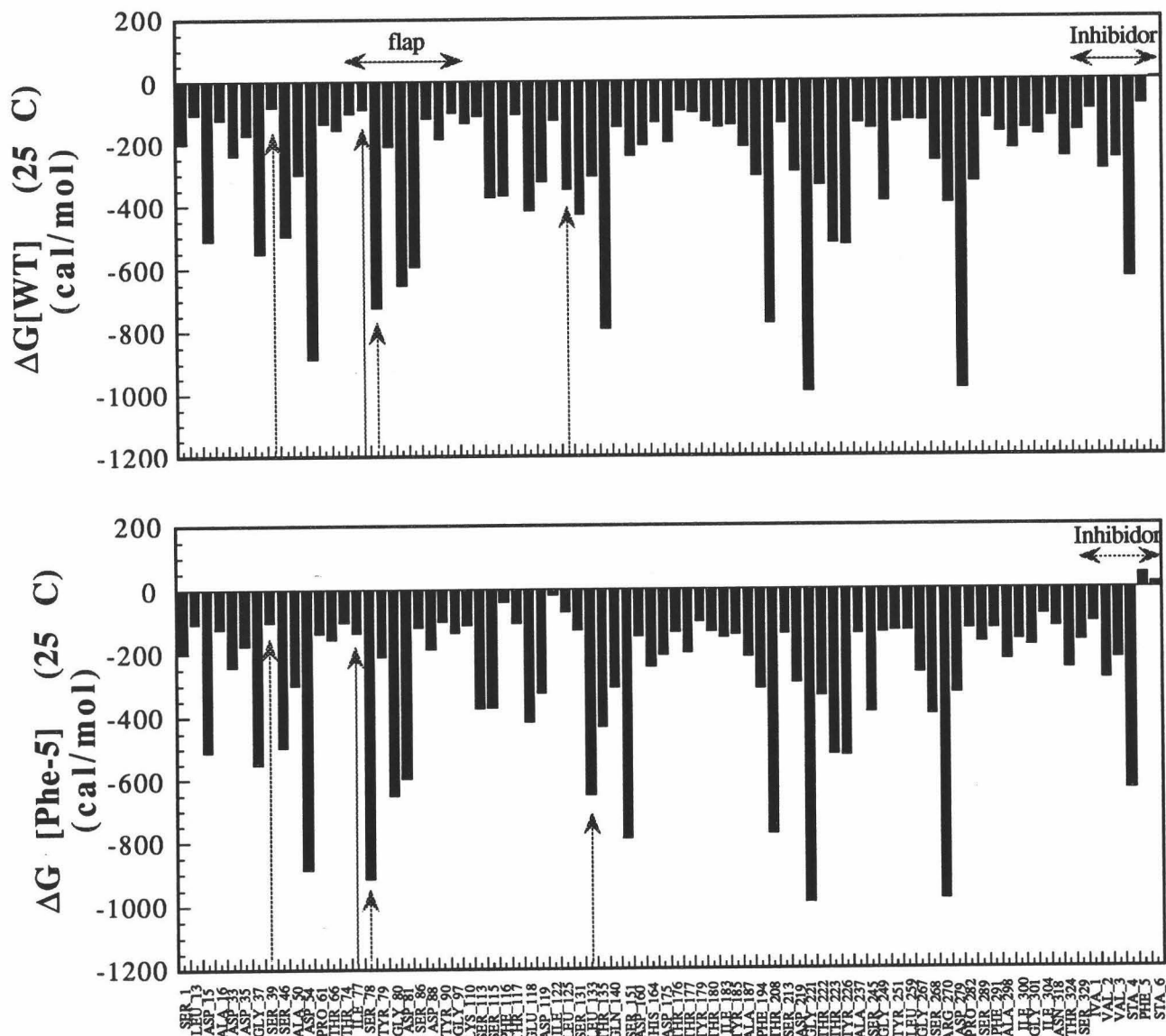
### *c.2.-Mutante A5F.*

De acuerdo con las predicciones, los péptidos mutantes A5F, A5Y, y A5W presentan un incremento en la afinidad por la enzima con respecto al inhibidor original debido a que el anillo aromático de estos aminoácidos se ajusta parcialmente a la cavidad hidrofóbica de la enzima, optimizando las interacciones con los aminoácidos Leu133, Ser78, Ile77 y Ser39, como se muestra en el panel superior derecho de la figura IV.B.17 para el caso concreto del mutante A5F. Como se puede observar, aunque el anillo aromático del resto de fenilalanina en posición 5 se encuentra sólo parcialmente enterrado en el interior de la proteína, las interacciones que establece con los restos de la enzima Leu 133, Ser 78, Ile 77 y Ser39, así como la deshidratación que induce en estos aminoácidos, proporcionan la mayor parte de la estabilización adicional.

En el panel inferior de la figura IV.B.18 se muestran, de forma similar a la descrita para la pepstatina original, los aminoácidos que contribuyen a la energía libre de unión con más de  $0.1 \text{ kcal}\cdot\text{mol}^{-1}$ , así como los restos cuyas contribuciones se incrementan más notablemente como consecuencia de la sustitución llevada a cabo en el inhibidor. En el panel superior de la figura IV.B.19. se muestran las diferencias ( $\Delta\Delta G = \Delta G(\text{wt}) - \Delta G(\text{mutante})$ ) entre las contribuciones de los restos individuales en el mutante A5F con respecto al complejo original.

La energía libre de Gibbs de unión para el mutante A5F es cerca de  $1 \text{ kcal}\cdot\text{mol}^{-1}$  más favorable que la correspondiente al complejo original. La contribución adicional de la entropía





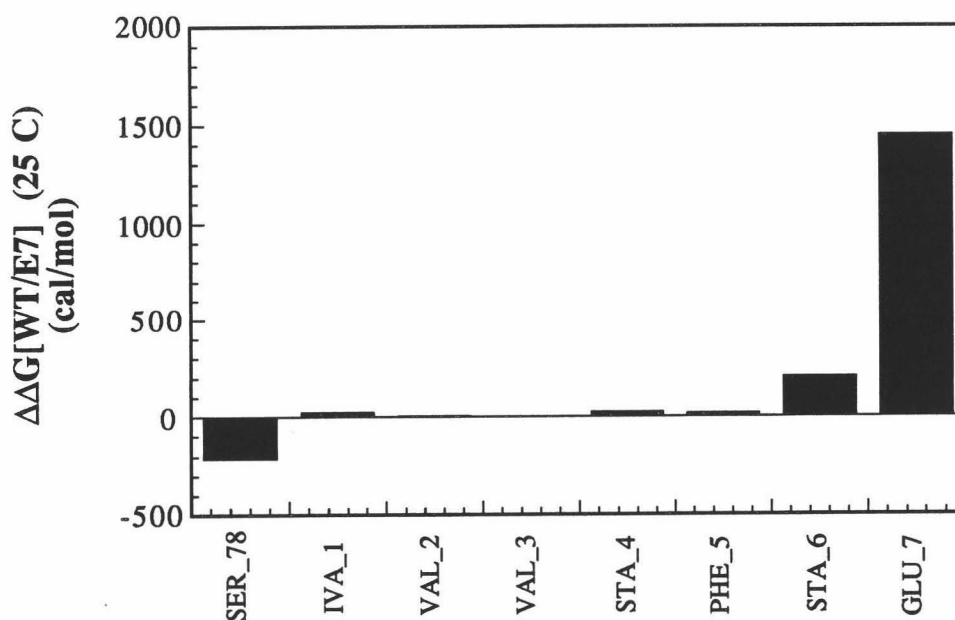
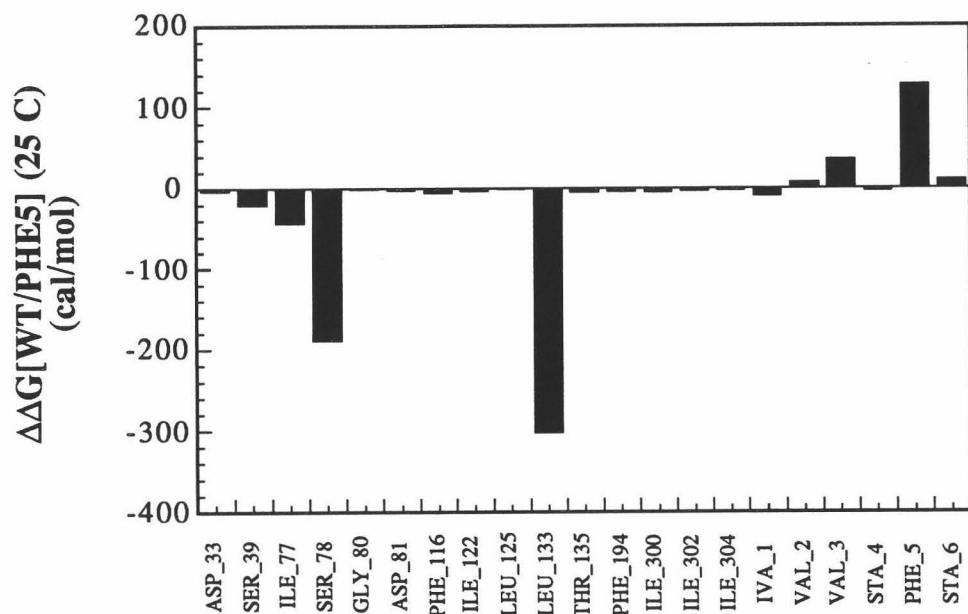
**Figura IV.B.18:** Contribuciones individuales a la energía libre de Gibbs genérica del proceso de unión de Pepstatina A (panel superior) y su derivado A5F (panel inferior) a Endothiapepsina. Se representan los restos de la enzima que contribuyen con más de  $0.1 \text{ kcal}\cdot\text{mol}^{-1}$  a la afinidad de unión. Las flechas señalan aquellos restos cuyas contribuciones se ven más afectadas por la introducción del resto de Fenilalanina en el inhibidor (Leu 133, Ser78, Ile 77, Ser39).

de deshidratación, procedente de la protección al disolvente de un mayor número de grupos hidrofóbicos, es cercana a  $4 \text{ kcal}\cdot\text{mol}^{-1}$ . Sin embargo, esta contribución entrópica se encuentra parcialmente compensada por un cambio en entalpía menos favorable en el nuevo complejo (aproximadamente  $2 \text{ kcal}\cdot\text{mol}^{-1}$  más positiva para A5F que para la pepstatina original) y una mayor pérdida de entropía conformacional (aproximadamente  $1 \text{ kcal}\cdot\text{mol}^{-1}$ ), a causa de las restricciones adicionales de los grados de libertad de la cadena lateral de Phe con respecto a Ala. El cambio entálpico para el mutante A5F es más desfavorable que para el inhibidor original debido a la mayor magnitud de la entalpía de deshidratación, que no se encuentra completamente compensada por las interacciones adicionales establecidas entre el péptido y la molécula de proteasa.

Como ocurre en todos los procesos de unión, como resultado de la existencia de diversas interacciones compensatorias (compensación entalpía/entropía, compensaciones entálpicas entre las interacciones intermoleculares y las correspondientes a la deshidratación, y compensación entrópica entre la entropía de solvatación y la conformacional), la magnitud del cambio global en energía libre es menor que el de las contribuciones independientes.

### c.3.-Mutante E7.

En el caso de la mutación E7, se ha procedido a la elongación del péptido en el extremo carboxilo terminal, de modo que, tras las etapas de minimización pertinentes, se observa que el resto glutámico adicional se encuentra dirigido hacia el exterior de la proteína, sin establecer interacciones significativas con ningún otro resto. Este hecho se refleja en la obtención de valores de entalpía similares a los observados con el péptido original, pepstatina A, por lo que puede afirmarse que la disminución en la energía libre de unión con respecto al inhibidor original tiene su origen fundamentalmente en la pérdida de entropía conformacional del resto glutámico como consecuencia de la unión del inhibidor a la enzima. Esta pérdida entrópica no se encuentra compensada por ningún otro tipo de interacción, entálpica o entrópica, por lo que conduce a un incremento significativo del cambio en energía libre y, consecuentemente, a la disminución de su afinidad de unión. En el panel inferior de la figura IV.B.17 se muestra la localización, predicha por el análisis termodinámico estructural, para el resto glutámico en el complejo con Endothiapepsina.



**Figura IV.B.19:** Contribuciones diferenciales con respecto al complejo con el inhibidor original (Pepstatina A) de los restos de la enzima y los inhibidores mutantes (Ala5Phe en el panel superior y Glu7 en el inferior).

En el panel inferior de la figura IV.B.19 se muestran las diferencias observadas entre las contribuciones de los restos individuales a la energía libre de unión en el complejo E7 con

respecto al original. Como se puede observar, no se producen cambios significativos en las contribuciones procedentes de los restos de la enzima, sino que el efecto desestabilizador de la mutación puede atribuirse casi por completo a la presencia del resto glutámico en la posición 7 del inhibidor.

#### IV.B.3.3.2.-Unión de sustratos a la proteasa del virus V.I.H.-1.

##### a.-Consideraciones generales.

Como se ha discutido anteriormente, la pérdida de sensibilidad de la enzima hacia sus inhibidores se produce debido a la aparición de cepas víricas que codifican moléculas de proteasa con mutaciones específicas, que disminuyen la afinidad por los inhibidores aunque manteniendo la actividad catalítica a niveles aceptables para la supervivencia del virus. Por tanto, el desarrollo de una nueva generación de inhibidores que permita dar solución al problema de la resistencia requiere la comprensión de las interacciones diferenciales de la enzima con inhibidores y sustratos.

En principio, la resistencia a un inhibidor determinado puede proceder de la reducción preferencial de la afinidad de unión con respecto a éste, o bien de un incremento en la actividad catalítica que compense parcialmente el efecto de la inhibición. Parece ser que, para la mayoría de los mutantes estudiados, las actividades catalíticas permanecen aproximadamente constantes en comparación con la actividad de la enzima nativa, observándose, sin embargo, una disminución significativa de la afinidad de unión hacia los inhibidores (Baldwin et al., 1995b; Hong et al., 1996; Lin et al., 1995).

Para algunas mutaciones la afinidad hacia el inhibidor disminuye hasta tres órdenes de magnitud, mientras que la constante de Michaelis ( $K_m$ ) para el sustrato sufre sólo una pequeña variación, de menos de un orden de magnitud (Baldwin et al., 1995b; Hong et al., 1996; Lin et al., 1995). En estos casos, es posible considerar la existencia de diferencias en las interacciones establecidas entre los sustratos e inhibidores con la enzima como el posible origen de la resistencia. De este modo, si un inhibidor tuviese la capacidad para imitar de forma precisa el modo de unión del sustrato a la enzima, la probabilidad de discriminación entre ambos

disminuiría drásticamente, dado que, en este caso, cualquier modificación que afectase a la afinidad de unión del inhibidor afectaría de modo similar a la unión del sustrato.

Las diferencias en el modo de unión de inhibidores y sustratos pueden deberse fundamentalmente a dos factores: bien a la existencia de ciertos restos en la enzima que son críticos para la unión de los inhibidores y no para los sustratos, o bien a la existencia de diferencias en cuanto a la naturaleza termodinámica de las interacciones implicadas en la unión de unos y otros con la enzima. En el primer caso, la situación es relativamente simple, dado que, si un resto determinado proporciona una contribución significativa a la unión del inhibidor, pero no a la del sustrato, su mutación provocará resistencia a dicho inhibidor. La segunda situación es más compleja debido a que, en este caso, sería necesario considerar las contribuciones relativas de cada grupo a la energética global del proceso de unión, así como el balance existente entre las contribuciones entálpicas y entrópicas, entre la entropía conformacional y de solvatación, etc.

#### *b.-Descripción de los modelos estructurales.*

La figura IV.B.20 muestra dos orientaciones diferentes de las estructuras tridimensionales obtenidas como resultado del proceso de modelado descrito en la sección IV.B.1.3.2. Se muestran en azul las dos estructuras generadas para el sustrato nonapeptídico natural.

Estas estructuras se han utilizado para el estudio de la energética de unión de los sustratos a la enzima, mediante la aplicación de la parametrización estructural de la energética y del formalismo descrito anteriormente. De este modo, se hace posible la comparación entre los determinantes moleculares de las afinidades de unión en ambos casos y el esclarecimiento de los posibles orígenes del fenómeno de la resistencia.

#### *c.-Análisis termodinámico-estructural del proceso de unión.*

En la tabla IV.B.5 se muestran los valores calculados para las distintas magnitudes termodinámicas correspondientes al proceso de unión entre la proteína y el sustrato obtenidos con ambos modelos. Asimismo se muestran los valores determinados experimentalmente en el laboratorio del Dr. Freire mediante calorimetría isotérmica de titulación a pH 5.5 y a diferentes temperaturas comprendidas en el intervalo entre 15 y 35 °C (Luque et al., 1998).

**Tabla IV.B.5: Análisis termodinámico estructural de la unión del sustrato natural a la proteasa del virus V.I.H.-1**

		4HVP	7HVP	Experimental
$\Delta C_p$	$\text{cal}\cdot(\text{Kmol})^{-1}$	-445	-489	-----
$\Delta H(25\text{ }^\circ\text{C})$	$\text{kcal}\cdot\text{mol}^{-1}$	8.5	11.4	10.5
$\Delta S(25\text{ }^\circ\text{C})$	$\text{cal}\cdot(\text{Kmol})^{-1}$	44.1	55.8	54.7
$\Delta G(25\text{ }^\circ\text{C})$	$\text{kcal}\cdot\text{mol}^{-1}$	-5.4	-6.1	-5.8
$K_d(25\text{ }^\circ\text{C})$	( $\mu\text{M}$ )	105	36.5	120

El análisis termodinámico-estructural proporciona un valor promedio entre ambos modelos para la energía libre de unión de  $-5.75 \pm 0.45 \text{ kcal}\cdot\text{mol}^{-1}$  a  $25\text{ }^\circ\text{C}$  y pH 5.5. Los valores predichos para  $\Delta G$  son consistentes con una afinidad de unión promedio de  $59.5\text{ }\mu\text{M}$ , en buena correspondencia con el valor de  $54\text{ }\mu\text{M}$  determinado experimentalmente en el laboratorio del Dr. Freire bajo idénticas condiciones experimentales, y el de  $150\text{ }\mu\text{M}$  descrito con anterioridad en la literatura (Dunn et al., 1994) a  $37\text{ }^\circ\text{C}$ .

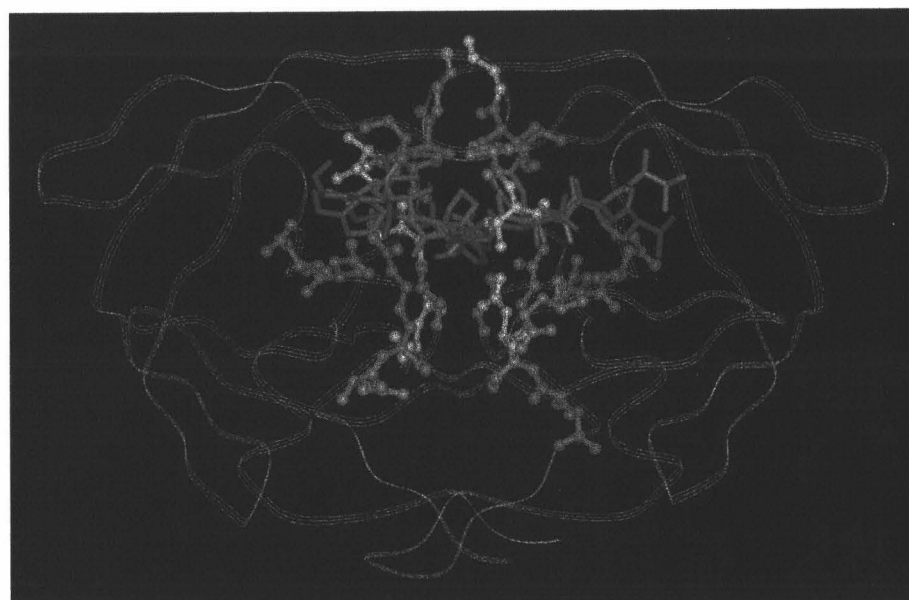
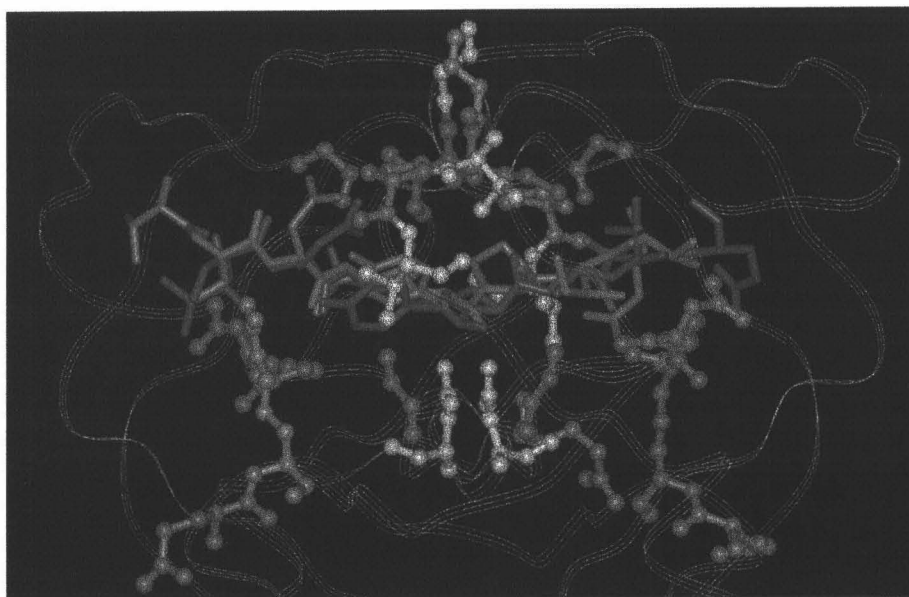
Asimismo, los valores obtenidos para los cambios de entalpía genérica y capacidad calorífica se encuentran de acuerdo con los resultados experimentales, de modo que los valores promedio entre ambos modelos son  $9.95 \pm 2.05 \text{ kcal}\cdot\text{mol}^{-1}$  y  $-467 \pm 31 \text{ cal}\cdot(\text{K}\cdot\text{mol})^{-1}$  y  $9.95 \pm 2.05 \text{ kcal}\cdot\text{mol}^{-1}$ , respectivamente. De acuerdo con estos resultados, tanto el análisis termodinámico estructural como los resultados experimentales indican que la unión de los sustratos a la proteasa se encuentra desfavorecida entálpicamente, por lo que las contribuciones favorables a la energía libre de Gibbs deben tener un origen entrópico.

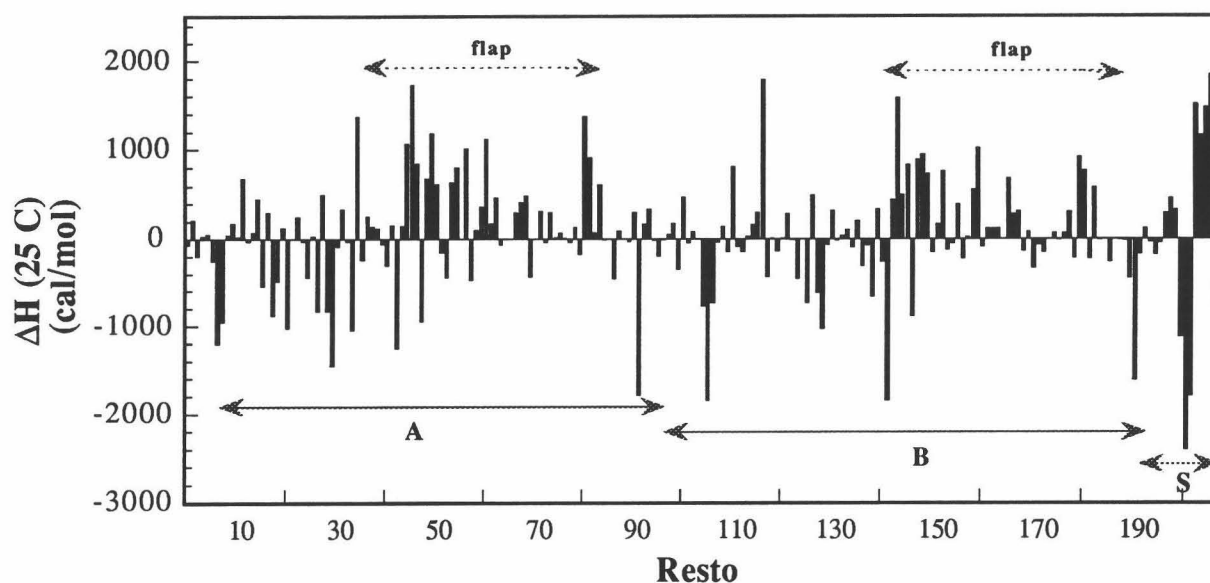
#### c.1.-Localización estructural de las principales contribuciones a la energética de unión.

El análisis de la distribución estructural de las contribuciones a la entalpía de unión, recogida en la figura IV.B.21, indica que la mayor parte de las contribuciones desfavorables tienen su origen en la reorganización de la región del flap en cada uno de los monómeros como consecuencia de la unión del ligando. Este cambio conformacional implica fundamentalmente a los restos correspondientes a la región del flap (restos en posiciones 40 a 60 en la secuencia) y la hebra beta adyacente correspondiente a los restos 60 a 84. Es interesante hacer notar que, al igual



*Figura IV.B.20: Localización de los sustratos modelados en este trabajo en la estructura tridimensional del complejo. Se representan aquellos restos de la enzima que contribuyen, según las predicciones, con más de 0.2 kcal·mol<sup>-1</sup> a la energía libre de Gibbs genérica para el proceso de unión, coloreados de acuerdo con la magnitud de sus contribuciones. Se muestran dos orientaciones diferentes, en las que los restos coloreados en rojo contribuyen con más de 0.8 kcal·mol<sup>-1</sup> a afinidad de unión, los naranjas entre 0.8 y 0.6 kcal·mol<sup>-1</sup>, los representados en amarillo entre 0.6 y 0.4 kcal·mol<sup>-1</sup>, y los blancos entre 0.4 y 0.2 kcal·mol<sup>-1</sup>.*

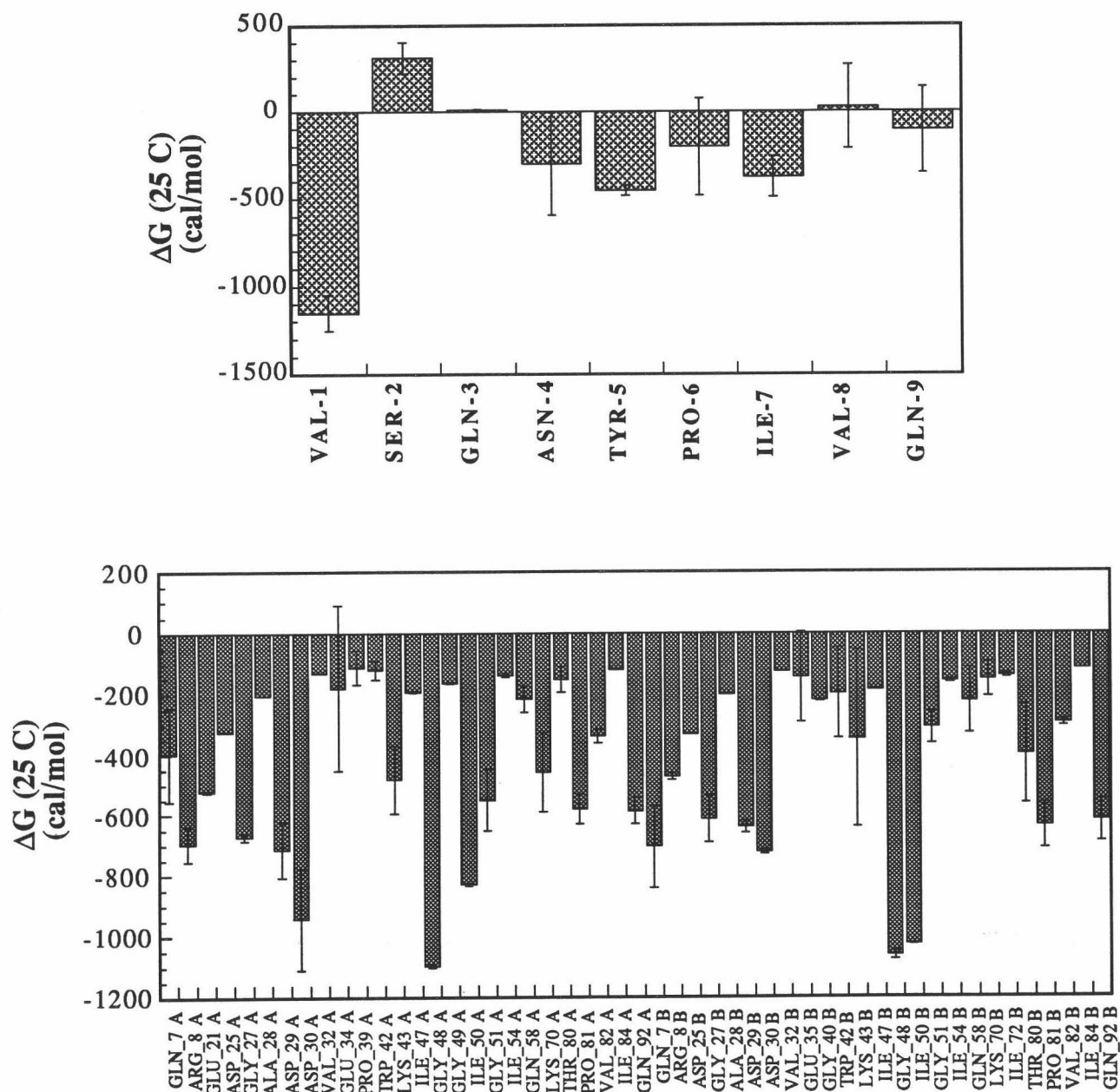




*Figura IV.B.21: Contribuciones individuales de los restos de la proteasa y del sustrato a la entalpía de unión. Se muestran los promedios entre los dos modelos construidos para el sustrato.*

que se observa en el caso de la unión del inhibidor peptídico acetil-pepstatina, el cambio en entalpía asociado a las interacciones entre la proteína y el sustrato *per se* es favorable y del orden de unas  $-3 \text{ kcal}\cdot\text{mol}^{-1}$ , aproximadamente.

En la figura IV.B.21, donde se recoge la estructura teórica del complejo entre los sustratos y la proteasa, se representan aquellos restos de la enzima que contribuyen, según las predicciones, con más de  $0.2 \text{ kcal}\cdot\text{mol}^{-1}$  a la energía libre de Gibbs genérica para el proceso de unión, coloreados de acuerdo con la magnitud de sus contribuciones. Se muestran dos orientaciones diferentes, en las que los restos coloreados en rojo contribuyen con más de  $0.8 \text{ kcal}\cdot\text{mol}^{-1}$  a afinidad de unión, los naranjas entre  $0.8$  y  $0.6 \text{ kcal}\cdot\text{mol}^{-1}$ , los representados en amarillo entre  $0.6$  y  $0.4 \text{ kcal}\cdot\text{mol}^{-1}$ , y los blancos entre  $0.4$  y  $0.2 \text{ kcal}\cdot\text{mol}^{-1}$ . Esta estructura sugiere la existencia de intensos contactos de van der Waals entre la proteína y el sustrato, particularmente fuertes en la región cercana al resto Val 1 del péptido, que corresponde a la posición P5, y que se encuentra relativamente distante del centro activo. Estas interacciones contribuyen favorablemente a la entalpía de unión.



**Figura IV.B.22:** Contribuciones individuales a la energía libre de unión del sustrato natural a la proteasa del virus V.I.H.-1. Se representa el promedio de los valores obtenidos con cada uno de los modelos, junto con su error estándar. En el panel superior se muestran las contribuciones de los restos del sustrato y en el inferior las procedentes de la enzima.

En el panel superior de la figura IV.B.22 se muestran las contribuciones de los restos del sustrato a la energía de Gibbs genérica. Los valores representados en la figura corresponden al promedio entre ambos modelos estructurales y sus errores estándar. Se puede observar que la mayor contribución procede del resto Val 1, seguido por Tyr 5, Pro 6 e Ile 7, mientras que, sorprendentemente, las contribuciones de los restantes aminoácidos son casi nulas. El análisis del mapa completo para la proteína (mostrado en el panel inferior de la misma figura, donde se representan los restos que contribuyen con más de  $0.1 \text{ kcal}\cdot\text{mol}^{-1}$ ) indica que gran parte de la energía de unión procede de la propia enzima.

### *c.2.-Especificidad de los sustratos.*

Experimentalmente se ha observado una notable disminución, casi de un orden de magnitud, en la constante de Michaelis ( $K_m$ ) correspondiente al heptapéptido, resultante de eliminar los restos Val1 y Gln9 de los extremos, con respecto al sustrato de nueve restos analizado en este trabajo (Dunn et al., 1994). Esta observación experimental es consistente con nuestras predicciones, que atribuyen un papel importante en la determinación de la energética de unión al resto Val en posición 1, de forma que, según nuestro cálculos, la mayor parte de las diferencias energéticas pueden atribuirse a la ausencia de dicho aminoácido.

De igual modo, la ausencia de contribuciones estabilizadoras por parte del resto Gln3 del sustrato está de acuerdo con las conclusiones desprendidas de los distintos estudios experimentales de especificidad, según los cuales, la naturaleza de los restos en posición P3 carece de especial relevancia. Asimismo, como ya se ha comentado, el análisis termodinámico estructural resalta el papel desempeñado por los restos del sustrato situados en posiciones alejadas del enlace peptídico hidrolizable, como es el caso de la Val-1 situada en posición P5, lo cual está también de acuerdo con las observaciones experimentales realizadas por Dunn y colaboradores (Dunn et al., 1994; Griffiths et al., 1992a). De este modo, en resumen, tanto los resultados experimentales como los obtenidos a partir de nuestra metodología indican que, en el sustrato natural analizado en este trabajo, las mayores contribuciones a la energética de unión proceden de los restos directamente adyacentes al enlace peptídico hidrolizable (P1, P1' y P2'), así como de los aminoácidos situados en posición P5.

#### IV.B.3.3.3.-Bases moleculares de la resistencia a fármacos en la proteasa del virus H.I.V.-1

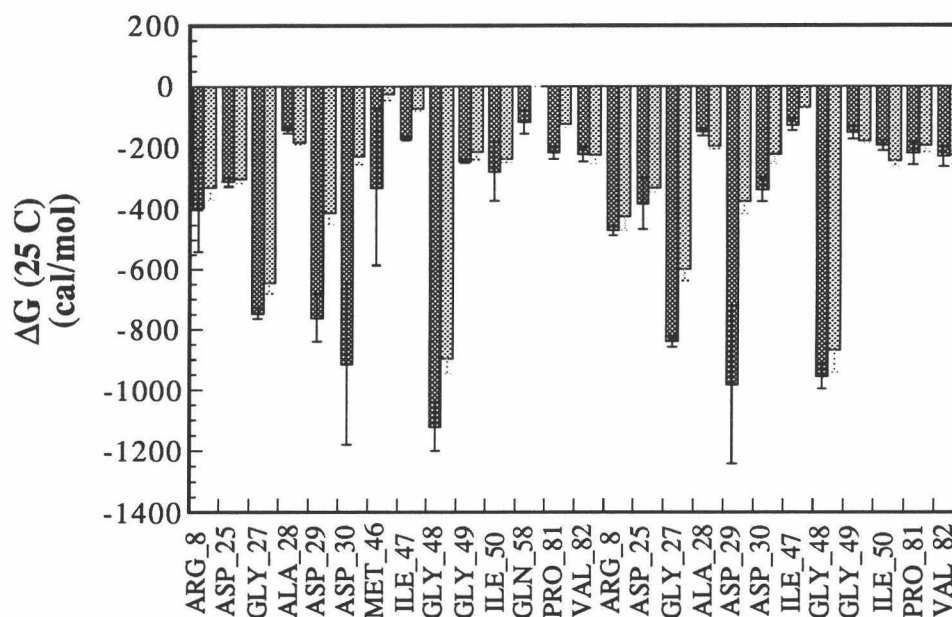
*a.-Análisis comparativo de la distribución estructural de las contribuciones a la energética de unión de sustratos peptídicos e inhibidores orgánicos.*

Una vez caracterizado termodinámicamente y estructuralmente el proceso de unión, tanto de inhibidores peptídicos como no peptídicos, a la proteasa es conveniente realizar un análisis comparativo entre ambos, con el fin de identificar aquellas características diferenciales que pudieran ser responsables del desarrollo de resistencia frente a los distintos fármacos.

Los inhibidores no peptídicos, cuya unión se ha estudiado en la sección IV.B.2.2.1, constituyen análogos de los sustratos, que han sido diseñados para competir por el mismo sitio de unión, por lo que se espera que las interacciones se establezcan esencialmente con el mismo grupo de restos de la proteasa con que lo hacen los sustratos peptídicos. El análisis termodinámico estructural llevado a cabo con ambos tipos de ligandos apoya esta hipótesis, como puede deducirse de la comparación entre las distribuciones estructurales de las contribuciones de la enzima a la energética de unión de sustratos e inhibidores representada en la Figura IV.B.23. Dicha gráfica contiene únicamente las contribuciones genéricas aportadas por los restos de la enzima, de modo que, con objeto de simplificar la representación, se han representado únicamente aquellos aminoácidos que contribuyen con más de 0.2 kcal·mol<sup>-1</sup> a la energética de unión.

Del análisis de la gráfica pueden extraerse varias conclusiones importantes. En primer lugar, reiterar el hecho de que, en ambos casos, las contribuciones proceden esencialmente de los mismos restos de la enzima, aunque con magnitudes diferentes para sustratos e inhibidores. Asimismo, puede observarse que, incluso aunque la afinidad de unión es mayor para los inhibidores peptídicos, las contribuciones intrínsecas aportadas por la enzima son mayores en el proceso de unión de los sustratos. De este modo, la suma de las contribuciones de los restos de la proteasa a la unión del sustrato es -11 kcal·mol<sup>-1</sup>, mientras que la suma total promedio para todos los inhibidores es  $-7.8 \pm 0.6$  kcal·mol<sup>-1</sup>. Como consecuencia, parece razonable afirmar que el sustrato posee mayor capacidad para establecer contactos intensos de van der Waals con la enzima que los inhibidores sintéticos, aún a pesar de la mayor afinidad de éstos. El origen de este comportamiento puede explicarse como consecuencia de la naturaleza química de dichos inhibidores en comparación con los sustratos peptídicos.





*Figura IV.B.23: Comparación de las contribuciones individuales de los restos de la proteasa del virus del V.I.H.-1 a la unión de inhibidores no peptídicos (Barras oscuras) y sustratos (barras claras). Se muestran los valores promedio para los distintos inhibidores y los dos modelos del sustrato natural así como los errores estándar correspondientes.*

#### **b.-Diferencias fundamentales en la energética de unión.**

Tanto los resultados experimentales como los obtenidos a partir de los análisis termodinámico-estructurales, indican la existencia de una clara diferencia en la termodinámica de unión de sustratos peptídicos e inhibidores orgánicos. El análisis termodinámico-estructural del proceso de unión de los trece inhibidores no peptídicos considerados en este trabajo predice un cambio entálpico desfavorable para el proceso de unión, con valores comprendidos entre 1.3 y 12.2 kcal·mol<sup>-1</sup> para los distintos inhibidores (Bardi et al., 1997). La única medida calorimétrica de la entalpía de unión de uno de estos inhibidores (SB-203386), aunque llevada a cabo con la proteasa del virus V.I.S., proporciona un valor de 6.4 kcal·mol<sup>-1</sup>, que se encuentra dentro del intervalo esperado (Hoog et al., 1996). Las entalpías de unión correspondientes a los sustratos son de la misma magnitud. Sin embargo, aunque en ambos casos las contribuciones entálpicas son

desfavorables, la afinidad de unión de los inhibidores no peptídicos se encuentran en el intervalo nanomolar, mientras que la correspondiente a los sustratos es del orden micromolar. Como consecuencia, tanto los inhibidores como los sustratos, deben compensar la contribución desfavorable de la entalpía con un cambio entrópico favorable al proceso de asociación.

Las principales contribuciones al cambio de entropía asociado a la unión de los ligandos son las correspondientes a la entropía de solvatación y al cambio de entropía conformacional, de signos opuestos entre sí (ver tabla IV.B.2 para los inhibidores no peptídicos y tabla IV.B.5 para los sustratos). La única contribución favorable, por lo tanto, procede del incremento en la entropía del disolvente, originado por la liberación de las moléculas de agua implicadas en la solvatación de las superficies hidrofóbicas que se ocultan al disolvente como consecuencia de la unión. Esta ganancia en entropía se compensa parcialmente por la disminución en entropía conformacional y translacional que implica la unión de la enzima y el ligando. Dado que los sustratos peptídicos poseen una mayor flexibilidad en disolución, su unión a la proteína implica la pérdida de un número significativo de grados de libertad. Por el contrario, los inhibidores orgánicos, en su mayoría pequeñas moléculas rígidas, poseen un menor número de grados de libertad, por lo que su unión a la enzima no supone una gran penalización entrópica.

Para los trece inhibidores orgánicos, el valor promedio para el cambio en entropía conformacional es de  $-20 \text{ cal}\cdot(\text{K}\cdot\text{mol})^{-1}$  con una desviación estándar de  $\pm 7 \text{ cal}\cdot(\text{K}\cdot\text{mol})^{-1}$  (Bardi et al., 1997), mientras que el valor correspondiente a la unión del sustrato nonapeptídico es de  $-67 \text{ cal}\cdot(\text{K}\cdot\text{mol})^{-1}$ . La diferencia en entropía conformacional entre ambos se traduce en  $14 \text{ kcal}\cdot\text{mol}^{-1}$  en energía libre de unión, diferencia que no queda completamente compensada por las interacciones entálpicas que favorecen la unión de los sustratos peptídicos. Como consecuencia, globalmente, los inhibidores no peptídicos poseen una mayor afinidad de unión debido, fundamentalmente, a su mayor rigidez en disolución.

### *c.-Hipótesis sobre el origen molecular de la resistencia.*

Los resultados presentados hasta ahora sugieren una posible explicación del origen molecular del fenómeno de la resistencia. Gracias a su elevada flexibilidad, los sustratos peptídicos son capaces de adaptar su conformación a los sutiles cambios conformacionales provocados en la estructura de la enzima como consecuencia de la aparición de una mutación determinada. Por el contrario, moléculas como los inhibidores sintéticos, caracterizados por una mayor rigidez

estructural, poseen una capacidad de adaptación muy limitada. Como resultado, mientras el sustrato peptídico es capaz de mantener fuertes contactos favorables con la proteasa mutante, el inhibidor no puede acomodarse a las distorsiones experimentadas por el sitio de unión.

En general, los inhibidores con baja flexibilidad poseerán, asimismo, una mayor selectividad hacia la proteína diana, aunque, como contrapartida, son susceptibles de perder su elevada afinidad a causa de pequeñas modificaciones en el sitio de unión. La elevada especificidad de los inhibidores sintéticos hacia la proteasa del virus V.I.H.-1 se debe, en parte, a su rigidez conformacional. Por ejemplo, el inhibidor peptídico acetil-pepstatina posee una mayor flexibilidad conformacional y, como consecuencia, su afinidad por los mutantes de la enzima disminuye en mucho menor grado que la de los inhibidores no peptídicos. Como contrapartida, este pentapéptido, que actúa como inhibidor de la mayoría de las proteasas aspárticas, posee una especificidad mucho menor.

Esta hipótesis se ve apoyada por la información estructural existente para algunos de los mutantes resistentes de la enzima. Por ejemplo, se ha demostrado que el efecto de la mutación V82A en la afinidad de unión no puede racionalizarse en términos de la eliminación de un grupo metilo en cada cadena de la proteasa, sino que, por el contrario, el análisis de la estructura cristalográfica revela la reorganización, extendida a lo largo de todo el sitio de unión, de los grupos del esqueleto peptídico (Baldwin et al., 1995b). Asimismo, el análisis termodinámico-estructural llevado a cabo en este trabajo (ver sección IV.B.2.2.1) indica que la energía libre de Gibbs se encuentra redistribuida en todo el sitio de unión. Resultados similares se han obtenido a partir del estudio de la unión entre el inhibidor U-89360E y los mutantes V82D y V82N (Hong et al., 1996). En ambos casos, la pérdida de afinidad de los inhibidores puede atribuirse a su incapacidad para adaptarse a un sitio de unión ligeramente distorsionado.

Como consecuencia, las diferencias termodinámicas observadas entre la unión de sustratos e inhibidores parecen identificar el balance relativo entre las contribuciones entálpicas y entrópicas a la energía libre de Gibbs de unión, y, particularmente, el papel desempeñado por la entropía conformacional en su determinación, como un factor esencial para la comprensión del fenómeno de la resistencia. De este modo, y en este contexto termodinámico, especificidad y resistencia pueden entenderse como dos caras de una misma moneda.

**V.-APÉNDICES**



**APENDICE A.2.****AREAS ACCESIBLES AL DISOLVENTE. ALGORITMOS DE CALCULO.****A.2.1.-Definición de área accesible y superficie molecular.**

Debido a la naturaleza difusa de la distribución radial de densidad electrónica que rodea cada núcleo atómico, la posición aparente de la superficie molecular depende de la técnica empleada para su examen de modo que, dependiendo del tipo de estudio es posible emplear diferentes definiciones de para las superficies moleculares (superficies de densidad constante, superficies equipotenciales, superficies de van der Waals, etc.).

Según el modelo de esferas rígidas una molécula cualquiera puede representarse mediante una colección de esferas solapadas entre sí y de radios equivalentes a los radios de van der Waals correspondientes a los átomos que representan. Aún a pesar de su naturaleza aproximada, este modelo proporciona una descripción adecuada de la superficie y volúmenes moleculares, fundamentalmente debido al carácter fuertemente abrupto del término repulsivo que describe las interacciones no covalentes en la función potencial.

Tomando como punto de partida este modelo, es posible distinguir tres tipos de superficies moleculares: la superficie de van der Waals, la superficie accesible al disolvente y la denominada superficie molecular (Richards, 1977; Pascual-Ahuir et al., 1994; Eisenhaber et al., 1995). La superficie de van der Waals corresponde a la superficie externa resultante de la representación de la molécula de acuerdo al modelo de esferas rígidas, mientras que la superficie accesible al disolvente, definida originariamente por Lee y Richards (Lee & Richards, 1971), puede describirse como la superficie generada por el centro de una molécula de disolvente o sonda, considerada a su vez como una esfera rígida, cuando se desplaza sobre la superficie de van der Waals. De este modo, ambas superficies coinciden si se considera un radio para la esfera sonda nulo. La superficie molecular, por otra parte, representa la región excluida al disolvente por la molécula y se define como la suma de dos contribuciones: la superficie de contacto, correspondiente a la porción de la superficie de van der Waals de cada átomo accesible a la esfera sonda y la superficie reentrante, definida como la superficie correspondiente a la parte inferior de la esfera sonda cuando ésta se encuentran con contacto con más de un átomo.

**A.2.2.- Radios de van der Waals.**



A pesar de la utilidad que la aproximación de esferas rígidas ha demostrado para la representación molecular, conviene hacer notar que la superficie de van der Waals de una molécula no se encuentra definida de un modo unívoco. Al contrario, dentro de ciertos límites, la elección de los radios de van der Waals asignados cada tipo de átomo es arbitraria, existiendo en la actualidad numerosos juegos de valores que difieren entre sí tanto en el procedimiento seguido para su obtención como en sus posibles aplicaciones (Bondi, 1964; Ramachandran & Sasisekharan, 1968; Chothia, 1975; Eisenberg & McLachlan, 1986; Shrake & Rupley, 1973; Rose et al., 1985; Silla et al., 1991; Lee & Richards, 1971; Richmond & Richards, 1978; Gibson & Scheraga, 1987; Ooi et al., 1987).

En general, los radios de van der Waals se encuentran estrechamente relacionados con la función potencial asociada a las interacciones no covalentes, de modo que generalmente la suma de los radios de dos átomos cualesquiera se iguala, bien a la distancia correspondiente al mínimo en la función potencial o bien a la menor distancia para la cual el potencial se hace cero. Considerando un potencial del tipo Lennard-Jones, estos dos valores difieren aproximadamente en un 12%. Asimismo, dado que existen distintas alternativas para la definición de los parámetros asociados a dichas funciones potenciales, es posible, partiendo del mismo juego de datos experimentales, obtener valores diferentes para los radios atómicos, por lo que es conveniente tomar en consideración el procedimiento seguido para la definición de cada conjunto de radios previamente a su aplicación a un problema determinado (Richards, 1985).

Dada la complejidad de los sistemas, cuando se trabaja con macromoléculas biológicas se recurre con frecuencia a la asignación de un único radio a pequeños grupos de átomos (átomos unificados o extendidos), generalmente consistente en un único átomo pesado y los átomos de hidrógeno unidos directamente al mismo, en lugar de considerar cada uno de los átomos individualmente. A pesar de que la relación entre dichos radios unificados y los correspondientes a los átomos individuales no es siempre obvia, su empleo se encuentra muy extendido en el estudio de procesos asociados a proteínas, y especialmente en el cálculo de superficies moleculares (Lee & Richards, 1971; Shrake & Rupley, 1973; Chothia, 1975; Richmond & Richards, 1978).

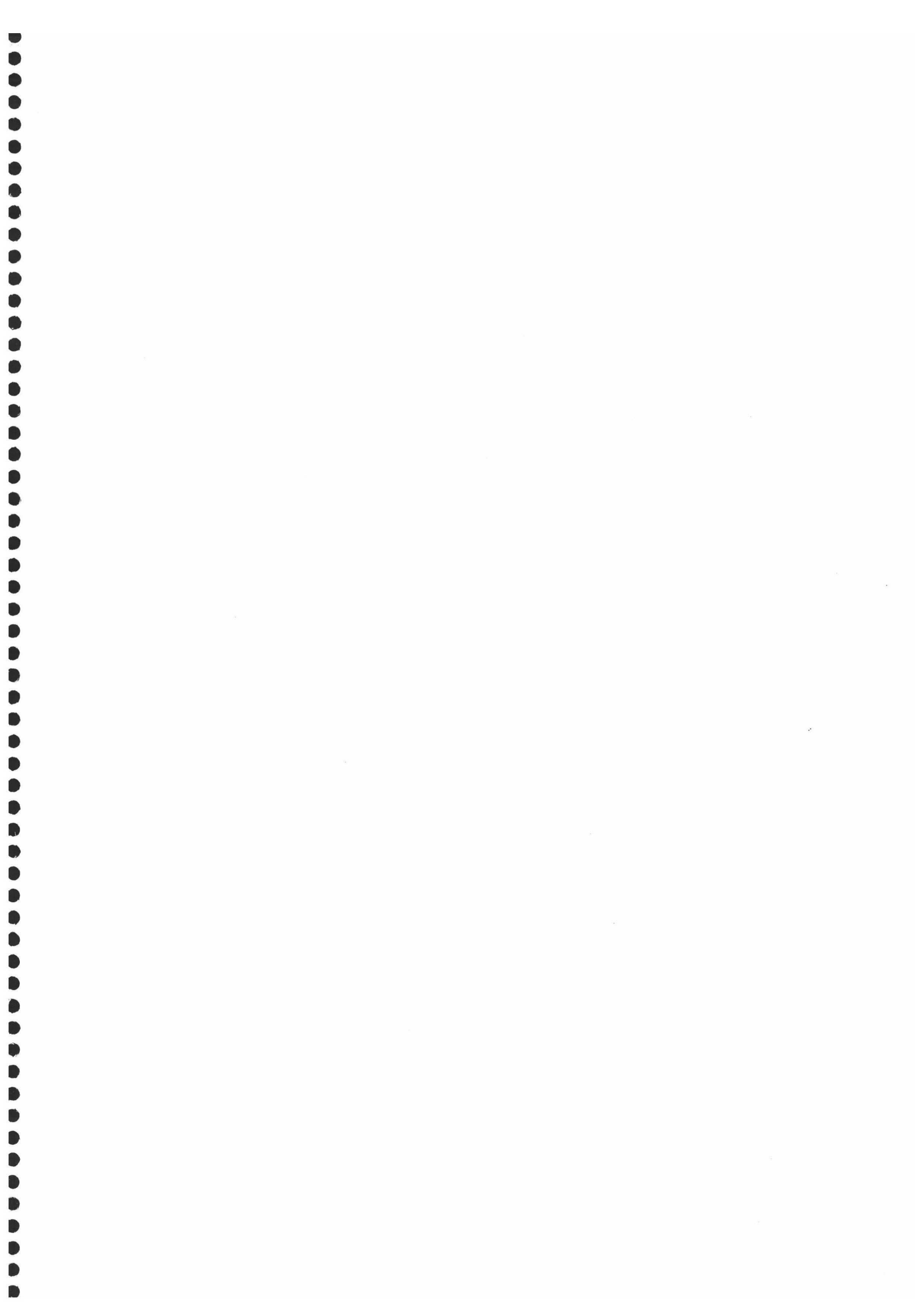
### **A.2.3.- Métodos de cálculo de superficies accesibles.**

Actualmente se dispone de un gran número de algoritmos que permiten el cálculo de las superficies moleculares a partir de las coordenadas atómicas derivadas mediante técnicas estructurales como espectroscopía de RMN o difracción de rayos X. Aunque algunos de ellos, con objeto de acelerar el proceso de cálculo y conseguir la minimización de los requerimientos computacionales, recurren a una representación simplificada de la molécula a nivel de aminoácidos u otros segmentos estructurales (Wodak & Janin, 1980; Zehfus et al., 1985), la gran mayoría se basan en la descripción de la molécula a nivel atómico de acuerdo con el modelo de esferas rígidas.

En general cabe considerar dos tipos de algoritmos según realicen el cálculo de las superficies de forma numérica o analítica. Los métodos analíticos (Connolly, 1983; Richmond, 1984; Gibson & Scheraga, 1987; Dodd & Theodorou, 1991; Perrot et al., 1992; Eisenhaber & Argos, 1993; Petitjean, 1994)), proporcionan, bien mediante el principio de inclusión-exclusión (Gibson, 1987; Petitjean, 1994), bien mediante la intersección de planos y esferas (Connolly, 1983; Richmond, 1984; Perrot et al., 1992; Eisenhaber & Argos, 1993; Dodd & Theodorou, 1991), una descripción precisa de la superficie molecular.

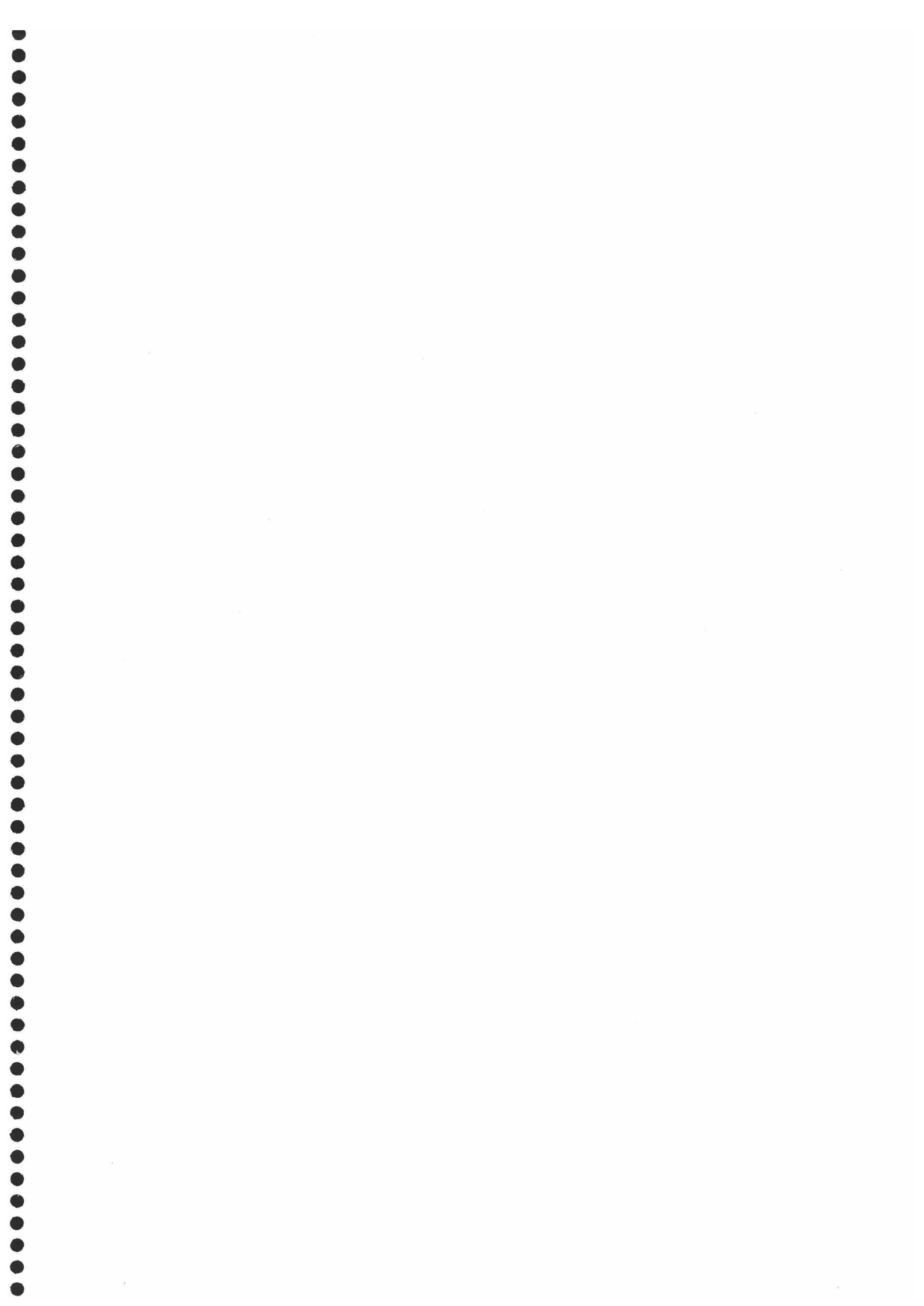
Estos métodos ofrecen además la posibilidad de obtener analíticamente la derivada del área con respecto a las coordenadas atómicas, tanto cartesianas (Richmond, 1984; Perrot et al., 1992) como internas (Eisenhaber & Argos, 1993), posibilitando su empleo en procesos de minimización y dinámica molecular. Sin embargo, su aplicación a macromoléculas se encuentra muy limitada debido al enorme incremento en el tiempo de computación con el número de esferas a considerar. Actualmente, la optimización de la velocidad de cálculo se ha convertido en uno de los objetivos fundamentales en el desarrollo de nuevos algoritmos (Perrot et al., 1992).

Las aproximaciones numéricas al cálculo de las superficies moleculares requieren, en general, menores tiempos de computación, por lo que han sido aplicados con mayor frecuencia al estudio de macromoléculas biológicas. Según el proceso seguido para aproximar la superficie molecular, cabe distinguir varios tipos de algoritmos: los basados en la generación de secciones cilíndricas de la molécula (Lee & Richards, 1971; Richmond & Richards, 1978), aquéllos que recurren al empleo de composiciones cúbicas (Müller, 1983; Pavlov & Fedorov, 1983; Meyer, 1988; Karfunkel & Eyraud, 1989) o bien los que consisten en la generación de distribuciones de puntos sobre las esferas atómicas (Shrake & Rupley, 1973; Wang & Levinthal, 1991; Pascual-Ahuir & Silla, 1991; Silla et al., 1991; LeGrand & Kenneth Jr, 1993; Abagyan et al., 1994; Eisenhaber et al., 1995; Bliznyuk & Gready, 1996).





## **VI.-CONCLUSIONES**



## V.-CONCLUSIONES

A partir de la interpretación y discusión de las metodologías de cálculo propuestas así como de los resultados obtenidos mediante su aplicación, que han sido presentados a lo largo de esta Memoria, se proponen las siguientes conclusiones:

1.-La accesibilidad al disolvente de una cadena polipeptídica en disolución se encuentra modulada por sus posibles cambios conformacionales, por lo que debe calcularse como una magnitud promedio determinada por las contribuciones de todas las posibles conformaciones ponderadas de acuerdo con su probabilidad. En este sentido se han determinado los promedios de Boltzmann para las áreas accesibles al disolvente de los distintos aminoácidos en conformación helicoidal. No se han encontrado grandes diferencias entre los promedios ponderados energéticamente y los valores obtenidos como la media aritmética de las accesibilidades de todas las conformaciones permitidas estéricamente, lo que indica la ausencia de grandes variaciones en accesibilidad dentro de los límites del espacio conformacional accesible a las distintas cadenas laterales.

2.- Las áreas accesibles al disolvente en el estado desplegado se han refinado mediante la aplicación de la parametrización estructural de la energética descrita en esta memoria al estudio de propensiones helicoidales de los distintos aminoácidos. Se ha comprobado la posibilidad de reproducir los datos de  $\Delta\Delta G$  procedentes de distintos estudios experimentales utilizando un único juego de valores para el área accesible al disolvente para los aminoácidos en el estado desplegado. Estos valores se encuentran comprendidos en el intervalo de accesibilidades que cada aminoácido puede presentar en el espacio conformacional permitido. En general los valores optimizados para las áreas expuestas al disolvente en el estado desplegado son un 85% menores que los calculados a partir de la media aritmética de las accesibilidades correspondientes a las conformaciones permitidas estéricamente. Se observa una tendencia general en las cadenas laterales a minimizar la superficie expuesta al disolvente con respecto a la conformación extendida.



3.- La utilización de ambos juegos de accesibilidades (tanto los correspondientes a los promedios de Boltzmann en conformación helicoidal, como los valores optimizados para el estado desplegado) ha posibilitado la elaboración de una escala de propensión helicoidal que se encuentra en excelente concordancia con las distintas escalas, tanto experimentales como calculadas mediante otros procedimientos semiempíricos, existentes en la literatura.

4.-La aplicación de la parametrización estructural de la energética permite llevar a cabo la disección de las contribuciones de las distintas magnitudes termodinámicas a los valores de propensión helicoidal de los distintos aminoácidos. Las propensiones helicoidales se encuentran compendidas en un intervalo de  $0.8 \text{ kcal}\cdot\text{mol}^{-1}$ , de modo que Ala es el aminoácido con mayor tendencia helicoidal y Gly el de carácter más fuertemente desestabilizador de la hélice. Los valores de energía libre son el resultado de un delicado balance entre términos entrópicos y entálpicos que se compensan entre sí.

El cambio de entalpía promedio predicho para la formación de una hélice alfa expuesta al disolvente y en ausencia de interacciones terciarias es  $-0.7 \text{ kcal}\cdot(\text{mol}\cdot\text{res})^{-1}$  a  $25^\circ\text{C}$ , valor cercano a los obtenidos experimentalmente. En general las diferencias existentes entre los valores de entalpía correspondientes a los distintos aminoácidos proceden de la deshidratación parcial de las cadenas laterales como consecuencia de la formación de la hélice, así como de los efectos que éstas puedan tener sobre la hidratación del esqueleto peptídico.

En cuanto a las contribuciones entrópicas se observa que la pérdida de entropía conformacional del esqueleto peptídico y de las cadenas laterales como consecuencia de la formación de la hélice contribuye notablemente a la determinación de las diferencias entre la propensiones helicoidales de los distintos aminoácidos, aunque este efecto se encuentra parcialmente compensado por la entropía de hidratación.

5.- Los resultados anteriores demuestran que, una vez optimizadas las accesibilidades de los distintos aminoácidos en el estado desplegado, la parametrización estructural de la energética de plegamiento de proteínas descrita en esta memoria es suficientemente sensible como para analizar los efectos producidos por cambios en aminoácidos individuales. Su aplicación a la optimización de las conformaciones adoptadas por las cadenas laterales de los distintos

aminoácidos en conformación helicoidal indica que asimismo posee la resolución necesaria para identificar correctamente la posición de los mínimos energéticos en el espacio conformacional. Por lo tanto, la aplicación de esta metodología al desarrollo de técnicas de diseño molecular se encuentra justificada.

6.- La metodología de diseño molecular descrita en la memoria se ha aplicado al diseño de inhibidores peptídicos de Endothiapepsina derivados del inhibidor universal de las proteasas aspárticas Pepstatina A.

-En primer lugar se ha procedido a la generación de nuevos inhibidores mediante la sustitución de la Ala en posición 5 del inhibidor por distintos aminoácidos. Este resto ha sido identificado como el que aporta las menores contribuciones a la afinidad de unión. De los doce inhibidores generados, aquéllos en los que se han introducido aminoácidos con cadenas laterales aromáticas en dicha posición son los que muestran el mayor incremento en la afinidad de unión. Los resultados experimentales confirman estas predicciones, de modo que la diferencia entre las constantes predichas en este trabajo y las obtenidas experimentalmente es  $0.2 \text{ kcal}\cdot\text{mol}^{-1}$ .

-La energía libre de Gibbs de unión para el mutante con Fenilalanina es cerca de  $1 \text{ kcal}\cdot\text{mol}^{-1}$  más favorable que la correspondiente al complejo original, que es el resultado de una significativa contribución favorable de la entropía de deshidratación, cercana a  $4 \text{ kcal}\cdot\text{mol}^{-1}$ , que se encuentra parcialmente compensada por la entalpía de deshidratación y la mayor pérdida de entropía conformacional de la cadena lateral de Fenilalanina con respecto al resto de Alanina original.

-La elongación del péptido en el extremo carboxilo terminal, donde se ha añadido un resto Glutámico, conduce a la disminución de la afinidad de unión del inhibidor debido a que el nuevo resto no establece nuevas interacciones con los restos de la enzima, de modo que la pérdida de entropía conformacional de este aminoácido como consecuencia del proceso de unión se encuentra descompensada.

7.- La parametrización estructural de la energética descrita en esta memoria se ha aplicado al estudio termodinámico-estructural de la unión de distintos inhibidores tanto peptídicos como no peptídicos a la proteasa del virus V.I.H. tipo 1.

-Las energías libres de unión de los trece inhibidores no peptídicos considerados en este trabajo se predicen con un error estándar de  $0.3 \text{ kcal}\cdot\text{mol}^{-1}$ , y el análisis de regresión lineal entre los valores de  $\Delta G$  calculados y los experimentales origina una pendiente de 0.982 y un coeficiente de correlación de 0.85, no observándose desviaciones sistemáticas entre los datos. Por tanto, los cálculos estructurales poseen suficiente precisión como para permitir el examen de las diferentes contribuciones a la energética de unión.

-El proceso de unión de estos compuestos a la enzima se encuentra dominado por el efecto hidrofóbico, de modo que las mayores contribuciones a la energía libre de Gibbs de unión se originan como consecuencia de la deshidratación de un elevado porcentaje de superficie apolar. Las contribuciones entálpicas son desfavorables, debido fundamentalmente a las contribuciones procedentes de la entalpía de hidratación. Una situación similar se observa para la unión del pentapéptido Acetil-pepstatina.

-El análisis termodinámico-estructural del proceso de unión nos ha permitido identificar los restos de la enzima que contribuyen más significativamente a la afinidad. Asimismo, el estudio de las constantes de estabilidad individuales de los restos de la enzima ha posibilitado la clasificación de las regiones de la enzima según su estabilidad estructural. Los resultados de ambos análisis indican que el sitio de unión está constituido por restos pertenecientes a las regiones más y menos estables de la molécula, de modo que una parte del sitio de unión se encuentra performado con anterioridad a la unión del ligando, mientras que otra se constituye posteriormente como consecuencia de la presencia de éste.

-El análisis del proceso de unión del inhibidor no peptídico A-77003 al mutante resistente V82A de la proteasa. Los cálculos predicen una reducción de 3.7 veces en la afinidad de unión con respecto a la enzima silvestre, de acuerdo con las observaciones experimentales. Asimismo, los resultados del análisis indican que el efecto de la mutación

no puede asignarse a una posición determinada, sino que las diferencias en energía libre se encuentran redistribuidas a lo largo de todo el sitio de unión. Por lo tanto, queda claro que la interpretación de los efectos de las diversas mutaciones requieren un análisis global de la situación, incluso cuando la mutación se encuentra localizada en el sitio de unión.

8.-Con objeto de esclarecer los posibles orígenes de la resistencia de los distintos fármacos desarrollada por la proteasa del virus V.I.H.-1 se ha procedido al análisis de los determinantes moleculares de las afinidades de unión de los sustratos a la enzima.

-Para ello se requiere el desarrollo de modelos estructurales para los complejos entre la proteasa un sustratos peptídicos de secuencia correspondiente a uno de los sitios de corte de los precursores polipeptídicos, que se ha llevado a cabo utilizando la metodología de diseño desarrollada en esta memoria.

-Los valores termodinámicos predichos por la parametrización estructural se encuentran en buena correspondencia con los obtenidos experimentalmente, siendo  $59.5 \mu\text{M}$  el valor calculado para la afinidad de unión. El análisis termodinámico-estructural proporciona un valor indica que la unión de los sustratos a la proteasa se encuentra desfavorecida entálpicamente, debido a las contribuciones desfavorables procedentes de la reorganización de la región del "flap" como consecuencia del proceso de asociación, aunque el cambio de entalpía asociado a las interacciones entre la proteína y el sustrato son favorables per se y del orden de  $-3 \text{ kcal}\cdot\text{mol}^{-1}$

-El análisis de las contribuciones de los restos individuales a la afinidad de unión indica la existencia de extensos contactos entre la proteasa y el sustrato, especialmente en la región correspondiente a la posición P5 del inhibidor, que contribuyen favorablemente a la entalpía de unión. De acuerdo con estos resultados, en los sustratos peptídicos las mayores contribuciones a la energética de unión proceden de los restos directamente adyacentes al enlace peptídico hidrolizable (P1, P1', y P2') así como de los situados en posición P5.



9.- Con el fin de identificar las características diferenciales de la unión de inhibidores no peptídicos con respecto a los sustratos, que pudieran ser responsables del desarrollo de resistencia frente a los distintos fármacos.

-La afinidad de unión de los inhibidores no peptídicos se encuentra en el intervalo nanomolar, mientras que la correspondiente a los sustratos es del orden micromolar. Dado que las contribuciones globales de la entalpía de unión son desfavorables, y de magnitud similar. La única contribución favorable es el cambio de entropía del disolvente asociado a la exclusión de superficies hidrofóbica del contacto con el disolvente, y que se encuentra parcialmente compensada por la disminución de la entropía conformacional y translacional como consecuencia del proceso de unión.

-La diferencia en afinidad por la enzima existente entre inhibidores y sustratos puede justificarse de acuerdo a su naturaleza. Así,, dado que los sustratos peptídicos poseen una mayor flexibilidad en disolución, su unión a la proteína implica la pérdida de un número significativo de grados de libertad, mientras que pasa los inhibidores orgánicos, que son moléculas rígidas de pequeño tamaño, el proceso de asociación no supone una gran penalización entrópica.

-A la luz de estos resultados ha sido posible proponer una hipótesis plausible sobre el origen molecular de la resistencia.. La elevada flexibilidad de los sustratos les permite adaptar su conformación en el complejo a los sutiles cambios provocados por las mutaciones en la estructura de la enzima, mientras que estos inhibidores de marcada rigidez no pueden acomodarse a las distorsiones experimentadas por el sitio de unión. De este modo, la aparición de una mutación en la enzima puede disminuir drásticamente la afinidad de unión de los inhibidores, manteniendo la actividad catalítica a niveles aceptables para la supervivencia del virus.

## **VII.-BIBLIOGRAFÍA**



## BIBLIOGRAFÍA:

- Abad-Zapatero, C., Rydel, T., & Erickson, J. (1990) *Proteins* 8, 62-81.
- Abagyan, R., Totrov, M., & Kuznetsov, D. (1994) *J Comput Chem* 15, 488-506.
- Abdel-Meguid, S., Metcalf, B., Carr, T., Demarsh, P., DesJarlais, R., Fichse, S., Green, D., Ivanoff, L., Lambert, D., Murthy, K., Petteway, S., Pitts, W., Tomaszek, T., Winborne, E., Zhao, B., Deryer, G., & Meek, T. (1994) *Biochemistry* 33, 11671-77.
- Altmann, K., Wojcik, J., Vasquez, M., & Scheraga, H. (1990) *Biopolymers* 30, 107-20.
- Andreeva, N., Zdanov, A., Gustchina, A., & Fedorov, A. (1984) *J Biol Chem* 259, 11353-11365.
- Antonov, V. (1985) in *Aspartic proteases and their inhibitors*. (Kostka, V. e. B., Ed.) pp 263-320, Walter de Gruyter.
- Antonov, V., Ginodman, L., Kapitannikov, Y., Barshevskaya, T., Gurov, A., & Rumsh, L. (1978) *FEBS Lett* 88, 87-90.
- Antonov, V., Ginodman, L., Rumsh, L., Kapitannikov, Y., Barshevskaya, T., Yavashev, L., Gurova, A., & Volkova, L. (1981) *Eur J Biochem* 117, 195-200.
- Antonson, P., Kammerer, R., Schulthess, T., Hänisch, G., & Engel, J. (1995) *J Mol Biol* 250, 74-9.
- Åquist, J., Luecke, H., Quicho, F., & Warshel, A. (1991) *Proc Natl Acad Sci U.S.A* 88, 2026-30.
- Armstrong, K., & Baldwin, R. (1993) *Proc Natl Acad Sci* 90, 11337-40.
- Armstrong, K., Fairman, R., & Baldwin, R. (1993) *J Mol Biol* 230, 284-91.
- Aurora, R., Srinivasan, R., & Rose, G. (1994) *Science* 264, 1126-9.
- Babé, L., Pichuantes, S., & Craik, C. (1991) *Biochemistry* 30, 106-11.
- Babé, L., Rosé, J., & Craik, C. (1995) *Proc Natl Acad Sci* 92, 10069-73.
- Backbro, K., Lowgren, S., Osterlund, K., Atepo, J., Unge, T., Hulten, BonhamNM, SchaalW, Karlen, A., & Hallberg, A. (1997) *J Med Chem* 40, 898-902.
- Badasso, M. (1994) , University of London, UK.
- Bailey, D., & Cooper, J. (1994) *Protein Sci* 3, 2129-43.
- Bailey, D., Cooper, J., Veerapandian, B., Blundell, T., Atrash, B., Jones, D., & Szelke, M. (1993) *Biochem J* 289 ( Pt 2), 363-71.
- Balbaa, M., Cunningham, A., & Hofmann, T. (1993) *Arch Biochem Biophys* 306, 297-303.
- Baldwin, E., Bhat, T., Gulnik, S., Liu, B., Topol, I., Kiso, Y., Minoto, T., Mitsuya, H., & Erickson, J. (1995a) *Structure*, 581-90.
- Baldwin, E., Bhat, T., Liu, B., Pattabiraman, N., & Erickson, J. (1995b) *Nat Struct Biol* 2, 244-9.
- Baldwin, R. (1986) *Proc Natl Acad Sci* 83, 8069-72.
- Baldwin, R. (1995) *Biophys Chem* 55, 127-35.
- Bardi, J., Luque, I., & Freire, E. (1997) *Biochemistry* 36, 6588-96.

- Barkholt, V. (1987) *Eur J Biochem* 167, 327-38.
- Barrish, J., Gordon, E., Alam, M., Lin, P., Bisacchi, G., Chen, P., Cheng, P., Fritz, A., Greytok, J., & Hermsmeier, M. (1994) *J Med Chem* 37, 1758-68.
- Bell, J., Bechtel, W., Sauer, U., Baase, W., & Matthews, B. (1992) *Biochemistry* 31, 3590-96.
- Ben-Naim, A. (1994) *Current Opinion in Structural Biology* 4, 264-8.
- Benson, S. (1968) *Thermochemical Kinetics. Methods for the estimation of thermochemical data and rate parameters*, John Wiley and Sons, New York.
- Berendsen, H. (1991) *Curr Op Struct Biol* 1, 191-5.
- Beveridge, A., & Heywood, G. (1993) *Biochemistry* 32, 3325-33.
- Beveridge, D., & DiCapua, F. (1989) *Ann Rev Biophys Biophys Chem* 18, 431-92.
- Bierzynski, A., Kim, P., & Baldwin, R. (1982) *Proc Natl Acad Sci USA* 79, 2470-4.
- Billich, A., & Winkler, G. (1991) *Arch Biochem Biophys* 290, 186.
- Billich, S., Knoop, M., Hansen, J., Strop, P., Sedlacek, J., Mertz, R., & K, M. (1988) *J Biol Chem* 263, 17905-17908.
- Blaber, M., Zhang, X., Lindstrom, J., Pepiot, S., Baase, W., & Matthews, B. (1994) *J Mol Biol* 235, 600-24.
- Blaber, M., Zhang, X., & Matthews, B. (1993) *Science* 260, 1637-40.
- Blinder, S. (1975) *J Am Chem Soc* 97, 978-81.
- Bliznyuk, A., & Gready, J. (1996) *J Comput Chem* 17, 962-69.
- Blundell, T., Cooper, J., Foundling, S., Jones, D., Atrash, B., & Szelke, M. (1987) *Biochemistry* 26, 5585-5590.
- Blundell, T., Jenkins, J., Pearl, L., Sewell, T., & Pederson, V. (1985) in *Aspartic proteinases and their inhibitors* (Kotska, V. e., Ed.) pp 151-161, Walter de Gruyter, Berlin.
- Blundell, T., Jenkins, J., Sewell, B., Pearl, L., Cooper, J., Tickle, I., Veerapandian, B., & Wood, S. (1990) *J Mol Biol* 211, 919-41.
- Blundell, T., Sewell, B., & McLachlan, A. (1979) *Biochim Biophys Acta* 580, 24-31.
- Bondi. (1964) *J Phys Chem* 68, 441-51.
- Bone, R., Vacca, J., Anderson, P., & Holloway, M. (1991) *J Am Chem Soc* 113, 9382-84.
- Bott, R., Subramanian, E., & Davies, D. (1982) *Biochemistry* 21, 6956-62.
- Bradley, E., Thomason, J., Cohen, F., Kosen, P., & Kuntz, I. (1990) *J Mol Biol* 215, 607-22.
- Brown, J., & Klee, W. (1971) *Biochemistry* 10, 470-6.
- Bruch, M., Dhingra, M., & Gierasch, L. (1991) *Proteins: Struct, Funct and Genet* 10, 130-9.
- Cabani, S., Gianni, P., Mollica, V., & Lepori, L. (1981) *J Solution Chem* 10, 563-95.
- Cameron, C., Grinde, B., Jacques, P., Jentoft, J., Leis, J., Wlodawer, A., & Weber, I. (1993) *J Biol Chem* 268, 11711-20.

- Cameron, C., Ridky, T., Shulenin, S., Leis, J., Weber, I., Copeland, T., Wlodawer, A., Burstein, H., Bizub-Bender, D., & Skalka, A. (1994) *J Biol Chem* 269, 11170-77.
- Capaldi, R., & Vandekooi, G. (1972) *Proc Natl Acad Sci* 69, 930-932.
- Chakrabartty, A., & Baldwin, R. (1995) *Adv Prot Chem* 46, 141-76.
- Chakrabartty, A., Doig, A., & Baldwin, R. (1993a) *Proc Natl Acad Sci* 90, 11332-6.
- Chakrabartty, A., Kortemme, T., & Baldwin, R. (1994) *Protein Sci* 3, 843-52.
- Chakrabartty, A., Kortemme, T., Padmanabhan, S., & Baldwin, R. (1993b) *Biochemistry* 32, 5560-5.
- Chakrabartty, A., Schellman, J., & Baldwin, R. (1991) *Nature* 351, 586-8.
- Chen, Z., Li, Y., Schock, H., Hall, D., Chen, E., & Kuo, L. (1995) *J Biol Chem* 270, 21433-6.
- Cheng, Y., Yin, F., Foundling, S., Bomlstrom, D., & Kettner, C. (1990) *Proc Natl Acad Sci* 87, 9660-4.
- Chothia. (1975) *Nature* 254, 304-8.
- Chou, P., & Fasman, G. (1974) *Biochemistry* 13, 221-22.
- Chou, P., & Fasman, G. (1978) *Annu Rev Biochem* 47, 251-76.
- Collins, J., Burt, S., & Erickson, J. (1995) *Nat Struct Biol* 2, 334-8.
- Condra, J., Schleif, W., Blahy, O., Gabryelski, L., Graham, D., Quintero, J., Rhodes, A., Robbins, H., Roth, E., Shivaprakash, M., & et, a. (1995) *Nature* 374, 569-71.
- Connolly, M. (1983) *J Appl Cryst* 16, 548-58.
- Cooper, J., Foundling, S., Blundell, T., Arrowsmith, R., Harris, C., & Champness, J. (1988) in *Topics in medicinal chemistry*. (Leeming, P., Ed.) pp 308-313, R Soc Chem Special Publication.
- Cooper, J., Foundling, S., Blundell, T., Boger, J., Jupp, R., & Kay, J. (1989) *Biochemistry* 28, 8596-603.
- Cooper, J., Foundling, S., Hemmings, A., Blundell, T., Jones, D., Hallett, A., & Szelke, M. (1987) *Eur J Biochem* 169, 215-221.
- Cooper, J., Khan, G., Taylor, G., Tickle, I., & Blundell, T. (1990) *J Mol Biol* 214, 199-222.
- Cooper, J., Quail, W., Frazao, C., Foundling, S., Blundell, T., Humblet, C., Lunney, E., Lowther, W., & Dunn, B. (1992) *Biochemistry* 31, 8142-50.
- Cornell, W., Howard, A., & Kollman, P. (1991) *Curr Op Struc Biol* 1, 201-212.
- Creamer, T., & Rose, G. (1992) *Proc Natl Acad Sci* 89, 5937-41.
- Creamer, T., & Rose, G. (1994) *Proteins: Struct, Funct and Genet* 19, 85-97.
- Creamer, T., & Rose, G. (1995) *Protein Science* 4, 1305-14.
- Creamer, T., Srinivasan, R., & Rose, G. (1995) *Biochemistry* 34, 16245-50.
- Creamer, T., Srinivasan, R., & Rose, G. (1997) *Biochemistry* 36, 2832-35.

- Cushman, M., Golebiewski, W., McMahon, J., Buckheit, R., Clanton, D., Weislow, O., Haugwitz, R., Bader, J., Graham, L., & Rice, W. (1994) *J Med Chem* 37, 3040-50.
- Cushman, M., Golebiewski, W., Pommier, Y., Mazumder, A., Reymen, D., De, C. E., Graham, L., & Rice, W. (1995) *J Med Chem* 38, 443-52.
- D'Aquino, J., Gómez, J., Hilser, V., Lee, K., Amzel, L., & Freire, E. (1996) *Proteins: Struct, Funct and Genet.* 25, 143-56.
- Dann, J., Stammers, D., Harris, C., Arrowsmith, R., Davies, D., Hardy, G., & Morton, J. (1986) *Biochem Biophys Res Commun* 134, 71-7.
- Darke, P., Leu, C., Davis, L., Heimbach, J., Diehl, R., Hill, W., Dixon, R., & Sigal, I. (1989) *J Biol Chem* 264, 2307-2312.
- Darke, P., Nutt, R., Brady, S., Garsky, V., Ciccarone, T., Leu, C., Lumma, P., Friedinger, R., Veber, D., & Sigal, I. (1988) *Biochem Biophys Res Commun* 156, 297-303.
- Davies, D. (1990) *Annual Review of Biophysical Chemistry* 19, 189-215.
- Debouck, C., Gorniak, J., Strickler, J., Meek, T., Metcalf, B., & Rosenberg, M. (1987) *Proc Natl Acad Sci* 84, 8903-6.
- Debouck, C., & Metcalf, B. (1990) *Drug Development Research* 21, 1-17.
- Deslongchamps, P. (1975) *Tetraedron* 31, 2463-2490.
- Dhanaraj, V., DeAlwis, C., Frzao, C., Badasso, M., Sibanda, B., Tickle, I., Cooper, J., Driessen, H., Newman, M., Aguilar, C., Wood, S., Blundell, T., Hobart, P., Goehagan, K., Ammirati, M., Danley, D., O'Connor, B., & Hoover, D. (1992) *Nature* 357, 466-472.
- Dill, K., Bromberg, S., Yue, K., Fiebig, K., Yee, D., Thomas, P., & Chan, H. (1995) *Protein Science* 4, 561-602.
- Dodd, L., & Theodorou, D. (1991) *Mol Phys* 72, 1313-54.
- Doig, A., & Baldwin, R. (1995) *Protein Science* 4, 1325-36.
- Doig, A., Chakrabartty, A., Klingler, T., & Baldwin, R. (1994) *Biochemistry* 33, 3396-403.
- Dreyer, G., Lambert, D., Meek, T., Carr, T., Tomaszek, T., & al., e. (1992) *Biochemistry* 31, 6646-59.
- Dreyer, G., Metcalf, B., Tomaszek, T., Carr, T., Chandler, A., Hyland, L., Fakhoury, S., Magaard, V., Moore, M., Strickler, J., Debouck, C., & Meek, T. (1989) *Proc. Natl. Acad. Sci* 86, 9752-6.
- Dunn, B., Gustchina, A., Wlodawer, A., & Kay, J. (1994) *Methods in Enzymology* 241, 254-78.
- Dunn, B., Jimenez, M., Parten, B., Valler, M., Rolph, C., & Kay, J. (1986) *Biochem J* 237, 899-906.
- Dyson, H., & Wright, P. (1993) *Current Opinion in Structural Biology* 3, 60-5.
- Eftink, M., Anusiem, A., & Biltonen, R. (1983) *Biochemistry* 22, 3884-96.
- Eisenberg, D., & McLachlan, A. (1986) *Nature* 319, 199.
- Eisenberg, D., Weiss, R., & Terwilliger, T. (1982) *Nature* 299, 371-4.

- Eisenhaber, F., & Argos, P. (1993) *J Comput Chem* 14, 1272-80.
- Eisenhaber, F., Lijnzaad, P., Argos, P., Sanders, C., & Scharf, M. (1995) *J Comput Chem* 16, 273-84.
- el-Farrash, M., Kuroda, M., Kitazaki, T., Masuda, T., Kato, K., Hatanaka, M., & Harada, S. (1994) *J Virol* 68, 233-9.
- Erickson, J. (1993) *Perspectives in Drug Discovery and Design* 1, 109-28.
- Erickson, J., Neidhart, D., VanDrie, J., Dempf, D., Wang, X., Norbeck, D., Plattner, J., Rittenhouse, J., Turon, M., Wideburg, N., Kohlbrenner, W., Simmer, R., Helfrich, R., Paur, D., & Knigge, M. (1990) *Science* 257, 527-534.
- Fairman, R., Shoemaker, K., York, E., Stewart, J., & Baldwin, R. (1990) *Biophys Chem* 37, 107-19.
- Fassler, A., Grutter, R., Tintelnot-Blomley, M., Alteri, E., Bold, G., & Lang, M. (1993) *Bioorganic and medicinal chemistry letters* 3, 2837-42.
- Fauchère, J., & Pliska, V. (1983) *Eur J Med Chem Chim Ther* 18, 169-75.
- Filimonov, V., Pfeil, W., Tsalkova, T., & Privalov, P. (1978) *Biophys Biochem* 9, 117-22.
- Finkelstein, A., & Janin, J. (1989) *Protein Eng* 3, 1-3.
- Fitzgerald, P., McKeever, B., VanMiddlesworth, J., Springer, J., Heimbach, J., Leu, C., Herber, W., Dixon, R., & Darke, P. (1990) *J Biol Chem* 265, 14209-19.
- Forood, B., Feliciano, E., & Nambiar, K. (1993) *Proc Natl Acad Sci U S A* 90, 838-42.
- Foundling, S., Cooper, J., Watson, F., Cleasby, A., Pearl, L., Sibanda, B., Hemmings, A., Wood, S., Blundell, T., Valler, M., & et, a. (1987) *Nature* 327, 349-352.
- Freire, E. (1995) *Annu Rev Biophys Biomol Struct* 24, 141-65.
- Freire, E., Haynie, D., & Xie, D. (1993) *Proteins* 17, 111-23.
- Freire, E., & Murphy, K. (1991) *J Mol Biol* 222, 687-98.
- Freire, E., & Xie, D. (1994) *Biophys Chem* 51, 243-51.
- Fruton, J. (1970) *Adv Enzymol* 33, 401-433.
- Fruton, J. (1976) *Adv Enzymol* 46, 1-36.
- Gans, P., Lyu, P., Manning, M., Woody, R., & Kallenbach, N. (1991) *Biopolymers* 31, 1605-14.
- García-Moreno, B. (1995) *Methods in enzymology*. 259, 512-38.
- García-Moreno, B., Dwyer, J., Gittis, A., Lattman, E., Spencer, D., & Stites, W. (1997) *Biophysical Chemistry* 64, 211-24.
- Gelb, M., Svaren, H., & Abeles, R. (1985) *Biochemistry* 24, 1813-1817.
- Gelin, B. (1993) in *Computer simulation of biomolecular systems. Theoretical and experimental applications*. (van Gunsteren, W., Weiner, P., & Wilkinson, A., Eds.), ESCOM Science publishers BV, Leiden.

- Ghosh, A., Lee, H., Thompson, W., Culberson, C., Holloway, M., McKee, S., Munson, P., Duong, T., Smith, A., Darke, P., Zugay, J., Emini, E., Schleif, W., & Anderson, P. (1994) *J Med Chem* 37, 1177-88.
- Giam, C., & Boros, I. (1988) *J Biol Chem* 263, 14617-20.
- Gibbs, J., & DiMarzio, E. (1959) *J Chem Phys* 30, 271.
- Gibson, K., & Scheraga, H. (1987) *Mol Phys* 62, 1247-65.
- Gilson, M., Given, J., Bush, B., & McCammon, J. (1997) *Biophys J* 72, 1047-69.
- Gnabre, J., Brady, J., Clanton, D., Ito, Y., Dittmer, J., Bates, R., & Huang, R. (1995) *Proc Natl Acad Sci* 92, 11239-43.
- Goldblum, A. (1988) *Biochem Biophys Res Commun* 157, 450-6.
- Goldblum, A., Rayan, A., Fliess, A., & Glick, M. (1993) *J Chem Inf Comput Sci* 33, 270-4.
- Gómez, J., & Freire, E. (1995) *J Mol Biol* 252, 337-50.
- Gómez, J., Hilser, V., Xie, D., & Freire, E. (1995) *Proteins: Struct Funct and Genet* 22, 204-12.
- Gómez, J., Semo, N., & Freire, E. (1997) in *Structure and function of Aspartic Proteinases: Retroviral and Cellular Enzymes* (James, M., Ed.), Plenum, New York.
- Griffiths, J., Phylip, L., Konvalinka, J., Strop, P., Gustchina, A., Wlodawer, A., Davenport, R., Briggs, R., Dunn, B., & Kay, J. (1992a) *Biochemistry* 31, 5193-200.
- Griffiths, J., Phylip, L., Konvalinka, J., Strop, P., Gustchina, A., Wlodawer, A., Davenport, R., Briggs, R., Dunn, B., & Kay, J. (1992b) *Biochemistry* 31, 5139.
- Griko, Y., Freire, E., Privalov, G., Van Dael, H., & Privalov, P. (1995) *J Mol Biol* 252.
- Griko, Y., Freire, E., & Privalov, P. (1994a) *Biochemistry* 33, 1889-99.
- Griko, Y., Makhatadze, G., Privalov, P., & Hartley, R. (1994b) *Protein Science* 3, 669-76.
- Griko, Y., Privalov, P., Sturtevant, J., & Venyaminov, S. (1988) *Proc Natl Acad Sci* 85, 3343-7.
- Grinde, B., Cameron, C., Leis, J., Weber, I., Wlodawer, A., Burstein, H., Bizub, D., & Skalka, A. (1992a) *J Biol Chem* 267, 9481-90.
- Grinde, B., Cameron, C., Leis, J., Weber, I., Wlodawer, A., Burstein, H., & Skalka, A. (1992b) *J Biol Mol* 267, 9491-8.
- Gulnik, S., Suvorov, L., Liu, B., Yu, B., Anderson, B., Mitsuya, H., & Erickson, J. (1995) *Biochemistry* 34, 9282-7.
- Hallet, A., Jones, D., Atrash, B., Szelke, M., Leckie, B., Beattie, S., Dunn, B., Valler, M., Rolph, C., Kay, J., Foundling, S., Wood, S., Pearl, L., Watson, F., & Blundell, T. (1985) in *Aspartic proteinases and their inhibitors*. (Kotska, V., Ed.) pp 467-478, Walter de Gruyter, Berlin.
- Harper, E., & Rose, G. (1993) *Biochemistry* 32, 7605-9.
- Harte, W., & Beveridge, D. (1993) *J Am Chem Soc* 115, 3883-6.



- Henderson, L., Benveniste, R., Sowder, R., Copeland, T., Schultz, A., & Oroszlan, S. (1988) *J Virol* 62, 2587.
- Hermans, J., Anderson, A., & Yun, R. (1992) *Biochemistry* 31, 5646-53.
- Hilser, V., & Freire, E. (1996a) *Proteins: Structure, Function and Genetics*.
- Hilser, V., & Freire, E. (1996b) *J Mol Biol* 262, 756-72.
- Hilser, V., Gómez, J., & Freire, E. (1996) *Proteins: Struct, Funct and Genet.* 26, 123-33.
- Hilser, V., Townsend, B., & Freire, E. (1997) *Biophys Chem* 64, 69-79.
- Hirst, J., & Sternberg, M. (1992) *Biochemistry* 31, 7211-8.
- Ho, D., Toyoshima, T., Mo, H., Kempf, D., Norbeck, D., Chen, C., Wideburg, N., Burt, S., Erickson, J., & Singh, M. (1994) *J Virol* 68, 2016-20.
- Hofmann, T., Allen, B., Bendiner, A., Blum, M., & Cunningham, A. (1988) *Biochemistry* 27, 1140-1146.
- Hofmann, T., Hodges, R., & James, M. (1984) *Biochemistry* 23, 635-43.
- Hol, W., van Duijnen, P., & Berendsen, H. (1978) *Nature* 273, 443-6.
- Holley, L., & Karplus, M. (1989) *Proc Natl Acad Sci USA* 86, 152-6.
- Hong, L., Treharne, A., Hartsuck, J., Foundling, S., & Tang, J. (1996) *Biochemistry* 35, 10627-33.
- Hoog, S., Towler, E., Zhao, B., Doyle, M., Debouck, C., & Abdel-Meguid, S. (1996) *Biochemistry* 35, 10279-86.
- Hoog, S., Zhao, B., Winborne, E., Fisher, S., Green, D., DesJarlais, R., Newlander, K., Callahan, J., Moore, M., Huffman, W., & et, a. (1995) *J Med Chem* 38, 3246-52.
- Horovitz, A., Matthews, J., & Fersht, A. (1992) *J Mol Biol* 227, 560-8.
- Hsu, I., Delbaere, L., James, M., & Hofmann, T. (1977) *Nature* 266, 140-145.
- Hughson, F., Barrick, D., & Baldwin, R. (1991) *Biochemistry* 30, 4113-8.
- Hui, J., Tomasselli, A., Zurcher-Neely, H., & Heinrickson, R. (1990) *J Biol Chem* 265, 21386-89.
- Hulten, J., Bonham, N., Nillroth, U., Hansson, T., Zuccarello, G., Bouzide, A., Aqvist, J., Classon, B., Danielson, U., Karlen, A., Kvarnstrom, I., Samuelsson, B., & Hallberg, A. (1997) *J Med Chem* 40, 885-97.
- Huyghues-Despointes, B., Scholtz, J., & Baldwin, R. (1993) *Protein Science* 2, 80-5.
- Hyland, L., Tomaszek, T., & Meek, T. (1991a) *Biochemistry* 30, 8454-63.
- Hyland, L., Tomaszek, T., Roberts, G., Carr, S., Maggaard, V., Bryan, H., Fakhoury, S., Moore, M., Minnich, M., Culp, J., DesJarlais, R., & Meek, T. (1991b) *Biochemistry* 30, 8441-53.
- Ido, E., Han, H., Kezdy, F., & Tang, J. (1991) *J Biol Chem* 266, 24358-66.
- Ihara, S., Ooi, T., & Takahashi, S. (1982) *Biopolymers* 21, 131-45.
- Iijima, H., Dunbar, J., & Marshall, G. (1987) *Proteins: structure, function and genetics* 2, 330-9.
- Iliadis, G., Brzezinski, B., & Zundel, G. (1996) *Biophys J* 71, 2840-2847.

- Iliadis, G., Zundel, G., & Brzezinski, B. (1994) *FEBS Lett* 352, 315-7.
- IUPAC-IUB. (1970) *J Biol Chem* 245, 6489.
- Jacobsen, H., Yasargil, K., Winslow, D., Craig, J., Krohn, A., Duncan, I., & Mous, J. (1995) *Virology* 206, 527-34.
- James, M., & Sielecki, A. (1983) *J Mol Biol* 163, 299-361.
- James, M., & Sielecki, A. (1985) *Biochemistry* 24, 3701-13.
- James, M., & Sielecki, A. (1986) *Nature* 319, 33-38.
- James, M., Sielecki, A., Hayakawa, K., & Gelb, M. (1992) *Biochemistry* 31, 3872-3886.
- James, M., Sielecki, A., & Hofman, R. (1985) in *Aspartic proteases and their inhibitors* (Kostka, V., Ed.) pp 163-177, Walter de Gruyter, Berlin.
- James, M., Sielecki, A., Salituro, F., Rich, D., & Hofmann, T. (1982) *Proc Nat Acad Sci* 79, 6137-6141.
- Janin, J. (1995) *Proteins: Structure, Function and Genetics* 21, 30-9.
- Janin, J., Wodak, S., Levitt, M., & Maigrett, B. (1978) *J Mol Biol* 125, 357-86.
- Janz, G. (1958) *Estimation of thermodynamic properties of organic compounds*, Academic Press, New York.
- Jaskolksi, M., Tomasselli, A., Sawyer, T., Staples, D., Heinrichson, R., & al., e. (1991) *Biochemistry* 30, 1600-9.
- Jenkins, J., Blundell, T., Tickle, I., & Ungaretti, L. (1975) *J Mol Biol* 99, 583-590.
- Jiménez, M., Bruix, M., González, C., Blanco, F., Nieto, J., Herranz, J., & Rico, M. (1993) *Eur J Biochem* 211, 569-81.
- Jiménez, M., Muñoz, V., Rico, M., & Serrano, L. (1994) *J Mol Biol* 242, 487-96.
- Jorgensen, W., & Swenson, C. (1985) *J Am Chem Soc* 107, 1489096.
- Jorgensen, W., & Tirado-Rives, J. (1988) *J Am Chem Soc* 110, 1657-66.
- Jupp, R., Dunn, B., Jacobs, J., Vlasuk, G., Arcuri, K., Veber, D., Perlow, E., Payne, L., Boger, J., De Laszlo, S., Chakravarty, P., TenBroeke, J., Hangauer, D., Ondeyka, K., Greenlee, W., & Kay, J. (1989) *Biochem J* 265.
- Kabsch, W., & Sander, C. (1983) *Biopolymers* 22, 2577-637.
- Kaplan, A., Michael, S., Wehbie, R., Knigge, M., Paul, D., Everitt, L., Kempf, D., Norbeck, D., Erickson, J., & Swanstrom, R. (1994) *Proc Natl Acad Sci U S A* 91, 5597-601.
- Karfunkel, H., & Eyraud, V. (1989) *J Comp Chem* 10, 628.
- Karplus, M., & Weaver, D. (1994) *Protein Science* 3, 650-68.
- Katoh, I., Ikawa, Y., & Yoshinaka, Y. (1989) *J Virol* 63, 2226-32.
- Katoh, I., Yasunaga, T., Ikawa, Y., & Yoshinaka, Y. (1987) *Nature* 329, 654-6.
- Kauzmann, W. (1959) *Adv Protein Chem* 14, 1-63.
- Kemp, D., Boyd, J., & Muendel, C. (1991) *Nature* 352, 451-4.
- Kempf, D. (1994) *Methods Enzymol* 241, 334-54.

- Kempf, D., Codacovi, L., Wang, X., Kohlbrenner, W., Wideburg, N., Saldivar, A., Vasavanonda, S., Marsh, K., Bryant, P., Sham, H., & et, a. (1993) *J Med Chem* 36, 320-30.
- Kempf, D., Marsh, K., Denissen, J., McDonald, E., Vasavanonda, S., Flentge, C., Green, B., Fino, L., Park, C., Kong, X., & et, a. (1995) *Proc Natl Acad Sci U S A* 92, 2484-8.
- Kempf, D., Norbeck, D., Codacovi, L., Wang, X., Kohlbrenner, W., Wideburg, N., Paul, D., Knigge, M., Vasavanonda, S., Craig-Kennard, A., Saldivar, A., Rosenbrook, W., Clement, J., Plattner, J., & Erickson, J. (1990) *J Med Chem* 33, 2687-9.
- Kempf, K., Marsh, K., Paul, k., Knigge, M., Norbeck, D., Kohlbrenner, W., Codacovi, L., Vasabanonda, S., Bryant, P., Wang, X., Wideburg, N., Clement, J., Plattner, J., & Erickson, J. (1991) *Antimicrobial Agents and Chemotherapy* 35, 2209-14.
- Kervinen, J., Thanki, N., Zdanov, A., Tino, J., Barrish, J., Lin, P., Colonna, R., Riccadi, K., Samanta, H., & Wlodawer, A. (1996) *Protein Pept Lett* 2, 399.
- Khechinashvili, N., Privalov, P., & Tiktopulo, E. (1973) *FEBS Lett* 30, 57-60.
- Kim, E., Baker, C., Dwyer, M., Murcko, M., Rao, B., Tung, R., & Navia, M. (1995) *J Am Chem Soc* 117, 1181-82.
- Kim, P., & Baldwin, R. (1990) *Annu Rev Biochem* 59, 631-60.
- King, P. (1993) in *Computer simulation of biomolecular systems. Theoretical and experimental applications.* (van Gunsteren, W., Weiner, P., & Wilkinson, A., Eds.) pp 267-314, ESCOM Science publishers BV, Leiden.
- King, R., Winslow, D., Garber, S., Scarnati, H., Bachelor, L., Stack, S., & Otto, M. (1995) *Antiviral Research* 28, 13-24.
- Kitchin, J., Bethell, R., Cammack, N., Dolan, S., Evans, D., Holman, S., Holmes, D., McMeekin, P., Mo, C., Nieland, N., & et, a. (1994) *J Med Chem* 37, 3707-16.
- Kneller, D., Cohen, F., & Langridge, R. (1990) *J Mol Biol* 214, 171-182.
- Kohl, N., Emini, E., Schleif, W., Davis, L., Heimbach, J., Dixon, R., Scolnick, E., & Sigal, I. (1988) *Proc Natl Acad Sci* 85, 4686-4690.
- Konvalinka, J., Strop, P., Velek, J., Cerna, V., Kostka, V., Phylip, L., Richards, A., Dunn, B., & Kay, J. (1990) *FEBS Lett* 268, 35.
- Krausslich, J., Ingraham, R., Skoog, M., Wimmer, E., Pallai, P., & Carter, C. (1989) *PNAS* 86, 807-11.
- Krohn, A., Redshaw, S., Ritchie, J., Graves, B., & Hatada, M. (1991) *J Med Chem* 34, 3340-2.
- Kuwajima, K., & Baldwin, R. (1983) *J Mol Biol* 169, 299-323.
- Lam, P., Jadhav, P., Eyermann, C., Hodge, C., Ru, Y., Bachelor, L., Meek, J., Otto, M., Rayner, M., Wong, Y., & et, a. (1994) *Science* 263, 380-4.
- Lapatto, R., Blundell, T., Hemmings, A., Overington, J., Wilderspin, A., Wood, S., Merson, J., Whittle, P., Danley, D., Geoghegan, K., Hawrylik, S., Lee, S., Scheld, K., & Hobart, P. (1989) *Nature* 342, 299-302.

- Lazaridis, T., Archontis, G., & Karplus, M. (1995) *Adv Prot Chem* 47, 231-96.
- Leach, S., Nemethy, G., & Scheraga, H. (1966) *Biopolymers* 4, 369-407.
- Lee, B., & Richards, F. (1971) *J Mol Biol* 55, 379-400.
- Lee, K., Xie, D., Freire, E., & Amzel, M. (1994) *Proteins: Struct, Funct and Genet.* 20, 68-84.
- LeGrand, S., & Kenneth Jr, M. (1993) *J Comput Chem* 14, 349-52.
- Lesser, G., & Rose, G. (1990) *Proteins: Struct Funct and Genet* 8, 6-13.
- Levitt, M. (1974) *J Mol Biol* 82, 393-420.
- Levitt, M. (1978) *Biochemistry* 17, 4277-85.
- Levitt, M., & Greer, J. (1977) *J Mol Biol* 114, 181-293.
- Liff, M., Lyu, P., & Kallenbach, N. (1991) *J Am Chem Soc* 113, 1014-9.
- Lifson, S., & Roig, A. (1961) *J Chem Phys* 34, 1963-74.
- Lin, L., Pinker, R., Phillips, G., & Kallenbach, N. (1994) *Protein Science* 3, 1430-5.
- Lin, Y., Fusek, M., Lin, X., Hartsuck, J., Kezdy, F., & Tang, J. (1992) *J Biol Chem* 267, 18413-8.
- Lin, Y., Lin, X., Hong, L., Foundling, S., Heinrikson, R., Thaisrivongs, S., Leelamanit, W., Raterman, D., Shah, M., Dunn, B., & Tang, J. (1995) *Biochemistry* 34, 1143-52.
- Livingstone, J., Spolar, R., & Record Jr, M. (1991) *Biochemistry* 30, 4237-44.
- Lockhart, D., & Kim, P. (1993) *Science* 260, 198-202.
- Loeb, D., Hutchison III, C., Edgell, M., Farmerie, W., & Swanstrom, R. (1989a) *J. Virol.* 63, 111-121.
- Loeb, D., Swanstrom, R., Everitt, L., Manchester, M., Stamper, S., & Hutchison III, C. (1989b) *Nature* 340, 397-400.
- Lunney, E., Hagen, S., Domagala, J., Humblet, C., Kosinski, J., Tait, B., Warmus, J., Wilson, M., Ferguson, D., Hupe, D., & et, a. (1994) *J Med Chem* 37, 2664-77.
- Lunney, E., Hamilton, H., Hodges, J., Kaltenbronn, J., Repine, J., Badasso, M., Cooper, J., Dealwis, C., Wallace, B., Lowther, W., & et, a. (1993) *J Med Chem* 36, 3809-20.
- Luque, I., & Freire, E. (1998) *Methods in Enzymology in press.*
- Luque, I., Gómez, J., Semo, N., & Freire, E. (1997) *Proteins: Structure, Function and Genetics* 29, 1-12.
- Luque, I., Mayorga, O., & Freire, E. (1996) *Biochemistry* 35, 13681-8.
- Luque, I., Todd, M., Gómez, J., Semo, N., & Freire, E. (1998) *Biochemistry in press.*
- Lyu, P., Liff, M., Marky, L., & Kallenbach, N. (1990) *Science* 250, 669-73.
- Lyu, P., Marky, L., & Kallenbach, N. (1989) *J Amer Chem Soc* 111, 2733-4.
- Lyu, P., Wemmer, D., Zhou, H., Pinker, R., & Kallenbach, N. (1993) *Biochemistry* 32, 421-5.
- Madhusoodan, V., Hoseu, T., Narayana, B., Kempf, D., Baldwin, E., Liu, B., Gulnik, S., Wideburg, N., Norbeck, D., Appert, K., & Erickson, J. (1994) *J Am Chem Soc* 116, 847-55.

- Makhatadze, G., Clore, G., Gronenborn, A., & Privalov, P. (1994) *Biochemistry* 33, 9327-32.
- Makhatadze, G., Kim, K., Woodward, C., & Privalov, P. (1993) *Protein Science* 2, 2028-36.
- Makhatadze, G., & Privalov, P. (1993a) *J Mol Biol* 232, 639-59.
- Makhatadze, G., & Privalov, P. (1993b) *J Mol Biol* 232, 660-79.
- Makhatadze, G., & Privalov, P. (1995) *Adv Prot Chem* 47, 308-425.
- Margolin, N., Heath, W., Osborne, E., Lai, M., & Vlahos, C. (1990) *Biochem Biophys Res Commun* 167, 554.
- Markowitz, M., Mo, H., Kempf, D., Norbeck, D., Bhat, T., Erickson, J., & Ho, D. (1995) *J Virol* 69, 701-6.
- Marqusee, S., & Baldwin, R. (1987) *Proc Natl Acad Sci* 84, 8898-902.
- Marqusee, S., Robbins, V., & Baldwin, R. (1989) *Proc Natl Acad Sci* 86, 5286-90.
- Martin, J. (1992) *Antiviral Research* 17, 265-78.
- Martínez, J., El Harrou, M., Filimonov, V., Mateo, P., & Fersht, A. (1994) *Biochemistry* 33, 3919-26.
- Mateo, P., & Privalov, P. (1981) *FEBS Lett* 123, 189-92.
- Mathews, C. (1993) *Annu Rev Biochem* 62, 653-83.
- McCammion, J. (1991) *Curr Op Struct Biol* 1, 196-200.
- McGregor, M., Islam, S., & Sternberg, M. (1987) *J Mol Biol* 198, 295-310.
- McPhee, F., Good, A., Kuntz, I., & Craik, C. (1996) *Proc Natl Acad Sci* 93, 11477-81.
- McQuade, T., Tomasselli, A., Liu, L., Karacostas, V., Moss, B., Sawyer, T., Heinrikson, R., & Tarpley, W. (1990) *Proc Natl Acad Sci* 87, 7472.
- Meek, E., Dayton, B., Metcalf, B., Dreyer, G., Strickler, J., Gorniak, J., Rosenberg, M., Moore, M., Magaard, V., & Debouck, C. (1989) *Proc Natl Acad Sci* 86, 1841-5.
- Meek, T., Rodriguez, E., & Angeles, T. (1994) *Methods in Enzymology* 241, 127-56.
- Merutka, G., Lipton, W., Shalongo, W., Park, S., & Stellwagen, E. (1990) *Biochemistry* 29, 7511-5.
- Meyer, A. (1988) *J Comput Chem* 9, 18-24.
- Miller, M., Schneider, J., Sathyanarayana, B., Toth, M., Marshall, G., Clawson, L., Selk, L., Kent, S., & Wlodawer, A. (1989a) *Science* 246, 1149-52.
- Miller, N., Jaskolski, M., Mohana Rao, J., Leis, J., & Wlodawer, A. (1989b) *Nature* 337, 576-579.
- Miller, S., Janin, J., Lesk, A., & Chothia, C. (1987) *J Mol Biol* 196, 641-56.
- Mitchinson, C., & Baldwin, R. (1986) *Proteins: Structure, Function and Genetics* 1, 23-33.
- Moore, M., Bryan, W., Fakhoury, S., Magaard, V., Huffman, W., Dayton, B., Meek, T., Hyland, L., Dreyer, G., Metcalf, B., Strickler, J., Gorniak, J., & Debouck, C. (1989) *Biochem Biophys Res Commun* 159, 420-5.
- Moutouh, L., Corbeil, J., & Richman, D. (1996) *Proc Natl Acad Sci U S A* 93, 6106-11.

- Müller, J. (1983) *J Appl Cryst* 16, 74-82.
- Muñoz, V., Blanco, F., & Serrano, L. (1995a) *Protein Science* 4, 1577-86.
- Muñoz, V., Blanco, F., & Serrano, L. (1995b) *Nature Structural Biology*.
- Muñoz, V., & Serrano, L. (1994) *Proteins: Struct, Funct and Genet* 20, 301-11.
- Muñoz, V., & Serrano, L. (1995a) *Biochemistry* 34, 15301-6.
- Muñoz, V., & Serrano, L. (1995b) *J Mol Biol* 245, 275-96.
- Muñoz, V., & Serrano, L. (1995c) *J Mol Biol* 245, 297-308.
- Muñoz, V., & Serrano, L. (1995d) *Curr Opin Biotechnol* 6, 382-6.
- Murphy, K., Bhakuni, V., Xie, D., & Freire, E. (1992) *J Mol Biol* 227, 293-306.
- Murphy, K., & Freire, E. (1992) *Adv Protein Chem* 43, 313-61.
- Murphy, K., & Gill, S. (1990) *Thermochim Acta* 172, 11-20.
- Murphy, K., & Gill, S. (1991) *J Mol Biol* 222, 699-709.
- Murphy, K., Privalov, P., & Gill, S. (1990) *Science* 247, 559-61.
- Murphy, K., Xie, D., García, C., Amzel, M., & Freire, E. (1993) *Proteins: Struct, Funct, and Genet* 15, 113-20.
- Murphy, K., Xie, D., Thompson, K., Amzel, M., & Freire, E. (1994) *Proteins: Struct Funct and Genet*. 18, 63-7.
- Myers, J., & Pace, C. (1996) *Biophys J* 71, 2033-39.
- Myers, J., Pace, C., & Scholtz, J. (1997a) *Proc Natl Acad Sci U S A* 94, 2833-7.
- Myers, J., Pace, C., & Scholtz, J. (1997b) *Biochemistry* 36, 10923-9.
- Nagai, K. (1960) *J Phys Soic Jpn* 15, 407.
- Navia, M., Fitzgerald, P., Mckeever, B., Leu, C., Heimbach, J., Herber, W., Sigal, I., Darke, P., & Springer, J. (1989) *Nature* 337, 615-20.
- Newman, M., Safro, M., Frazao, C., Khan, G., Zdanov, A., Tickle, I., Blundell, T., & Andreeva, N. (1991) *J Mol Biol* 221, 1295-309.
- Newmark, A., & Knowles, J. (1975) *J Am Chem Soc* 97, 3557-3559.
- Novotny, J., Bruccoleri, R., & Saul, F. (1989) *Biochemistry* 28, 4735-49.
- Nozaki, Y., & Tanford, C. (1971) *J Biol Chem* 246, 2211-7.
- Nugiel, D., Jacobs, K., Cornelius, L., Chang, C., Jadhav, P., Holler, E., Klabe, R., Bacheler, L., Cordova, B., Garber, S., Reid, C., Logue, K., Gorey-Feret, L., Lam, G., Erickson-Viitanen, S., & Seitz, S. (1997) *J Med Chem* 40, 1465-74.
- O'Neil, K., & DeGrado, W. (1990) *Science* 250, 646-51.
- Okamoto, Y. (1994) *Proteins: Structure, Function, and Genetics* 19, 14-23.
- Oobatake, M., & Ooi, T. (1992) *Prog Biophys Mol Biol* 59, 237-84.
- Ooi, T., & Oobatake, M. (1991) *Proc Natl Acad Sci* 88, 2859-63.
- Ooi, T., Oobatake, M., Nemethy, G., & Scheraga, H. (1987) *Proc Natl Acad Sci* 84, 3086-90.
- Osterhout, J., Baldwin, R., York, E., Stewart, J., Dyson, H., & Wright, P. (1989) *Biochemistry* 28.



- Otto, M., Garber, S., Winslow, D., Reid, C., Aldrich, P., Jadhav, P., Patterson, C., Hodge, C., & Cheng, Y. (1993) *Proc Natl Acad Sci U S A* 90, 7543-7.
- Pace, N., & Scholtz, J. (1997) *Biophysical Journal*.
- Padmanabhan, S., & Baldwin, R. (1991) *J Mol Biol* 219, 135-7.
- Padmanabhan, S., & Baldwin, R. (1994) *J Mol Biol* 241, 706-13.
- Padmanabhan, S., Marqusee, S., Ridgeway, T., Laue, T., & Baldwin, R. (1990) *Nature* 344, 268-70.
- Padmanabhan, S., York, E., Gera, L., Stewart, J., & Baldwin, R. (1994) *Biochemistry* 33, 8604-9.
- Park, S., Shalongo, W., & Stellwagen, E. (1993a) *Biochemistry* 32, 12901-5.
- Park, S., Shalongo, W., & Stellwagen, E. (1993b) *Biochemistry* 32, 7048-53.
- Pascual-Ahuir, J., & Silla, E. (1991) *J Comp Chem* 12, 1047.
- Pascual-Ahuir, J., Silla, E., & Tuñón, I. (1994) *J Comput Chem* 15, 1127-38.
- Pauling, L., Corey, R., & Branson, H. (1951) *Proc Natl Acad Sci* 37, 205.
- Pavlov, M., & Fedorov, B. (1983) *Biopolymers* 22, 1507-22.
- Pearl, L. (1987) *FEBS Lett* 214, 8-12.
- Pearl, L., & Blundell, T. (1984) *FEBS Lett* 174, 96-101.
- Pearl, L., & Taylor, W. (1987) *Nature* 329, 351-4.
- Pechik, I., Gustchina, A., Andreeva, N., & Fedorov, A. (1989) *FEBS Lett* 247, 118-22.
- Peller, L. (1959) *J Phys Chem* 63, 1194.
- Perrot, G., Cheng, B., Gibson, K., Vila, J., Palmer, K., Nayeem, A., Maigre, B., & Scheraga, H. (1992) *J Comput Chem* 13, 1-11.
- Petitjean, M. (1994) *J Comput Chem* 15, 507-23.
- Pettit, S., Simsic, J., Loeb, D., Everitt, L., Hutchinson, C., & Swanstrom, R. (1991) *J Biol Chem* 266, 14539.
- Phylip, L., Richards, A., Kay, J., Konvalinka, J., Strop, P., Blaha, I., Velek, J., Dostka, V., Ritchie, A., Bradhurst, A., Farmerie, W., Scarborough, P., & Dunn, B. (1990) *Biochem Biophys Res Commun* 171, 439-444.
- Pinker, R., Lin, L., Rose, G., & Kallenbach, N. (1993) *Protein Science* 2, 1099-105.
- Podlogar, B., Farr, R., Friedrich, D., Tarnus, C., Huber, E., Cregge, R., & Schirlin, D. (1994) *J Med Chem* 37, 3684-92.
- Pohorille, A., & Pratt, L. (1986) *Methods in Enzymology* 127, 64-78.
- Polgar, L. (1987) *FEBS Lett* 219, 1-4.
- Polgar, L., Szeltner, Z., & Boros, I. (1994) *Biochemistry* 33, 9351-7.
- Poorman, R., Tomasselli, A., Heinrison, R., & Kezdy, F. (1991) *J Biol Chem* 266, 14554-64.
- Presta, L., & Rose, G. (1988) *Science* 240, 1632-41.

- Privalov, P., Griko, Y., Venyaminov, S., & Kutysenko, V. (1986) *J Mol Biol* 190, 487-98.
- Privalov, P., & Khechinashvili, N. (1974) *J Mol Biol* 86, 665-84.
- Privalov, P., & Makhatadze, G. (1990) *J Mol Biol* 213, 385-91.
- Privalov, P., & Makhatadze, G. (1992) *J Mol Biol* 224, 715-23.
- Privalov, P., & Makhatadze, G. (1993) *J Mol Biol* 232, 660-79.
- Privalov, P., Tiktopulo, E., & Khechinashvili, N. (1973) *Int J Pept Protein Res* 5, 229-237.
- Qian, H. (1993) *Biopolymers* 33, 1605-16.
- Qian, H. (1994) *Biophys J* 67, 349-55.
- Qian, H., & Schellman, J. (1992) *J Phys Chem* 96, 3987-94.
- Rajagopalan, T., Stein, W., & Moore, S. (1966) *J Biol Chem* 241, 4295-4297.
- Ramachandran, G., & Sasisekharan, V. (1968) *Adv Prot Chem* 23, 284-438.
- Rao, B., & Singh, U. (1991) *J Am Chem Soc* 113.
- Reid, K., Lindley, P., & Thornton, J. (1985) *FEBS Letters* 190, 209-13.
- Reynolds, C., King, P., & Richards, G. (1992) *Molecular Physics* 76, 251-75.
- Rich, D., Bernatowicz, M., Agarwal, N., Kawai, M., Salituro, F., & Schmidt, P. (1985) *Biochemistry* 24, 3165-73.
- Rich, D., Green, J., Toth, M., Marshall, G., & Kent, S. (1990) *J Med Chem* 33, 1285-8.
- Rich, D., Sun, C., Vara-Prasad, J., Pathiasseril, A., Toth, M., Marshall, G., Clare, M., Mueller, R., & Houseman, K. (1991) *J Med Chem* 34, 1222-5.
- Rich, D., Sun, E., & Singh, J. (1977) *Biochem Biophys Res Commun* 74, 762-7.
- Rich, D., Sun, E., & Ulm, E. (1980) *J Med Chem* 23, 27-33.
- Richards, A., Broadhurst, A., Ritchie, A., Dunn, B., & Kay, J. (1989a) *FEBS Lett* 253, 214-6.
- Richards, A., Phylip, L., Farmeries, W., Scarborough, P., Alvarez, A., Dunn, B., Hirel, P., Konvalinka, J., Strop, P., Pavlickova, L., Kostka, V., & Kay, J. (1990) *J Biol Chem* 265, 7733-6.
- Richards, A., Roberts, R., Dunn, B., Graves, M., & Kay, J. (1989b) *FEBS Lett* 247, 113-117.
- Richards, F. (1977) *Ann Rev Biophys Bioeng* 6, 151-76.
- Richards, F. (1985) *Methods in Enzymology* 115, 440-64.
- Richardson, J. (1981) *Adv Protein Chem* 34, 167-338.
- Richardson, J., & Richardson, D. (1988) *Science* 240, 1648-52.
- Richmond, T. (1984) *J Mol Biol* 178, 63-89.
- Richmond, T., & Richards, F. (1978) *J Mol Biol* 119, 537-55.
- Ridky, T., & Leis, J. (1995) *J Biol Chem* 270, 29621-3.
- Roberts, N. (1995) *AIDS* 9 Suppl 2, 27-S32.

- Roberts, N., Martin, J., Kinchington, D., Bradhurst, A., Craig, J., Duncan, I., Galpin, S., Handa, B., Kay, J., Krohn, A., Lamber, R., Merrett, J., Mills, J., Parkes, K., Reshaw, S., Ritchie, A., Taylor, D., Thomas, G., & Machin, P. (1990) *Science* 248, 358-361.
- Rohl, C., Chakrabartty, A., & Baldwin, R. (1996) *Protein Science* 5, 2623-37.
- Rohl, C., Scholtz, J., York, E., Stewart, J., & Baldwin, R. (1992) *Biochemistry* 31, 1263-9.
- Romines, K., Watenpaugh, K., Howe, W., Tomich, P., Lovasz, K., Morris, J., Janakiraman, M., Lynn, J., Horng, M., Chong, K., & et, a. (1995a) *J Med Chem* 38, 4463-73.
- Romines, K., Watenpaugh, K., Tomich, P., Howe, W., Morris, J., Lovasz, K., Mulichak, A., Finzel, B., Lynn, J., Horng, M., & et, a. (1995b) *J Med Chem* 38, 1884-91.
- Rose, G., Geselowitz, A., Lesser, G., Lee, R., & Zehfus, M. (1985) *Science* 229, 834.
- Rose, R., Craik, C., Douglas, N., & Stroud, R. (1996a) *Biochemistry* 35, 12933-44.
- Rose, R., Gong, Y., Greytok, J., Bechtold, C., Terry, B., Robinson, B., Alam, M., Colonno, R., & Lin, P. (1996b) *Proc Natl Acad Sci U S A* 93, 1648-53.
- Sali, A., Veerapandian, B., Cooper, J., Foundling, S., Hoover, D., & Blundell, T. (1989) *EMBO J* 8, 2179-88.
- Sampath-Kumar, P., & Fruton, J. (1974) *Proc Natl Acad Sci* 71, 1070-1072.
- Sancho, J., Neira, J., & Fersht, A. (1992) *J Mol Biol* 224, 749-58.
- Schechter, I., & Berger, A. (1967) *Biochem Biophys Res Commun* 27, 157-162.
- Schellman, J. (1955a) *C. R. Trav. Lab. Carlsberg, Ser. Chim.* 29, 230-259.
- Schellman, J. (1955b) *C. R. Trav. Lab. Carlsberg, Ser. Chim.* 29, 223-229.
- Schellman, J. (1958) *J Phys Chem* 62, helices  
non two states.
- Schirling, D., Baltzer, S., Van Dorsselaer, V., Weber, F., Weill, C., Altenburger, J., Neises, B., Flynn, G., Remy, J., & Tarnus, C. (1993) *Biomed Chem Lett* 3, 253.
- Schmidt, P., Holladay, M., Salituro, F., & Rich, D. (1985) *Biochem Biophys Res Commun* 129, 597-602.
- Schneider, J., & Kent, S. (1988) *Cell* 54, 363-8.
- Scholtz, J., Marqusee, S., Baldwin, R., York, E., Stewart, J., Santoro, M., & Bolen, D. (1991) *Proc Natl Acad Sci U S A* 88, 2854-8.
- Scholtz, J., Qian, H., Robbins, V., & Baldwin, R. (1993) *Biochemistry* 32, 9668-76.
- Seale, J., Srinivasan, R., & Rose, G. (1994) *Protein Science* 3, 1741-5.
- Searle, M., Williams, D., & Gerhard, U. (1992) *J Am Chem Soc* 114, 10697-704.
- Serrano, L., & Fersht, A. (1989) *Nature* 342, 196-9.
- Serrano, L., Neira, J., Sancho, J., & Fersht, A. (1992a) *Nature* 356, 453-5.
- Serrano, L., Sancho, J., Hirshberg, M., & Fersht, A. (1992b) *J Mol Biol* 227, 544-59.
- Shalongo, W., & Stellwagen, E. (1995) *Protein Science* 4, 1161-6.

- Sham, H., Zhao, C., Marsh, K., Betebenner, D., Lin, S., Rosenbrook, W., Herrin, T., Li, L., Madigan, D., Vasavanonda, S., Molla, A., Saldivar, A., McDonald, E., Wideburg, N., Kempf, D., Norbeck, D., & Plattner, J. (1996) *Biochem Biophys Res Commun* 225, 436-40.
- Sharp, K. (1993) in *Computer simulation of biomolecular systems. Theoretical and experimental applications*. (van Gunsteren, W., Weiner, P., & Wilkinson, A., Eds.), ESCOM Science publishers BV, Leiden.
- Sheridan, R., Levy, R., & Salemme, F. (1982) *Proc Natl Acad Sci* 79, 4545-9.
- Shoemaker, K., Fairman, R., Schultz, D., Robertson, A., York, E., Stewart, J., & Baldwin, R. (1990) *Biopolymers* 29, 1-11.
- Shoemaker, K., Kim, P., Brems, D., Marqusee, S., York, E., Chaiken, I., Stewart, J., & Baldwin, R. (1985) *Proc Natl Acad Sci* 82, 2349-53.
- Shoemaker, K., Kim, P., York, E., Stewart, J., & Baldwin, R. (1987) *Nature* 326, 563-7.
- Shoeman, R., Honer, B., Stoller, T., Kesselmeier, C., Miedel, M., Traub, P., & Graves, M. (1990) *Proc Natl Acad Sci* 87, 6336-40.
- Shrake, A., & Rupley, J. (1973) *J Mol Biol* 79, 351-71.
- Sielecki, A., Hayakawa, K., Fujinaga, M., Murphy, M., Fraser, M., Muir, A., Carilli, C., Lewicki, J., Baxter, J., & James, M. (1989) *Science* 243, 1346-1351.
- Siepmann, J. (1993) in *Computer simulation of biomolecular systems. Theoretical and experimental applications*. (van Gunsteren, W., Weiner, P., & Wilkinson, A., Eds.), ESCOM Science publishers BV, Leiden.
- Silla, E., Tuñón, I., & Pascual-Ahuir, J. (1991) *J Comput Chem* 12, 1077-88.
- Silva, A., Cachau, R., Sham, H., & Erickson, J. (1996) *J Mol Biol* 255, 321-346.
- Skulnick, H., Johnson, P., Aristoff, P., Morris, J., Lovasz, K., Howe, W., Watenpaugh, K., Janakiraman, M., Anderson, D., Reischer, R., Schwartz, T., Banitt, L., Tomich, P., Lynn, J., Horng, M., Chong, K., Hinshaw, R., Dolak, L., Seest, E., Schwende, F., Rush, B., Howard, G., Toth, L., Wilkinson, K., Romines, K., & et, a. (1997) *J Med Chem* 40, 1149-64.
- Skulnick, H., Johnson, P., Howe, W., Tomich, P., Chong, K., Watenpaugh, K., Janakiraman, M., Dolak, L., McGrath, J., Lynn, J., & et, a. (1995) *J Med Chem* 38, 4968-71.
- Smith, R., Brereton, I., Chai, R., & Kent, S. (1996) *Nat Struct Biol* 3, 946-50.
- Sneddon, S., & Tobias, D. (1992) *Biochemistry* 31, 3947-55.
- Spinelli, S., Liu, Q., Alzari, P., Hirel, P., & Poljak, R. (1991) *Biochimie* 73, 1391-96.
- Spolar, R., Livingston, J., & Record Jr, M. (1992) *J Biol Chem* 267, 3947-55.
- Spolar, R., & Record Jr, M. (1994) *Science* 263, 777-84.
- Stellwagen, E., Park, S., Shalongo, W., & Jain, A. (1992) *Biopolymers* 32, 1193-200.
- Straatsma, T., & McCammon, J. (1992) *Annu Rev Phys Chem* 43, 407-35.

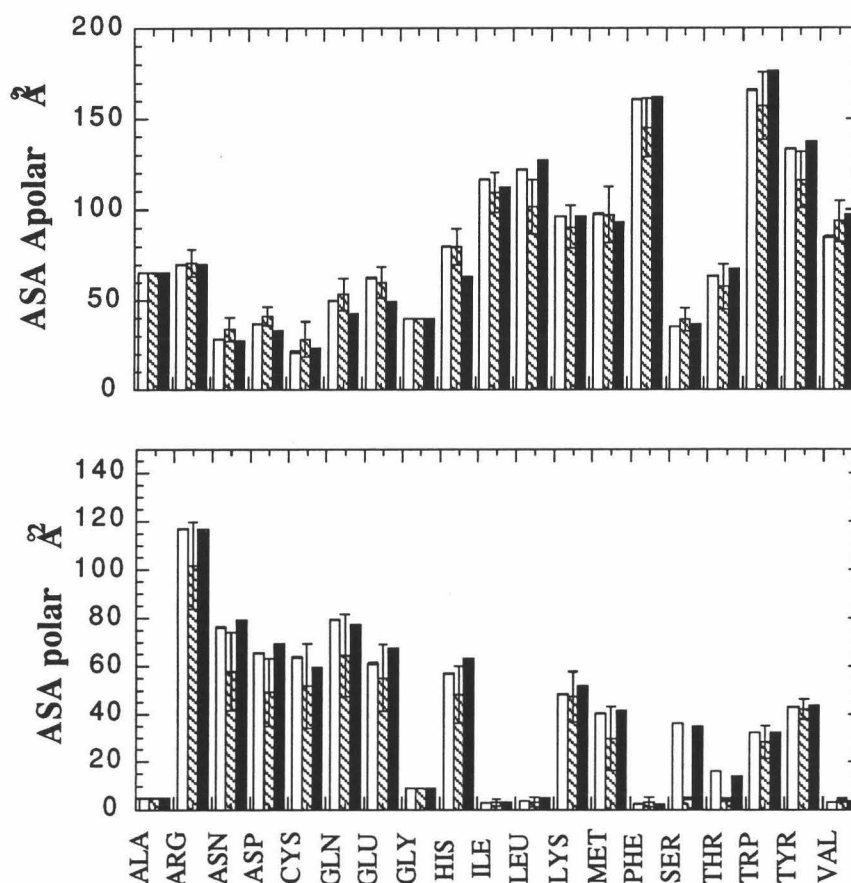
- Straatsma, T., Zacharias, M., & McCammon, J. (1993) in *Computer simulation of biomolecular systems. Theoretical and experimental applications.* (van Gunsteren, W., Weiner, P., & Wilkinson, A., Eds.), ESCOM Science publishers BV, Leiden.
- Strickler, J., Gorniak, J., Dayton, B., Meek, T., Moore, M., Magaard, V., Malinowski, J., & Debouk, C. (1989) *Proteins: Struct Funct and Gen* 6, 139-54.
- Subramanian, E., Swan, I., Liu, M., Davies, D., Jenkins, J., Ticle, I., & Blundell, T. (1977) *Proc Natl Acad Sci* 74, 556-559.
- Sueki, M., Lee, S., Powers, S., Denton, J., Konishi, Y., & Scheraga, H. (1984) *Macromolecules* 17, 148-55.
- Suguna, K., Bott, R., Padlan, E., Subramanian, E., Sheriff, S., Cohen, G., & Davies, D. (1987a) *J Mol Biol* 196, 877-900.
- Suguna, K., Padlan, E., Bott, R., Boger, J., Parris, K., & Davies, D. (1992) *Proteins: Struct, Funct, and Genet* 13, 195-205.
- Suguna, K., Padlan, E., Smith, C., Carlson, W., & Davies, D. (1987b) *Proc. Natl. Acad. Sci.* 82, 7009-7013.
- Swain, A., Miller, M., Green, J., Rich, D., Kent, S., & Wlodawer, A. (1990) *Proc Natl Acad Sci* 87, 8805-9.
- Szelke, M. (1985) in *Aspartic proteinases and their inhibitors* (Kotska, V. E., Ed.) pp 421-441, De Gruyter, Berlin.
- Szelke, M., Leckie, B., Hallett, A., Jones, D., Sueiras, J., Atrash, B., & Lever, A. (1982) *Nature* 299, 555-7.
- Szeltner, Z., & Polgar, L. (1996) *J Biol Chem* 271, 5458-63.
- Takahashi, M., Wang, T., & Hofmann, T. (1974) *Biochem Biophys Res Commun* 57, 39-46.
- Takahashi, S., Kim, E., Hibino, T., & Ooi, T. (1989) *Biopolymers* 28, 995-1009.
- Tanford, C. (1962) *J Am Chem Soc* 84, 4240-7.
- Tang, J. (1971) *J Biol Chem* 246, 4510-4517.
- Tang, J., James, M., Hsu, I., Jenkins, J., & Blundell, T. (1978) *Nature* 271, 618-621.
- Tang, J., & Wong, R. (1987) *J Cell Biochem* 33, 53-63.
- Tatunashvili, L., & Privalov, P. (1986) *Biofizika* 31, 578-81.
- Thaisrivongs, S., Watenpaugh, K., Howe, W., Tomich, P., Dolak, L., Chong, K., Tomich, C., Tomasselli, A., Turner, S., Strohbach, J., & et, a. (1995) *J Med Chem* 38, 3624-37.
- Thanki, N., Rao, J., Foundling, S., Howe, W., Moon, J., & al., e. (1992) *Protein Science* 1, 1061-72.
- Thompson, S., Murthy, K., Zhao, B., Winborne, E., Green, D., Fisher, S., DesJarlais, R., Tomaszek, T. J., Meek, T., Gleason, J., & et, a. (1994) *J Med Chem* 37, 3100-7.
- Thompson, W., Fitzgerald, P., Holloway, M., Emini, E., Darke, P., & al., e. (1992) *J Med Chem* 35, 1685-701.

- Tidor, B., & Karplus, M. (1991) *Biochemistry* 30, 3217-28.
- Tiktopulo, E., & Privalov, P. (1978) *FEBS Lett* 91, 57-8.
- Tirado-Rives, J., & Jorgensen, W. (1991) *Biochemistry* 30, 3864-71.
- Tischenko, V., & Gorodnov, B. (1979) *Biofizika* 24, 334-5.
- Tischenko, V., Tiktopulo, E., & Privalov, P. (1974) *Biofizika* 19, 400-404.
- Tisdale, M. (1996) *International Antiviral News* 4.
- Tomasselli, A., Hui, J., Adams, L., Chosay, J., Lowery, D., Greenberg, B., Yem, A., Deible, J., Zurcher-Neely, H., & Heinrikson, R. (1991) *J Biol Chem* 266, 14548-53.
- Tomasselli, A., Huy, J., Sawyer, T., Staples, D., FitzGerald, D., Chaudhary, V., Pastan, I., & Heinrikson, R. (1990a) *J Biol Chem* 265, 408-13.
- Tomasselli, A., Olsen, M., Hui, J., Staples, D., Sawyer, T., Heinrickson, R., & Tomich, C. (1990b) *Biochemistry* 29, 264-269.
- Tong, L., Pav, S., Pargellis, C., Do, F., Lamarre, D., & Anderson, P. (1993) *Proc Natl Acad Sci* 90, 8387-91.
- Tozser, J., Blaha, I., Copeland, T., Wondrak, W., & Oroszlan, S. (1991a) *FEBS Letters* 281, 77-80.
- Tozser, J., Gustchina, A., Weber, I., Blaha, I., Wondrak, E., & Oroszlan, S. (1991b) *FEBS Lett* 279, 356.
- Tozser, J., Weber, I., Gustchina, A., Blaha, I., Copeland, T., Louis, J., & Oroszlan, S. (1992) *Biochemistry* 31, 4793.
- Vacca, J. (1994) *Methods Enzymol* 241, 311-34.
- Vacca, J., Guare, J., deSolms, S., Sanders, W., Giuliani, E., & al., e. (1991) *J Med Chem* 34, 1225-8.
- van Gunsteren, W., Beutler, T., Fraternali, F., King, P., Mark, A., & Smith, P. (1993) in *Computer simulation of biomolecular systems. Theoretical and experimental applications.* (van Gunsteren, W., Weiner, P., & Wilkinson, A., Eds.), ESCOM Science publishers BV, Leiden.
- van Gunsteren, W., Luque, F., Timms, D., & Torda, A. (1994) *Annu Rev Biophys Biomol Struct* 23, 847-63.
- Vara Prasad, J., Para, K., Tummino, P., Ferguson, D., McQuade, T., Lunney, E., Rapundalo, S., Batley, B., Hingorani, G., Domagala, J., Gracheck, S., Bhat, R., Liu, B., Baldwin, E., Erickson, J., & Sawyer, T. (1995) *J. Med. Chem.* 38, 898-905.
- Vásquez, M., Pincus, M., & Scheraga, H. (1987) *Biopolymers* 26, 351-71.
- Vásquez, M., & Scheraga, H. (1988) *Biopolymers* 27, 41-58.
- Veerapandian, B., Cooper, J., Sali, A., & Blundell, T. (1990) *J Mol Biol* 216, 1017-29.
- Veerapandian, B., Cooper, J., Sali, A., Blundell, T., Rosati, R., Dominy, B., Damon, D., & Hoover, D. (1992) *Protein Sci* 1, 322-8.
- Viguera, A., & Serrano, L. (1995a) *Biochemistry* 34, 8771-9.
- Viguera, A., & Serrano, S. (1995b) *J Mol Biol* 251, 150-60.



- Wada, A. (1976) *Adv Biophys* 9, 1-13.
- Waltho, J., Feher, V., Merutka, G., Dyson, H., & Wright, P. (1993) *Biochemistry* 32, 6337-47.
- Wand, A., Roder, H., & Englander, S. (1986) *Biochemistry* 25, 1107-14.
- Wang, G., Li, S., Wideburg, N., Krafft, G., & Kempf, D. (1995) *J Med Chem* 38, 2995-3002.
- Wang, H., & Levinthal, C. (1991) *J Comput Chem* 12, 868-71.
- Wang, Y., Freedberg, D., Yamazaki, T., Wingfield, P., Stahl, S., Kaufman, J., Kiso, Y., & Torchia, D. (1996) *Biochemistry* 35, 9945-50.
- Watson, F., Wood, S., Tickle, I., Shearer, A., Sibanda, B., Newman, M., Khan, G., Foundling, S., Cooper, J., Veerapandian, B., & Blundell, T. (1988) in *18th Linderstrom-Lang Conference Proceedings*, Elsinore, Denmark.
- Weber, I., Miller, M., Jaskolski, M., Leis, J., Skalka, A., & Wlodawer, A. (1989) *Science* 243, 928-31.
- Weiner, S., Kollman, P., Case, P., Singh, U., Ghio, C., Alagona, G., Profeta, S., & Weiner, P. (1984) *J Am Chem Soc* 106, 765-84.
- Wesson, L., & Eisenberg, D. (1992) *Protein Science* 1, 227-235.
- Williams, R., Chang, A., Juretic, D., & Loughran, S. (1987) *Biochim Biophys Acta* 916, 200-4.
- Wimley, W., Creamer, T., & White, S. (1996) *Biochemistry* 35, 5109-24.
- Wlodawer, A., & Erickson, J. (1993) *Annu Rev Biochem* 62, 543-85.
- Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B., Baldwin, E., Weber, I., Selk, L., Clawson, L., Schneider, J., & Kent, S. (1989) *Science* 245, 616-621.
- Wodak, S., & Janin, J. (1980) *Proc Natl Acad Sci* 77, 1736-40.
- Wójcik, J., Altman, K., & Scheraga, H. (1990) *Biopolymers* 30, 121-34.
- Wolfenden, R. (1983) *Science* 222, 1087-93.
- Wolfenden, R., Andersson, L., Cullis, P., & Southgate, C. (1981) *Biochemistry* 20, 849-55.
- Wondrak, E., Louis, J., & Oroszlan, S. (1991) *FEBS Lett* 280, 34-346.
- Workman, R., & Burkitt, D. (1979) *Archives of Biochemistry and Biophysics* 194, 157-164.
- Wright, P., Dyson, H., & Lerner, R. (1988) *Biochemistry* 27, 7167-75.
- Xie, D., Fox, R., & Freire, E. (1994) *Protein Science* 3, 2175-84.
- Xie, D., & Freire, E. (1994a) *Proteins: Struct Funct and Genet* 19, 291-301.
- Xie, D., & Freire, E. (1994b) *J Mol Biol* 242, 62-80.
- Xie, D., Haynie, D., & Freire, E. (1993) *Proteins: Struct, Funct and Genet* 17, 111-23.
- Xie, L., Xie, J., Kashiwada, Y., Cosentino, L., Liu, S., Pai, R., Cheng, Y., & Lee, K. (1995) *J. Med. Chem.* 38, 3003-8.

- Yamakazi, T., Nicholson, L., Torchia, D., Wingfield, P., Stahl, S., Kaufman, J., Eyermann, C., Hodge, C., Lam, P., Ru, Y., Jadhav, P., Chang, C., & Weber, P. (1994) *JACS* 116, 10791-2.
- Yang, A., & Honig, B. (1995) *J Mol Biol* 252, 351-65.
- Yang, J., Spek, E., Gong, Y., Zhou, H., & Kallenbach, N. (1997) *Protein Science* 6, 1264-72.
- Yang, P., & Rupley, J. (1979) *Biochemistry* 18, 2654-61.
- Zehfus, M., Seltzer, J., & Rose, G. (1985) *Biopolymers* 24, 2511-19.
- Zhang, Z., Poorman, R., Maggiora, L., Heinrikson, R., & Kezdy, F. (1991) *J Biol Chem* 266, 15591-4.
- Zhou, N., Kay, C., Sykes, B., & Hodges, R. (1993) *Biochemistry* 32, 6190-7.
- Zielenkiewicz, P., & Saenger, W. (1992) *Biophys J* 63, 1483-6.
- Zimm, B., & Bragg, J. (1959) *J Chem Phys* 31, 526-35.



*Figura IV.A.2 Valores promedio de las áreas apolares y polares accesibles al disolvente para los distintos aminoácidos. Se representan el promedio de Boltzmann (barras blancas), la media aritmética entre las conformaciones estéricamente permitidas (barras rayadas) y el valor de accesibilidad en el mínimo energético. Las barras de error representan las correspondientes desviaciones estándar.*

**IV.A.3.2.-Cálculo de áreas accesibles promedio en conformaciones desplegadas.**

En las figuras IV.A.3 y IV.A.4 se muestran los perfiles de áreas accesibles apolares y polares, respectivamente, en función de los ángulos  $\phi$  y  $\psi$  del resto en la posición central de los tripéptidos Ala-X-Ala para algunos aminoácidos (Gly, Ala, Leu e Ile). En estos mapas conformacionales, las regiones no permitidas estéricamente se representan en color morado, mientras que, en las zonas accesibles, los valores de las áreas se muestran de acuerdo con el código de colores reflejado en las barras bajo cada gráfica.

Estas gráficas muestran cómo las principales restricciones conformacionales a la rotación de los ángulos del esqueleto peptídico del aminoácido en posición central aparecen como consecuencia de la presencia del carbono beta de su cadena lateral. De este modo, se puede

*Figura IV.A.3.: Perfiles de área accesible apolar del resto en la posición central de tripéptidos Ala-X-Ala en función de los ángulos  $\phi$  y  $\psi$  de su esqueleto peptídico. Se muestran los perfiles para los aminoácidos Gly (panel superior izquierdo), Ala (panel superior derecho), Leu (panel inferior izquierdo) e Ile (panel inferior derecho). Las accesibilidades se muestran de acuerdo con el código de color bajo cada figura. Las zonas en azul representan las regiones del espacio conformacional no permitidas estéricamente.*

*Figura IV.A.4: Perfiles de área accesible apolar del resto en la posición central de tripéptidos Ala-X-Ala en función de los ángulos  $\phi$  y  $\psi$  de su esqueleto peptídico. Se muestran los perfiles para los aminoácidos Gly (panel superior izquierdo), Ala (panel superior derecho), Leu (panel inferior izquierdo) e Ile (panel inferior derecho). Las accesibilidades se muestran de acuerdo con el código de color bajo cada figura. Las zonas en azul representan las regiones del espacio conformacional no permitidas estéricamente. (Página siguiente)*