



UNIVERSIDAD NACIONAL DE LA PLATA  
FACULTAD DE CIENCIAS EXACTAS  
DEPARTAMENTO DE QUÍMICA

**Soluciones diluidas de aminoácidos en agua.  
Teoría y simulación.**

Danilo Germán Renzi

2004

Esta Tesis para optar al grado de Doctor de la Facultad de Ciencias Exactas de la Universidad Nacional de La Plata, ha sido realizada en el Instituto de Física de Líquidos y Sistemas Biológicos (IFLYSIB) CONICET-UNLP bajo la dirección del Dr. Fernando Vericat, desempeñándose como Asesor Científico por la Facultad de Ciencias Exactas el Dr. Eduardo E. Mola.

## **AGRADECIMIENTOS.**

Quiero expresar mi más sincero agradecimiento a todas aquellas personas e instituciones que con su participación y colaboración hicieron posible la realización de este trabajo de Tesis.

Agradezco muy especialmente al *Dr. Fernando Vericat*, quien además de dirigir este trabajo con total responsabilidad y sapiencia, me honró con su vasta cordialidad y generosidad. Gracias a Rita y toda su familia por recibirme tan afectuosamente en cada viaje.

A mis amigos *César Stoico* y *Manuel Carlevaro*, con quienes tuve y tengo la enorme fortuna de compartir mucho más que trabajo.

A *Luis Pugnaroni*, *Augusto Melgarejo*, *Ernesto Caffarena*, *Andrés Salvay*, *Diego Vallejo*, *Ileana Márquez*, *Alfredo López*, *Guillermo Vasco Zarragoicoechea*, *Alejandro Mesón* y *Eitel Peltzer* porque no bastándoles con ayudarme cotidianamente en el trabajo, se brindaron siempre con total generosidad y afecto.

A *Raúl*, *Víctor*, *Osvaldo*, *Mónica*, *Silvia*, *Cristina*, *Eduardo*, *Juan*, *Inés*, *Graciela*, *Viviana*, *Federico* y *Néstor* por la excelente predisposición mostrada cada vez que necesité de su colaboración, muchas gracias.

Agradezco profundamente a todo el personal del IFLYSIB, en particular a *Rosa*, *Rubén*, *María*, *Carlitos*, y muy especialmente a mi madre platense *Lola* por todo el amor y el cuidado de estos años.

A la *Universidad Nacional de Rosario*, al *Instituto de Física de Líquidos y Sistemas Biológicos* y a la *Universidad Nacional de La Plata* que con su aporte hicieron posible la concreción de este trabajo de Tesis.

A mis padres *José* y *Alicia*, a mi hermano *Gustavo*, a *Carolina*, *Emiliano* y *Agostina* por el amor, el apoyo, y el respaldo incondicional a todas mis decisiones. Gracias por permitirme disfrutar y sentir en todo momento, el privilegio de contar con una familia. Y por último, a *Victoria*, que se empeña en llenar de felicidad cada uno de mis días.

# ÍNDICE

<b>PREFACIO</b>	i
1 PROTEÍNAS, AMINOÁCIDOS Y AGUA: “UN LENGUAJE BIOLÓGICO”.	1
1.1 PROTEÍNAS: “El discurso”.	2
1.2 AMINOÁCIDOS: “Las palabras”.	4
1.3 AGUA: “El contexto”.	6
1.4 EFECTO HIDROFÓBICO: ¿”La sintaxis”?	13
1.5 REFERENCIAS DEL CAPÍTULO.	15
2 HERRAMIENTAS TEÓRICAS.	17
2.1 SIMULACIONES MOLECULARES.	17
2.1.1 DINÁMICA MOLECULAR.	21
2.2 TEORÍAS MECÁNICO–ESTADÍSTICAS.	25
2.2.1 ECUACIONES INTEGRALES Y TEORÍAS PERTURBATIVAS PARA LAS FUNCIONES DE CORRELACIÓN.	27
2.2.1.1 Ecuaciones integrales.	28
2.2.1.2 Teorías perturbativas.	30
2.2.2 TEORÍA DE PRATT Y CHANDLER.	32
2.3 REFERENCIAS DEL CAPÍTULO.	35
3 UN RESIDUO DE AMINOÁCIDO APOLAR EN AGUA: HIDRATACIÓN HIDROFÓBICA.	36
3.1 SIMULACIÓN MEDIANTE DINÁMICA MOLECULAR.	36
3.1.1 ENERGÍA LIBRE DE TRANSFERENCIA.	36
3.1.1.1 Escala de hidrofobicidad.	40
3.1.1.2 Relación lineal de Chothia.	43
3.1.2 SOLVATACIÓN.	45
3.1.2.1 Correlación soluto-solvente.	45
3.1.2.2 Puentes de hidrógeno.	50
3.2. TEORÍA MECÁNICO–ESTADÍSTICA.	53
3.2.1 APLICACIÓN DE LA TEORÍA DE PRATT Y CHANDLER.	53
3.2.1.1 Definiciones y ecuaciones del modelo.	53
3.2.1.2 Cálculo de la función de correlación radial.	56
3.2.2 NÚMERO DE COORDINACIÓN.	61
3.3 REFERENCIAS DEL CAPÍTULO.	63
4 DOS RESIDUOS APOLARES DISUELTOS EN AGUA: INTERACCIÓN HIDROFÓBICA.	66
4.1 TEORÍA MECÁNICO–ESTADÍSTICA.	66
4.1.1 TEORÍA DEL EFECTO HIDROFÓBICO.	67
4.1.1.1 Correlaciones para el sistema de referencia.	68
4.1.1.2 Efecto de las interacciones atractivas.	71
4.1.2 APLICACIÓN DE LA TEORÍA A RESIDUOS APOLARES.	75
4.1.2.1 Función de distribución de pares.	75
4.1.2.2 Potenciales efectivos de pares.	80
4.2 COMPARACIÓN CON SIMULACIÓN.	82
4.3 REFERENCIAS DEL CAPÍTULO.	84

5	EXTENSIÓN DE LA TEORÍA MECÁNICO–ESTADÍSTICA.	85
5.1	AMINOÁCIDOS POLARES Y CARGADOS.	85
5.1.1	POTENCIALES DE INTERACCIÓN DIRECTA.	86
5.1.2	FUNCIÓN DE CORRELACIÓN DE PARES.	88
5.1.3	POTENCIALES EFECTIVOS DE PARES.	91
5.1.3.1	Aminoácidos polares.	92
5.1.3.2	Aminoácidos cargados.	93
5.1.3.3	Aminoácidos pertenecientes a grupos diferentes.	95
5.2	POTENCIALES AL CONTACTO.	96
5.2.1	CORRELACIÓN DE LOS VALORES AL CONTACTO.	97
5.3	REFERENCIAS DEL CAPÍTULO.	99
6	COMENTARIOS FINALES.	101
	<b>APÉNDICES.</b>	105
A.	RESUMEN DE GROMOS96.	106
A.1	ECUACIONES DE MOVIMIENTO.	106
A.2	ALGORITMOS PARA LA RESOLUCIÓN DE LAS ECUACIONES DE MOVIMIENTO.	108
A.2.1	ALGORITMO SALTO DE RANA.	109
A.3	CAMPOS DE FUERZA.	110
A.4	CONDICIONES PERIÓDICAS DE CONTORNO.	115
A.5	ACOPLAMIENTO A BAÑOS TÉRMICOS Y DE PRESIÓN.	117
A.6	CONFIGURACIÓN Y VELOCIDADES INICIALES.	118
A.7	CÁLCULO DE ENERGÍAS LIBRES.	119
A.8	CICLO TERMODINÁMICO.	121
A.9	REFERENCIAS DEL APÉNDICE A.	122
B.	VALORES DE LOS PARÁMETROS USADOS EN LAS SIMULACIONES CON GROMOS96.	123
B.1	REFERENCIAS DEL APÉNDICE B.	125
C.	VALORES DE LOS PARÁMETROS UTILIZADOS EN EL CÁLCULO DE LAS FUNCIONES DE DISTRIBUCIÓN, MEDIANTE LA TEORÍA MECÁNICO–ESTADÍSTICA.	127

## Prefacio

El problema del plegamiento de proteínas, esto es, descubrir el mecanismo por el cual la secuencia de aminoácidos determina su estructura plegada (estado nativo), y por lo tanto, su función biológica, constituye uno de los principales desafíos pendientes para la Biología Molecular y la Biofísica de este siglo.

Entender el principio que lleva a una proteína a su plegamiento nativo en condiciones fisiológicas normales, posee muchísimo interés para ramas del conocimiento tales como: la Biología, la Medicina o la Farmacología. Sus consecuencias sobre los diversos aspectos de la vida del hombre en nuestro planeta, algunas inimaginables a esta altura de nuestro saber, pueden ser verdaderamente muy significativas.

Por la importancia e interés que presenta este problema, en los últimos años, se han utilizado las más diversas herramientas para abordarlo, tanto experimentales como teóricas. Desde el punto de vista teórico, el arsenal de enfoques, técnicas y métodos utilizados con ese fin es vastísimo. Por la complejidad que presenta, la mayoría se basa en cálculos numéricos más que analíticos y hacen un intensivo uso de computadoras, cuyas velocidades y memorias son cada vez mayores.

Algunas de las dificultades encontradas, se ponen de manifiesto en las “paradojas” que el plegamiento de las proteínas presenta. La más conocida es la planteada por Levinthal, según la cual, si la proteína hiciera una búsqueda exhaustiva por el espacio de configuraciones hasta encontrar la correspondiente al mínimo absoluto de energía libre, el tiempo que le requeriría alcanzar su estructura nativa desde el momento en que se sintetiza, sería enorme (del orden de la edad del Universo).

Para resolver esta paradoja, se ha propuesto que la proteína, al plegarse, siga un “camino” que la conduzca a la estructura nativa; y este camino, habría quedado definido por una apropiada selección de secuencias de aminoácidos ocurrida a lo largo de la evolución.

Sin embargo, esto lleva a una segunda paradoja: ¿Cómo fueron seleccionadas a lo largo de la evolución las secuencias de proteínas cuyas

estructuras nativas cumplen con la función biológica necesaria? El número de secuencias posibles para una proteína de longitud 100 es  $20^{100}$ , esto significa que no todas las proteínas pudieron formarse. Nuevamente resulta imposible pensar que se haya recorrido el espacio de secuencias completo antes de encontrar a aquellas apropiadamente plegadas.

En general, en los estudios teórico-computacionales se han usado dos tipos de modelos: los llamados *fenomenológicos*, que contemplan el paisaje energético globalmente sin tener en cuenta los detalles atómicos, y los *atómicos* que consideran el conjunto completo de átomos que constituyen a la molécula de proteína. Dentro de este último tipo de estudio podemos distinguir al menos dos líneas: (a) modelado por comparación, y (b) predicciones ab-initio.

En el primer caso, para una dada secuencia de estructura desconocida se buscan patrones en un banco de estructuras de proteínas (Protein Data Bank). Luego se confeccionan listas de la tendencia de los diferentes aminoácidos a formar parte de las distintas estructuras secundarias en las conformaciones nativas de esas proteínas, y se intenta ensamblar las estructuras secundarias predichas en estructuras terciarias, de manera que sea consistente con ciertos requisitos, tales como: una alta densidad de empaquetamiento, un centro no polar, y algunas otras restricciones características de las proteínas.

En los métodos de predicción ab-initio se procura alcanzar la estructura de una secuencia dada a partir del conocimiento de las funciones de energía entre diversos átomos o grupos de átomos, es decir, de los campos de fuerzas. En algunos estudios, este conocimiento se complementa con la suposición simplificadora de que los grupos de átomos pueden ocupar solamente los sitios de una red. El caso más general en el cual los átomos pueden, en principio, ocupar en forma continua las distintas regiones del espacio, requiere de mayor potencia de cálculo.

Los códigos computacionales conocidos como GROMOS y el más reciente GROMACS, son ejemplos de programas diseñados para seguir este enfoque. En ambos casos pueden simularse moléculas de interés biológico en general, y polipéptidos en particular, definiendo sus átomos o grupos de átomos, uniones y

ángulos. Las simulaciones pueden realizarse suponiendo a las moléculas en el “vacío” o disueltas en algún solvente de interés, como puede ser el agua.

La dinámica, de alguna manera, soslaya a la paradoja de Levinthal por cuanto ésta debería conducir al sistema determinísticamente por un camino adecuado que recorra una región limitada del espacio de configuraciones, sin necesidad de tener que visitarlo íntegramente para analizar la energía de todas las configuraciones posibles. Sin embargo, para resolver completamente el problema existen todavía serios inconvenientes con: el conocimiento de los campos de fuerza reales que entran en juego, y la velocidad de las computadoras.

Cualquiera sea el enfoque utilizado, tanto experimental como teórico, hay algo universalmente aceptado: *el solvente cumple un rol fundamental en el proceso de plegamiento y en la estabilidad de la estructura plegada*, aún cuando los detalles de esta contribución no se conozcan completamente.

Considerando el predominio que el agua tiene en el medio celular, es fácil comprender los esfuerzos realizados para desentrañar su influencia, particularmente en relación con los aminoácidos apolares y el denominado *efecto hidrofóbico*.

Si bien el problema del plegamiento requiere de tratamientos que abarquen al sistema completo en forma simultánea, estudiar el efecto del solvente sobre los aminoácidos individuales, contribuiría a entender mejor el rol de éste, tanto en el proceso de plegamiento como en la estabilidad de la estructura nativa. Más aún, estos estudios podrían utilizarse también en otros problemas de interés biológico, como por ejemplo ciertos mecanismos de reconocimiento entre moléculas.

En esta Tesis estudiamos soluciones diluidas de aminoácidos desde un punto de vista teórico y con simulación mediante Dinámica Molecular utilizando el paquete de programas GROMOS96.

La estructura del trabajo es la siguiente. En el primer capítulo hacemos una breve presentación de nuestros objetos principales de estudio, los aminoácidos y el agua. Repasamos el papel de los aminoácidos en relación a las proteínas y discutimos brevemente algunos aspectos del *efecto hidrofóbico*, referidos tanto a la *hidratación hidrofóbica* como a la *interacción hidrofóbica*.

En el capítulo 2 consideramos las principales herramientas que usaremos a lo largo de la Tesis: Dinámica Molecular mediante GROMOS96 y ecuaciones integrales para las correlaciones *residuo–agua* y *residuo–residuo*.

En el capítulo siguiente utilizamos estas herramientas para estudiar la *hidratación hidrofóbica* que experimenta un aminoácido apolar en un medio acuoso. Obtenemos las correlaciones entre los átomos del residuo y el agua a partir de las simulaciones y de la solución de las ecuaciones integrales correspondientes.

Además establecemos una nueva escala de hidrofobicidad mediante simulación con Dinámica Molecular, calculando la variación de energía libre que se obtiene al transferir un residuo de aminoácido apolar desde un solvente orgánico a agua.

El capítulo 4 está dedicado al estudio de la *interacción hidrofóbica*. El sistema ahora está formado por dos residuos apolares disueltos en agua. En esencia, usamos una teoría mecánico–estadística para calcular las funciones de distribución *residuo–residuo* y a partir de ellas los potenciales de fuerza media.

En el capítulo 5 extendemos la teoría desarrollada en los dos capítulos anteriores para incluir a los residuos de aminoácidos polares y cargados. Luego, comparamos los valores al contacto de nuestros potenciales de fuerza media con los obtenidos por otros métodos que usualmente son utilizados en cálculos de reconocimiento de estructuras u otras aplicaciones.

Finalizamos este trabajo discutiendo los resultados obtenidos en relación a su potenciales aplicaciones y las futuras líneas de investigación que procuraremos desarrollar a partir de aquí.

Casilda, septiembre de 2004

Danilo Germán Renzi

# 1. PROTEÍNAS, AMINOÁCIDOS Y AGUA: “UN LENGUAJE BIOLÓGICO”.

Hay un problema que, desde hace décadas, ha merecido y merece la atención de muchos de los investigadores que trabajamos en áreas como la Física Biológica o la Biología Molecular. Este problema es el llamado *problema del plegamiento de proteínas*, y consiste en descubrir el código por el cual la secuencia de aminoácidos de una proteína determina su estructura plegada (estado nativo) y por tanto, su función biológica.

Conocer dicha relación nos permitiría saber cómo un estado tan particular de la materia surge de las leyes subyacentes de la Química y la Física. Además, esta información podría utilizarse para identificar las estructuras de miles de secuencias de proteínas que son descubiertas todos los días en laboratorios de biotecnología, ya que experimentalmente resulta más fácil determinar la secuencia de aminoácidos que la forman, que la estructura tridimensional de su estado nativo. Esto ayudaría a saber cuáles son las secuencias útiles para determinados propósitos y cuáles no lo son.

Un problema relacionado con éste y de sumo interés es el de *diseño* de secuencias de proteínas, también llamado el problema inverso al de plegamiento. En un algoritmo de plegamiento ingresaríamos una secuencia de aminoácidos, y obtendríamos la estructura nativa correspondiente. En cambio, un algoritmo para el problema inverso usaría como entrada una estructura nativa deseada y daría a la salida una lista de secuencias que se pliegan a ella.

Desde luego, el problema del plegamiento y el inverso, son dos formulaciones distintas del mismo problema y ambos requieren el entendimiento de la relación existente entre secuencia de aminoácidos y estructura nativa.

Entre las diversas maneras en las que se ha abordado este problema, se encuentran las simulaciones numéricas con computadoras. Si se contara con un programa de computación capaz de simular el proceso que ocurre en la célula, se podría intentar reproducir el plegamiento de una proteína a su estado nativo tridimensional, a partir de la definición de su secuencia y del campo de fuerzas con el cual interactúan cada uno de los átomos que forman parte del sistema.

Lamentablemente, existen varias limitaciones para poder realizar esta tarea, algunas relacionadas con el tamaño del sistema (varios miles de átomos), otras con el tiempo de simulación necesario para representar el proceso de plegamiento (de milésimas a decenas de segundos) y otras con el conocimiento (o desconocimiento) de las interacciones entre los átomos involucrados.

Un sistema como éste requiere de una enorme potencia de cálculo, superior aún al de las computadoras más poderosas con las que se cuenta hoy en día. Por esta razón, aún cuando fuésemos capaces de representar con precisión la estructura de las moléculas y las interacciones presentes, sería imprescindible realizar algunas aproximaciones tendientes a simplificar ambos aspectos.

En el marco de este problema tan interesante y complejo, nos planteamos dar un pequeño paso hacia un mejor entendimiento de ciertas propiedades físicas de las moléculas que forman a las proteínas (aminoácidos), y sus interacciones efectivas cuando se hallan disueltas en agua.

En esta Tesis, consideramos algunos modelos aproximados de las moléculas de aminoácidos y del agua como solvente para estudiar ciertas propiedades estructurales y termodinámicas del sistema. Utilizamos dos herramientas de cálculo diferentes: la simulación mediante Dinámica Molecular, y una teoría mecánico-estadística de líquidos clásicos.

Para poner a nuestro trabajo en un contexto adecuado, a continuación exponemos brevemente algunos aspectos de interés de las proteínas, los aminoácidos y el agua. Podemos imaginar a estos tres elementos como integrantes de un “lenguaje biológico” que permite relacionar la información codificada en el ADN con las estructuras que las proteínas adquieren en el medio celular, y por ende, con su función vital.

## **1.1 PROTEÍNAS: “El discurso”.**

Las proteínas son macromoléculas biológicas muy abundantes e importantes (su nombre proviene del griego *protos*, “*primero*” o “*más importante*”), ya que participan de casi todos los procesos que tienen lugar en las células, además de formar parte de las mismas [1]. Cumplen con diversas funciones, como catalizar

reacciones celulares, transportar moléculas o formar estructuras básicas (piel, pelo, plumas o tendones).

Una proteína es una cadena polimérica lineal (que en una analogía lingüística podemos asimilar con el “discurso”) de decenas a miles de una clase especial de unidades monoméricas llamadas *aminoácidos* (las “palabras”). Todas las proteínas en todas las especies, independientemente de su función o actividad biológica, están constituidas por el mismo conjunto de 20 aminoácidos (el “léxico o vocabulario”) todos en la forma isomérica L.

Una proteína se distingue de otra por el tipo, número y orden (*secuencia*) que presentan sus aminoácidos en la cadena. Esto se conoce como la *estructura primaria* de la proteína.

Ahora bien, estas cadenas poliméricas no son libres de adoptar cualquier estructura tridimensional al azar. Los impedimentos estéricos y una gran variedad de interacciones débiles hacen que unos plegamientos sean más estables que otros. Aparecen así, algunas disposiciones regulares como las *hélices  $\alpha$*  y las *conformaciones  $\beta$* , que se denominan *estructuras secundarias* de las proteínas.

A su vez, estas estructuras secundarias se organizan, bajo condiciones fisiológicas normales (un medio acuoso con pH aproximadamente neutro, presión normal y temperaturas entre 20–40 °C), en una *única* estructura tridimensional compacta llamada *estructura terciaria* o *estado nativo* (en la analogía lingüística correspondería al “discurso coherente u oración”), que permite poner de manifiesto la función biológica de la proteína (la “semántica o significado”). Algunas proteínas poseen varias cadenas polipeptídicas y tienen un nivel adicional de estructura: la *estructura cuaternaria*.

Bajo condiciones especiales, como pueden ser altas temperaturas, variaciones de pH, o solvente no acuoso, la proteína se desnaturaliza o despliega y adquiere una conformación extendida.

En el año 1973 Anfinsen [2] llevó a cabo un experimento en el cual demostró que algunas proteínas desnaturalizadas eran capaces de recuperar su función biológica con sólo devolverle al medio las condiciones adecuadas, sin ninguna ayuda de otros elementos u organismos celulares. El experimento de Anfinsen

sugirió que (al menos para algunas proteínas) la secuencia de aminoácidos contiene toda la información necesaria para determinar el plegamiento nativo.

## 1.2 AMINOÁCIDOS: “Las palabras”.

Según dijimos, las secuencias de todas las proteínas conocidas se forman a partir de sólo 20 aminoácidos, cuya única diferencia se encuentra en el conjunto de átomos que forman su *cadena lateral* o *residuo* (recuadrado en la figura 1.1).

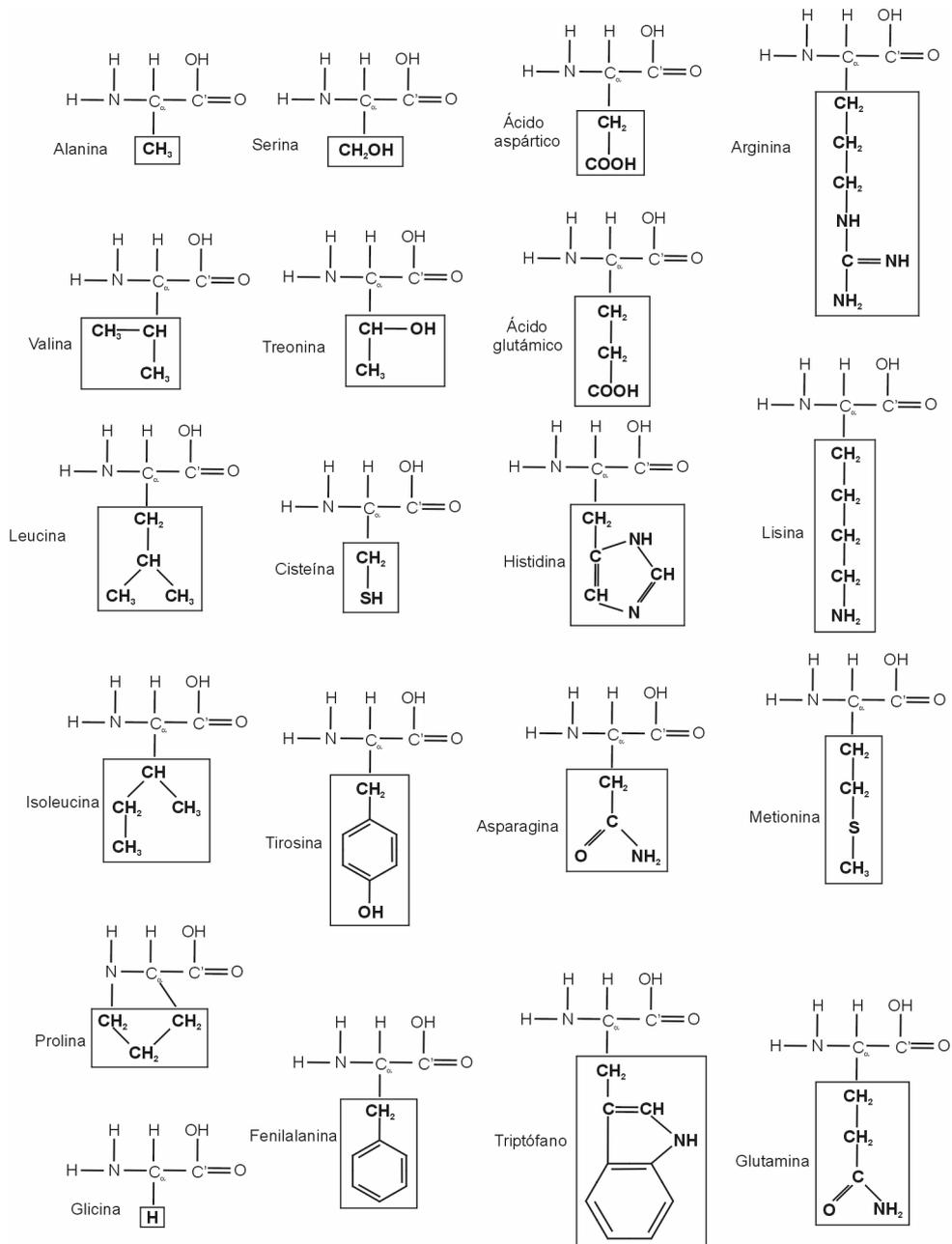


figura 1.1

El átomo clave en un aminoácido es el átomo de carbono llamado *alfa* ( $C_{\alpha}$ ). Ligados a él por uniones covalentes simples se encuentran otros cuatro átomos, uno de hidrógeno (H), uno de nitrógeno (N), un carbono carboxilo ( $C'$ ), y el primer átomo del residuo.

El átomo de carbono carboxilo ( $C'$ ) está conectado a un grupo hidroxilo (OH) por una unión covalente simple y a un átomo de oxígeno (O) por una unión covalente doble. El nitrógeno está unido a dos átomos de hidrógeno, formando un grupo amino  $NH_2$ . La *cadena principal* de una proteína está formada por el triplete de átomos N,  $C_{\alpha}$ ,  $C'$  de cada aminoácido que la compone [3].

Tabla 1.1: Denominación y clasificación de los aminoácidos [3].

AMINOÁCIDOS	NOMBRE	CÓDIGO DE 3 LETRAS	CÓDIGO DE 1 LETRA
<b>No polares o Apolares (hidrofóbicos)</b>	alanina	ala	A
	valina	val	V
	fenilalanina	phe	F
	leucina	leu	L
	isoleucina	ile	I
	metionina	met	M
	prolina	pro	P
<b>Cargados (hidrofílicos)</b>	ácido aspártico	asp	D
	ácido glutámico	glu	E
	lisina	lys	K
	arginina	arg	R
<b>Polares (hidrofílicos)</b>	serina	ser	S
	treonina	thr	T
	tirosina	tyr	Y
	histidina	his	H
	cisteína	cys	C
	asparagina	asn	N
	glutamina	gln	Q
	triptófano	trp	W
<b>Ninguno de los anteriores</b>	glicina	gly	G

Como detallamos en la Tabla 1.1, podemos clasificar a los aminoácidos de acuerdo a su polaridad en: *apolares*, *polares* y *cargados*. La mayoría de los investigadores consideran a la *glicina* carente de residuo, y por eso la dejan fuera de los conjuntos anteriores. Otros prefieren incluirla entre los aminoácidos apolares.

Dos moléculas de aminoácidos pueden unirse covalentemente a través de un enlace llamado *peptídico*, formando un *dipéptido*. Este enlace se forma por eliminación de los elementos del agua del grupo  $\alpha$ -carboxilo de un aminoácido y del grupo  $\alpha$ -amino del otro (figura 1.2).

De igual manera pueden unirse muchos aminoácidos para formar un *polipéptido*. Las cadenas de aminoácidos se denominan habitualmente *péptidos*. Los péptidos que se encuentran en la naturaleza varían en tamaño desde pequeñas moléculas que sólo contienen dos o tres aminoácidos, hasta macromoléculas que contienen decenas o cientos de ellos, como las proteínas.

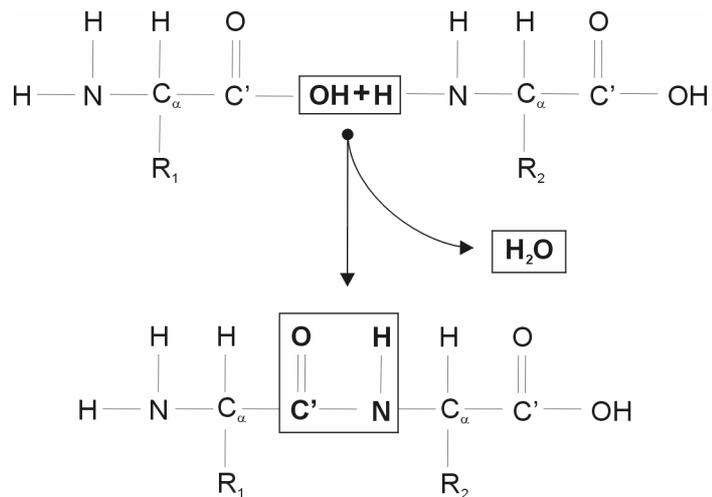


figura 1.2

### 1.3 AGUA: “El contexto”

Si bien las proteínas son solubles en varios solventes orgánicos, el agua es el solvente más importante en Biología. El agua, como componente fundamental del medio celular, participa de todos los procesos que ocurren dentro de la célula viva. En particular, tiene una influencia decisiva en el plegamiento de las proteínas, en la estabilidad del estado nativo y en la unión que éstas tienen con otras moléculas o átomos, relacionadas a su función biológica.

A pesar del rol fundamental que cumple como solvente fisiológico, y de ser el líquido más abundante sobre la Tierra, no existe todavía una teoría microscópica de agua en estado líquido y de soluciones acuosas que sea completamente satisfactoria. Después de muchos siglos de trabajo e investigación, y de miles de páginas escritas sobre ella, el agua sigue siendo un líquido que no deja de sorprendernos por sus propiedades únicas [4, 5].

La pregunta que naturalmente nos hacemos es, ¿por qué el agua tiene esas propiedades tan particulares que la distinguen de todos los demás solventes? Para intentar responderla, conviene observar algunos aspectos que caracterizan a los llamados líquidos simples, como por ejemplo argón en estado líquido, y compararlos con el correspondiente comportamiento mostrado por el agua.

En particular, el aspecto que nos interesa señalar aquí, es la estructura general del líquido tal cual es descrita en el espacio real por la función de correlación de pares, que es la transformada de Fourier del factor de estructura definido en el espacio de los momentos. El factor de estructura se puede obtener a partir de la intensidad total resultante de experimentos de difracción de rayos X o de neutrones, separándolo del factor de forma atómico.

Es bien conocido que los sólidos cristalinos tienen funciones de correlación que se extienden periódicamente con máximos bien determinados, que definen un orden estructural característico de largo alcance. En cambio, en los líquidos atómicos simples las funciones de correlación muestran típicamente oscilaciones cuyos máximos se atenúan rápidamente en unos pocos diámetros atómicos, definiendo un orden de corto alcance.

Podemos entender fácilmente este comportamiento si consideramos las interacciones entre los átomos (o más generalmente las moléculas) que forman a estos líquidos. Estas interacciones dependientes sólo de la distancia entre partículas, pueden ser modeladas a nivel de Born–Oppenheimer mediante potenciales del tipo de Lennard–Jones. Para pequeñas distancias de separación, el potencial es fuertemente repulsivo y de muy corto alcance, indicando la imposibilidad que tienen las partículas de penetrarse a medida que se acercan. Esta repulsión tiene su origen en el principio de exclusión de Pauli que impide el solapamiento de las nubes electrónicas de los átomos. A distancias mayores,

tenemos atracciones de largo alcance del tipo van der Waals, cuyo origen debe buscarse en las interacciones entre dipolos inducidos por fluctuaciones. En la zona intermedia los dos tipos de fuerzas compiten, determinando un mínimo de energía potencial.

Ahora bien, a diferencia del estado gaseoso, el estado líquido de un sistema se caracteriza por su gran densidad. Por ende, las partículas constituyentes están muy próximas entre sí. En éstas condiciones, la parte dominante del potencial interpartículas es la repulsiva, y consecuentemente, la que determina la estructura de corto alcance en los líquidos simples [6]. Es decir, la forma en que se ordenan los átomos o moléculas alrededor de uno cualquiera de ellos. Este comportamiento explica el éxito de las teorías perturbativas para describir a los líquidos simples. La parte repulsiva representa la parte principal o de referencia, y las interacciones atractivas desempeñan un rol secundario o perturbativo en el cálculo de la estructura, aún cuando su contribución pueda ser importante al considerar la termodinámica del sistema.

Estos comentarios nos permiten entender por qué el líquido de esferas rígidas ocupa un lugar estratégico en la teoría de líquidos simples. En éste modelo, los átomos se representan por esferas impenetrables que interactúan a través de un potencial que constituye el caso extremo de una fuerza infinitamente repulsiva y cuyo alcance es justamente el diámetro de las esferas. Este sencillo modelo da lugar a un orden de corto alcance muy próximo al que se observa en los líquidos simples. Alrededor de una partícula dada se pueden acomodar (dependiendo de la densidad) aproximadamente una decena de partículas vecinas en la primera capa que se extiende hasta una distancia cercana al mínimo de energía potencial. Este número está muy próximo a doce, que es el número máximo de esferas rígidas que pueden acomodarse alrededor de una dada. En consecuencia, el líquido de esferas rígidas es capaz de describir las características básicas del orden de corto alcance observado en los líquidos simples. Por esto y por su simplicidad, constituye el sistema de referencia casi universalmente usado en las teorías perturbativas de líquidos simples. En este sentido el líquido de esferas rígidas ocupa un lugar equivalente al que tienen el gas ideal para los gases o el oscilador armónico para los sólidos.

Estas ideas sobre la estructura de los líquidos simples dejan de ser válidas para el agua. Esto se debe a que la interacción entre moléculas de agua es mucho más compleja que la existente entre dos átomos de argón. Para comenzar, la molécula de agua no tiene un centro de simetría como los átomos de gases raros. Esto hace que las fuerzas dependan fuertemente de las orientaciones relativas, además de la distancia entre moléculas. Por otra parte, los núcleos y electrones que forman la molécula dan lugar a distribuciones de carga con momentos multipolares permanentes, donde el dipolar y el cuadrupolar son responsables de interacciones atractivas importantes que se suman a las de van der Waals y a las repulsivas. Por último, y quizás la diferencia esencial con los líquidos simples, un hidrógeno de una molécula de agua y uno de los orbitales aislados del oxígeno de otra, tienen la posibilidad de formar puentes de hidrógeno debido al carácter iónico de ambos. Este tipo de enlace, que es en cierto modo intermedio entre un enlace covalente y una interacción electrostática, puede pensarse como una interacción atractiva que depende específicamente de la orientación relativa de las moléculas involucradas y de su distancia.

Es muy probable que alrededor de una molécula de agua se acomoden otras cuatro formando puentes de hidrógeno con ella. Esto se debe a que la molécula de agua central tiene los dos orbitales aislados de su oxígeno dispuestos a unirse a los hidrógenos de otras dos moléculas vecinas, y además sus dos hidrógenos pueden formar uniones con los orbitales no comprometidos en el enlace covalente de otras dos moléculas. Esto, sumado a que los ángulos característicos de la molécula de agua tienen los valores adecuados, hace que en el agua líquida predomine una estructura tetraédrica tipo hielo, aunque con tiempos de vida más cortos.

Vemos entonces que los puentes de hidrógeno dan lugar a fuerzas atractivas muy intensas y de muy corto alcance que compiten con las fuerzas repulsivas en establecer la estructura del agua. De esta forma, la visión perturbativa de los líquidos simples deja de tener validez para el agua y otros líquidos “asociados”, cuyas moléculas son capaces de formar puentes de hidrógeno.

La complejidad de las interacciones entre moléculas de agua, en particular la formación de puentes de hidrógeno, además de invalidar el uso de teorías perturbativas, dificulta la definición de potenciales de tipo analítico al nivel de Born–Oppenheimer como se hizo con tanto éxito con el potencial de Lennard–Jones en el caso de líquidos atómicos simples. El mayor obstáculo para esto, es encontrar una expresión que describa adecuadamente los puentes de hidrógeno.

Recientemente [7, 8] se ha propuesto un potencial molecular que a las repulsiones y términos de interacción electrostática entre dipolos y cuadrupolos, le incorpora un término de tipo octupolar con el cual se intenta modelar la interacción “tetraédrica” debida a los puentes de hidrógeno. El modelo bautizado “Yukagua” permite algún tipo de tratamiento analítico y ha sido utilizado en simulaciones Monte Carlo, dando funciones de correlación átomo–átomo (O–O, O–H, H–H) que acuerdan razonablemente bien con las obtenidas mediante experimentos de difracción. Sin embargo, éste y otros pocos intentos que procuran modelar las interacciones entre moléculas de agua, mediante expresiones analíticas basadas en desarrollos multipolares con la finalidad de utilizarlos en teorías mecánico–estadísticas, se encuentran todavía en etapas preliminares.

La mayoría de los potenciales que han demostrado ser adecuados para describir las interacciones entre moléculas de agua, fueron diseñados para ser usados en simulaciones numéricas (Monte Carlo y Dinámica Molecular). Todos ellos modelan a las moléculas mediante sitios cargados que representan a átomos o grupos de átomos. Estos sitios interactúan con los de otra molécula, a través de potenciales radiales, adecuadamente parametrizados para reproducir propiedades estructurales y termodinámicas de algún estado particular (normalmente el hielo en ciertas condiciones) del que haya abundante datos experimentales. Dentro de este tipo de potenciales encontramos una gran cantidad de modelos que han sido intensivamente utilizados en las más diversas simulaciones. En particular, uno de los más populares es el modelo SPC/E (Extended Simple Point Charge) [9] desarrollado por Berendsen y colaboradores, al cual elegimos para realizar las simulaciones que presentaremos en esta tesis.

Más allá de los aspectos instrumentales, lo cierto es que las características tan particulares que presenta la interacción entre moléculas de agua, en especial

los puentes de hidrógeno, son causantes de que el agua líquida posea una estructura de tipo tetraédrica con aproximadamente cuatro vecinos. Esto la diferencia esencialmente de los líquidos simples que presentan, en general, un número de vecinos mayor, próximo al máximo de empaquetamiento de esferas rígidas. Ésta diferencia en la estructura del agua con respecto a otros potenciales solventes, es la responsable de las propiedades especiales del agua con sus múltiples consecuencias.

Según esta visión, el agua líquida tiene una estructura tetraédrica tan definida, que la pone en un lugar especial frente a todos los demás líquidos que actúan como solvente. Esto es así, aún en comparación con otros líquidos que también tienen la capacidad de agregarse a través de sus puentes de hidrógeno como los alcoholes, y por supuesto mucho más, frente a los líquidos simples.

En presencia de solutos apolares, el agua pone de manifiesto algunas de esas peculiaridades. Por ejemplo, la baja solubilidad de este tipo de soluto en agua, y el hecho observado experimentalmente, de que esa solubilidad disminuye al aumentar la temperatura. Podemos hacernos una idea aproximada de lo que ocurre en este caso si recordamos que la variación de energía libre al considerar el sistema soluto-agua con respecto al agua pura, puede escribirse  $\Delta G = \Delta H - T\Delta S$ . Puesto que la variación de entalpía prácticamente no cambia con la temperatura (a temperatura ambiente), el aumento de  $\Delta G$  con la temperatura es atribuido al segundo término de la relación anterior. Esto es posible si la variación entrópica  $\Delta S$  es negativa, lo que implica que el agregado del soluto vuelve más ordenado al sistema.

El aumento del orden al agregar una molécula apolar en agua se manifiesta en las proximidades de ésta y puede explicarse si aceptamos con Nemethy y Scheraga [10] que la formación de puentes de hidrógeno entre moléculas de agua permite dividir al conjunto total en 5 subconjuntos, de acuerdo al número de puentes de hidrógeno que forman: 0 (molécula libre), 1, 2, 3 ó 4. Si indicamos con  $E_i$  ( $i = 0,1,2,3,4$ ) a la energía de una molécula que forma  $i$  puentes de hidrógeno, se verifica que  $E_4 < E_3 < E_2 < E_1 < E_0$ . En presencia de la molécula apolar, ésta puede interactuar a través de fuerzas de van der Waals con las moléculas de agua y, en ese caso, su interacción con las moléculas que forman 0, 1, 2 y 3

enlaces equivale a reemplazar la interacción dipolo–dipolo que ésta ejerce con otra molécula de agua vecina y, en consecuencia, las correspondientes energías aumentan. Pero debido a que la estructura tetraédrica del agua da lugar a agregados formados de moléculas con 4 puentes de hidrógeno con una estructura tipo *clatratos*, es posible que la molécula apolar interactúe con este tipo de moléculas de agua y la interacción de van der Waals contribuya, como un “quinto enlace”, a disminuir la energía de éstas, favoreciendo así la estructuración del agua a su alrededor. En esta situación, generalmente se habla de solvatación y en el caso específico de agua de *hidratación*.

En vez de pensar lo que le ocurriría a una molécula de soluto apolar y analizar, como hicimos, qué pasa cuando ésta es obligada a sumergirse en agua, puede interesarnos considerar cuál es el comportamiento espontáneo relativo de algunas pocas de tales moléculas introducidas en agua. De acuerdo a lo señalado, al sumergirlas en agua éstas se hidratan aumentando la contribución a la entropía de las moléculas de agua alrededor de cada partícula apolar. Si consideramos, por ejemplo, dos de esas moléculas, entre ellas se ejercerán, por un lado, la interacción atractiva de van der Waals que se observaría en fase gaseosa, pero además aparecerá una fuerza atractiva extra debida al agua. Esta fuerza es tal que, cuando las dos moléculas estén a distancias relativamente cortas, tenderá a aproximarlas más. La razón es que de esa manera se tenderá a disminuir el área expuesta de cada molécula al agua con lo cual se dificultará la formación de la capa ordenada que las hidrata. En esa situación la variación entrópica entre el estado en que las dos moléculas tienden a juntarse y el estado en que las dos moléculas están suficientemente separadas, como para permitir que se formen sendas capas de hidratación, es positiva. Como resultado, la variación de energía libre  $\Delta G = \Delta H - T\Delta S$  disminuye determinando la dirección en que el sistema evolucionará espontáneamente hacia el equilibrio. Esto es, el efecto del agua se manifiesta en una interacción que favorece la atracción entre solutos apolares. Teniendo en cuenta la naturaleza de esta fuerza, se la denomina *interacción hidrofóbica*.

Ambos efectos, la hidratación y la interacción hidrofóbica, constituyen dos manifestaciones notables del agua frente a la presencia de solutos apolares que,

dada la naturaleza apolar de muchos de los residuos de aminoácidos que forman la secuencia de las proteínas, es de esperar sean de gran importancia en relación al plegamiento y estabilidad de las mismas.

#### **1.4 EFECTO HIDROFÓBICO: ¿“La sintaxis”?**

La información contenida en el ADN determina, en principio, la estructura primaria (“discurso impreciso”) de cada proteína, o sea, el número de aminoácidos (“palabras”) y su secuencia.

Sin embargo, la estructura primaria por sí sola no es suficiente para expresar la función biológica que la proteína debe cumplir (en nuestra analogía lingüística la “semántica o significado”). Se requiere de una estructura más elaborada para eso. La función biológica se manifiesta si la proteína se expresa mediante un “discurso” más refinado y coherente que sólo lo da la estructura plegada nativa (lingüísticamente la “oración o discurso coherente”) que la proteína adquiere en el “contexto” adecuado (solución acuosa en condiciones fisiológicas).

Si quisiéramos reproducir el proceso de plegamiento de las proteínas en el medio celular, deberíamos considerar al sistema completo formado por todos los átomos (de la proteína y del medio acuoso) con sus respectivas interacciones.

Como dijimos, actualmente es imposible intentar realizar esta tarea en forma exacta, tanto desde un punto de vista teórico como computacional. Por eso, para poder estudiar algunas propiedades, necesitamos considerar modelos simplificados de las moléculas que forman el sistema y de sus interacciones.

Entre las interacciones más importantes que participan del proceso de plegado y determinan la estabilidad de las proteínas, están las fuerzas repulsivas de exclusión, las fuerzas atractivas entre dipolos inducidos del tipo de van der Waals, entre multipolos de distintos órdenes (cargas, dipolos, cuadrupolos, etc), los puentes de hidrógeno y los puentes disulfuro. Pero además de éstas fuerzas, que son el resultado de la interacción directa entre los átomos o grupos atómicos, debemos considerar aquella que, según mencionamos, tiene su origen en la acción indirecta del agua sobre los solutos, particularmente los solutos apolares: la interacción hidrofóbica.

En 1936, Alfred Mirsky y Linus Pauling [11] propusieron a la formación de puentes de hidrógeno como “la” fuerza dominante en el plegamiento, y lograron predecir correctamente la existencia de estructuras secundarias como *hélices  $\alpha$*  o *conformaciones  $\beta$* , pero no tuvieron éxito con las estructuras terciarias.

Si consideramos que aquellos residuos capaces de formar puentes de hidrógeno en el interior de la proteína, también pueden hacerlo con el agua cuando se encuentran expuestos al medio, y que ambas interacciones son energéticamente del mismo orden, concluimos que la formación de puentes de hidrógeno no puede ser la principal fuerza responsable del plegamiento y estabilidad de las proteínas en solución.

Además, como la formación de puentes disulfuro y las interacciones entre residuos *cargados* son poco frecuentes, es razonable pensar que estas contribuciones tampoco pueden ser las determinantes de la estructura terciaria de las proteínas.

Usando estos argumentos, W. Kauzmann [12] propuso a la *interacción hidrofóbica* como la fuerza que determina el colapso de la proteína durante el proceso de plegamiento y su estabilidad en solución.

Esta propuesta de Kauzmann instaló a la *interacción hidrofóbica*, y en general al *efecto hidrofóbico*, en el centro de la escena. Si bien este efecto era conocido y estudiado en otros sistemas, a partir de aquí tomó gran relevancia e interés en el campo de los fenómenos biológicos.

Luego de su trabajo, surgieron diversos modelos simplificados pensados para estudiar las distintas facetas del plegamiento de proteínas [13, 14], donde en casi todos los casos, se definen interacciones que contienen implícitamente al *efecto hidrofóbico* como la causa principal del plegamiento de las proteínas en su estructura compacta.

En general, estos modelos son usados (con relativo éxito) para explicar diferentes aspectos del problema, como la dinámica de plegamiento, mutaciones de aminoácidos, dependencia con la temperatura, presión, etc.

Actualmente, el problema del plegamiento y estabilidad de las proteínas en condiciones fisiológicas normales sigue abierto, y desde luego, su resolución supera largamente las pretensiones de este trabajo de Tesis.

La valoración exacta de la incidencia que el *efecto hidrofóbico* tiene en el plegamiento y estabilidad de la estructura nativa, también sigue siendo una cuestión en discusión ya que no se cuenta con evidencia experimental que sea concluyente y definitiva al respecto, si bien existe un consenso generalizado sobre el rol protagónico que desempeña la acción del agua en los procesos biológicos.

Por este motivo, teniendo en cuenta la vigencia y el interés que todavía despierta estudiar el *efecto hidrofóbico* desde un punto de vista molecular para los sistemas biológicos en general, y para aminoácidos en particular, es que tomamos al *efecto hidrofóbico* en sus dos aspectos, la *hidratación* y la *interacción hidrofóbica*, como el eje alrededor del cual desplegamos el estudio de los aminoácidos disueltos en agua.

## 1.5 REFERENCIAS DEL CAPÍTULO.

- 1 A.L. Lehninger, D.L. Nelson and M.M. Cox, *Principios de Bioquímica* (Ediciones Omega S.A., Barcelona, 1993).
- 2 C.B. Anfinsen, *Science*, **181**, 223 (1973).
- 3 Branden and J. Tooze, *Introduction to Protein Structure* (Garland Publishing, Inc., New York, 1991).
- 4 D. Eisenberg and W. Kauzmann, *The Structure and Properties of Water* (Oxford University Press, London, 1969).
- 5 F. Franks (editor), *Water. A Comprehensive Treatise*. Vol. 1 (Plenum Press, New York, 1973).
- 6 J.P. Hansen and I.R. McDonald, *Theory of Simple Liquids*, (Academic Press, London, 1986).
- 7 L. Blum, F. Vericat and L. Degève, *Physica A*, **265**, 396 (1999).
- 8 C.M. Carlevaro, L. Blum and F. Vericat, *J. Chem. Phys.*, **119**, 5198 (2003).
- 9 H.J.C. Berendsen, J.R. Grigera and T.P. Straatsma, *J. Phys. Chem.*, **91**, 6269 (1987).
- 10 G. Nemethy and H.A. Scheraga, *J. Chem. Phys.*, **36**, 3382 (1962); **41**, 680 (1964).
- 11 A.E. Mirsky and L. Pauling, *Proc. Natl. Acad. Sci. USA*, **22**, 439 (1936).

12 W. Kauzmann, *Adv. Protein Chem.* **14**, 1 (1959).

13 T.E. Creighton (Editor), *Protein Folding* (W.H. Freeman and Company, New York, 1992).

14 H. Bohr and S. Brunak (Editores), *Protein Folds. A Distance-Based Approach* (CRS Press Inc., Boca Raton, Florida, 1996).

## **2. HERRAMIENTAS TEÓRICAS.**

A continuación exponemos algunos conceptos básicos relacionados con diversas técnicas teóricas, que utilizaremos en los próximos capítulos: simulaciones mediante Dinámica Molecular y una teoría mecánico–estadística desarrollada inicialmente para describir soluciones acuosas de solutos apolares.

Comenzamos introduciendo el concepto de Simulación Molecular (SM) con el que designamos en forma genérica a los dos métodos de cálculo más comunes, Monte Carlo (MC) y Dinámica Molecular (DM). Ambos permiten describir a los sistemas de muchas partículas desde los dos puntos de vista ergódicamente equivalentes: el probabilístico y el dinámico.

La primera parte de este capítulo está dedicada a dar una muy breve descripción de las características de la simulación mediante DM en términos generales. Muchos de los cálculos presentados en los capítulos siguientes fueron realizados utilizando el paquete de programas Gromos96 (GRONingen MOlecular Simulations) [1]. Algunos datos específicos de este software pueden encontrarse en el *Apéndice A* de esta Tesis.

La segunda parte introduce las ideas más importantes que subyacen en la teoría de los líquidos clásicos en general, y en el formalismo de Pratt y Chandler [2] en particular, basadas en el concepto de las funciones de correlación de pares.

### **2.1 SIMULACIONES MOLECULARES.**

La historia y el desarrollo de la Simulación Molecular, están íntimamente ligados a la historia y el desarrollo de las computadoras digitales. Después de la segunda guerra mundial, las computadoras estuvieron disponibles (por primera vez) para un espectro más amplio de investigadores, y uno de los primeros problemas en ser abordados por éstos fue la simulación numérica de líquidos densos.

La primera simulación de un líquido fue llevada a cabo por Metrópolis, Rosenbluth, Rosenbluth, Teller y Teller [3] sobre la computadora MANIAC en Los Álamos, introduciendo el método Monte Carlo. El nombre simulación Monte Carlo

fue dado por Metrópolis y Ulam, debido al intensivo uso que el método hace de los números al azar generados con la computadora.

La primera simulación con Dinámica Molecular fue realizada en 1957 por Alder y Wainwright [4], quienes estudiaron la dinámica de un sistema formado por esferas duras.

Si bien desde entonces se han realizado refinamientos y extensiones que permiten abordar problemas muy diversos usando simulaciones moleculares, los algoritmos básicos para MC y DM no han cambiado sustancialmente desde los años 50 [5, 6].

Para un modelo cuyo potencial entre partículas esté determinado a nivel de Born–Oppenheimer, las simulaciones moleculares permiten obtener un resultado que puede tomarse como exacto para problemas de Mecánica Estadística, respecto de las soluciones aproximadas a las que hay que recurrir cuando se los estudia desde el punto de vista teórico analítico, debido a la complejidad de la mayoría de los sistemas de interés. Esto es particularmente significativo cuando el sistema en estudio es una proteína inmersa en un medio acuoso. Ya comentamos sobre las dificultades que existen para desarrollar una teoría adecuada que describa al agua líquida, ya sea pura o en presencia de solutos simples. Es fácil imaginar cómo se multiplica el grado de dificultad si se consideran solutos más complejos como moléculas de aminoácidos o una proteína completa.

Las SM son técnicas esencialmente numéricas donde en cada paso podemos monitorear las posiciones (y velocidades en DM) de las distintas partículas o sitios que interactúan a través de cierto campo de fuerza. En este sentido las SM actúan como un “experimento” para el modelo, determinado por las interacciones presentes. La comparación de estos resultados de simulación con datos experimentales constituye una prueba (sin ambigüedades) de la bondad del modelo propuesto. Si surgieran discrepancias entre las medidas experimentales y los datos obtenidos a partir de SM, éstas se deberían, sin dudas, a fallas en el modelo utilizado para representar al sistema real en estudio.

Antes de que las computadoras se utilizaran para realizar simulaciones, la única manera de predecir las propiedades de un sistema molecular, era a partir de una teoría. Sin embargo, hay muy pocos sistemas en los que se pueden calcular

las propiedades de equilibrio en forma exacta. Para el resto, es inevitable tener que introducir aproximaciones. Como consecuencia de esto, muchas de las propiedades de sistemas reales debieron obtenerse sobre la base de teorías no exactas.

El conocimiento de las interacciones intermoleculares, aún cuando se consideren moléculas simples, es en general muy incompleto. Y aunque contáramos con suficiente información sobre las interacciones intermoleculares, esas teorías sólo permitirían una estimación más o menos buena de las propiedades de interés.

Esto plantea un difícil problema si pretendemos chequear la validez de una teoría particular a partir de la comparación directa con los datos experimentales. Si la teoría y los experimentos estuvieran en desacuerdo, podría ser porque la teoría es incorrecta o porque se tiene una estimación errónea de las interacciones intermoleculares, o por ambos motivos. Por lo tanto, sería muy bueno poder obtener resultados exactos para un dado modelo de sistema, sin tener que caer en teorías aproximadas. Las simulaciones numéricas hacen que esto sea posible.

Como dijimos, si comparamos las propiedades obtenidas en forma experimental con las calculadas a partir de simulación numérica, y observamos que no coinciden, entonces podemos decir que el modelo es inadecuado, y por lo tanto, debemos mejorar la estimación de las interacciones intermoleculares. Ahora bien, si al comparar los resultados de simulación surgidos de cierto modelo del sistema con las predicciones de una teoría aplicada a ese mismo modelo, encontráramos que la teoría y la simulación están en desacuerdo, concluiríamos que las aproximaciones usadas en la teoría no son válidas. En este caso, la simulación juega el rol de "*experimento*" diseñado para chequear la teoría.

Esta aplicación de la simulación es de una enorme utilidad e importancia. Tanto es así, que puede decirse que ha cambiado la manera en que se construyen nuevas teorías. En la actualidad, es raro que una teoría se utilice sin antes ser chequeada por una simulación.

En la figura 2.1 representamos este doble rol desempeñado por la simulación. En un caso como test del modelo frente a los resultados

experimentales, y en el otro, de las teorías frente al “experimento simulado” para un determinado modelo.

En los siguientes capítulos tratamos de aprovechar estas dos funciones de las SM y las usamos tanto para estudiar algunos aspectos estructurales de los aminoácidos disueltos en agua, como para verificar la bondad de los resultados obtenidos a partir de una teoría mecánico–estadísticas de líquidos clásicos.

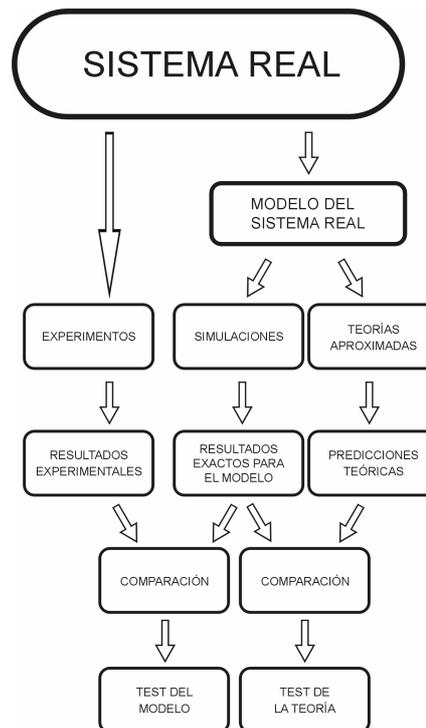


figura 2.1

Las SM permiten relacionar en forma directa los detalles microscópicos de un sistema con las propiedades macroscópicas de interés experimental. Por esta razón, resulta conveniente usar esta técnica para estudiar muchas propiedades fisicoquímicas de aminoácidos y proteínas en solución, no sólo por el indudable interés académico de este tipo de información, sino también por su potencial aplicación tecnológica.

En ciertos casos, se quieren conocer algunos aspectos del sistema en condiciones experimentalmente difíciles de alcanzar, como por ejemplo muy altas presiones; o de difícil acceso experimental, como el comportamiento de una molécula en un poro o en una membrana; o determinados detalles muy sutiles de la estructura o movimiento molecular que son muy difíciles de medir. En todos

estos casos las SM han demostrado ser una poderosa herramienta, que permite estudiar fenómenos, que de otra manera, sería imposible o muy costoso observar.

### **2.1.1 DINÁMICA MOLECULAR.**

Como mencionamos anteriormente, bajo la denominación Simulaciones Moleculares incluimos a las dos técnicas principales: Monte Carlo y Dinámica Molecular.

En el método de Monte Carlo se generan aleatoriamente una serie de configuraciones de todas las moléculas que constituyen el sistema en estudio. Estas configuraciones se eligen de manera que estén *pesadas* por el factor de Boltzmann y se calculan los promedios correspondientes de las magnitudes de interés que son funciones de esas configuraciones. En este sentido se trata de una versión numérica del enfoque de Gibbs basado en los “ensambles” estadísticos.

En cambio, en la Dinámica Molecular se obtienen las configuraciones y momentos (lineales y angulares) resolviendo las correspondientes ecuaciones de movimiento clásicas a partir de condiciones iniciales elegidas generalmente al azar. Es decir, se estudia la evolución temporal del sistema y luego se calculan los promedios temporales de las magnitudes de interés. Se trata, por lo tanto, de un enfoque dinámico más próximo al proyecto inicial de Boltzmann.

En principio, las ecuaciones diferenciales de movimiento con condiciones iniciales que plantea la DM, se pueden resolver usando cualquier algoritmo estándar de diferencias finitas. Sin embargo, en la práctica, muchos de los algoritmos comúnmente usados como el Runge–Kutta convencional, son computacionalmente muy costosos porque requieren múltiples evaluaciones de fuerzas.

La evaluación de las fuerzas es el paso que más tiempo consume en una simulación y cualquier algoritmo que requiera más de una evaluación de las fuerzas por paso temporal es inviable. Por este motivo, los algoritmos posibles forman un pequeño subconjunto, algunos de los cuales fueron desarrollados específicamente para ser usados en DM.

Un buen algoritmo debería cumplir con un conjunto de pautas como las siguientes:

- (a) ser rápido y requerir poca memoria,
- (b) permitir el uso de un paso temporal largo,
- (c) reproducir la trayectoria clásica en forma tan precisa como sea posible,
- (d) satisfacer las leyes de conservación para la energía y momentos,
- (e) ser reversible en el tiempo.

No todos los puntos anteriores son igualmente importantes para una DM. Aunque a primera vista la velocidad parece importante, en realidad no lo es, ya que el tiempo necesario para integrar las ecuaciones de movimiento en la mayoría de los sistemas, es pequeño comparado con el tiempo requerido en el cálculo de las interacciones.

Por eso, es más importante mantener la precisión para pasos temporales grandes, ya que cuanto mayores sean éstos, menos evaluaciones de fuerzas serán necesarias hasta alcanzar el tiempo total de simulación.

El problema es que cuando aumentamos el paso temporal  $\delta t$ , la trayectoria que siguen las moléculas se aleja de la correcta. En cualquier caso, es poco razonable pensar que un algoritmo pueda seguir indefinidamente la trayectoria clásica exacta. Cualquier perturbación por pequeña que parezca, aún el pequeño error asociado con la precisión aritmética finita, causará que la trayectoria surgida de la simulación, se aparte de la trayectoria clásica con la que inicialmente coincidía.

Afortunadamente esto no es crítico, ya que buscamos soluciones de las ecuaciones de movimiento que sean esencialmente exactas para tiempos comparables con el tiempo de correlación de interés, de modo que podamos calcular a la función de correlación temporal en forma precisa. Por tanto, no necesitamos trayectorias exactas, sino más bien, cuidar de que la energía se conserve.

Si incrementamos el paso temporal  $\delta t$ , resulta más difícil cumplir con la conservación de la energía total, y tendremos que encontrar una solución de compromiso entre velocidad y precisión para la simulación. El algoritmo ideal

sería aquel que permite un paso de tiempo grande, mientras conserva la energía en forma aceptable.

Por otra parte, las ecuaciones de Newton son reversibles en el tiempo y también deberían serlo los algoritmos. Sin embargo, si invirtiéramos los momentos de todas las partículas en un dado instante, el sistema no regresaría hacia atrás sobre sus pasos en el espacio de las fases repitiendo su trayectoria, aún cuando llevemos a cabo la simulación con una precisión numérica infinita. Éstos algoritmos sería reversibles sólo en el límite de un paso temporal  $\delta t$  infinitamente pequeño.

Un esquema simplificado de los pasos que sigue una DM, sería el siguiente:

(a) se leen los parámetros que especifican las condiciones de la corrida (temperatura inicial, número de partículas, densidad, paso temporal, etc.),

(b) se inicializa el sistema (posiciones y velocidades para todas las partículas),

(c) se realiza el cálculo de las fuerzas sobre todas las partículas,

(d) se integran las ecuaciones de movimiento de Newton. Este paso y el anterior constituyen el corazón de la simulación, y se repiten hasta que se haya calculado la evolución temporal del sistema durante la longitud total de tiempo deseado,

(e) finalmente, se calculan e imprimen los promedios de las cantidades medidas y se detiene la simulación.

Cuando este esquema se aplica a ciertos sistemas, como los líquidos simples formados por átomos, éstos actúan como entidades individuales que interactúan entre sí a través de potenciales centrales del tipo Lennard–Jones.

En el caso de moléculas, como el agua y los aminoácidos, existen dos enfoques posibles. Uno trata a cada molécula como una unidad. En este caso debemos caracterizar a sus configuraciones mediante la posición (por ejemplo dando las coordenadas de su centro de masa) y su orientación (definiendo ángulos adecuados, por ejemplo los ángulo de Euler). En esta descripción a nivel molecular debemos resolver el sistema de ecuaciones de movimiento para la posición del centro de masa y los ángulos de Euler dando como condiciones iniciales las posiciones, ángulos y velocidades de los centros de masas y los

momentos angulares. Este camino implica integrar numéricamente un sistema muy grande de ecuaciones que acopla posiciones y ángulos. Además, requiere el conocimiento de complicados potenciales de interacción a nivel de Born–Oppenheimer que dependen tanto de las posiciones como de las orientaciones.

El segundo enfoque es, por lejos, el que ha sido más utilizado para simular sistemas constituidos por moléculas complejas. Se trata de un enfoque atómico donde las unidades dinámicas individuales son los átomos que forman a las moléculas. Éstos, se mueven interactuando entre sí vía fuerzas puramente centrales que requieren sólo de sus posiciones, como en los líquidos atómicos. Pero ahora, hay que simular los enlaces covalentes de manera que en cada etapa del cálculo, las moléculas se reconstituyan con los átomos que corresponde, a las distancias y ángulos adecuados para conservar la identidad molecular.

En algunos casos, los enlaces covalentes se simulan agregando a los potenciales de interacción atómicos (entre los pares que forman parte del enlace) potenciales del tipo *oscilador armónico* con constantes suficientemente grandes. En general, en moléculas poliatómicas no lineales, es necesario considerar no sólo los efectos de estiramiento y compresión que pueden simularse con resortes, sino también movimientos de deformación que cambian los ángulos entre los enlaces y modifican los ángulos de torsión.

Las restricciones holonómicas implícitas en los enlaces covalentes pueden ser tratadas construyendo un sistema de coordenadas generalizadas que obedecen ecuaciones de movimiento libre en las cuales las restricciones aparecen en forma no explícita. Una subrutina numérica que realiza en forma muy eficiente esta dinámica con restricciones es la conocida con el nombre SHAKE [7].

En el próximo capítulo, utilizamos DM para estudiar un sistema compuesto por 1 aminoácido y 511 moléculas de agua dentro de una caja rectangular con las dimensiones adecuadas para que la solución esté a la presión deseada. La herramienta que utilizamos para esto es el programa GROMOS96 que maneja las restricciones holonómicas usando la subrutina SHAKE.

## 2.2 TEORÍAS MECÁNICO–ESTADÍSTICAS.

Además de las simulaciones mediante DM, utilizamos ciertas técnicas mecánico–estadísticas desarrolladas para estudiar sistemas continuos de muchas partículas. Cuando decimos sistemas continuos nos referimos a una descripción en donde las partículas pueden ocupar todas las configuraciones posibles de un espacio compacto, a diferencia de los estudios de sistemas en redes (donde ocupan sólo sitios discretos).

La mecánica estadística intenta describir el comportamiento macroscópico de un sistema constituido por muchas partículas a partir de sus interacciones microscópicas mediante un enfoque probabilístico. El comportamiento macroscópico a que nos referimos se resume en las magnitudes termodinámicas del sistema de interés.

En la mecánica estadística clásica las interacciones entre átomos y moléculas vienen dadas por potenciales al nivel de Born–Oppenheimer, es decir, a un nivel donde se supone que los electrones acompañan instantáneamente a los núcleos perdiendo su individualidad.

Consideremos un sistema compuesto de  $m$  especies de moléculas de las cuales  $N_i$  son del tipo  $i$  con  $(i = 1, 2, \dots, m)$  de manera que el número total de moléculas es  $N = \sum_{i=1}^m N_i$ .

Indicamos genéricamente con  $\vec{X}_{\alpha_i} \equiv (\vec{r}_{\alpha_i}, \vec{\Omega}_{\alpha_i})$  a la configuración de la molécula  $\alpha$  del tipo  $i$ . Aquí  $\vec{r}_{\alpha_i}$  denota la posición del centro de masa y  $\vec{\Omega}_{\alpha_i} \equiv (\alpha_{\alpha_i}, \beta_{\alpha_i}, \gamma_{\alpha_i})$  los ángulos de Euler que determinan su orientación. Para el caso de sistemas formados por átomos o moléculas esféricas, es suficiente establecer sólo la posición  $\vec{r}_{\alpha_i}$  del centro de masa.

La configuración del sistema queda especificada dando las configuraciones de todas las moléculas:  $\vec{X}^N \equiv (\vec{X}^{N_1}, \vec{X}^{N_2}, \dots, \vec{X}^{N_m})$  donde indicamos con  $\vec{X}^{N_i} \equiv (\vec{X}_1, \vec{X}_2, \dots, \vec{X}_{N_i})$  la configuración de todas las moléculas de la especie  $i$ .

Para que el modelo quede completamente determinado, tenemos que dar su energía potencial de interacción  $U(\vec{X}^N)$ . Para moléculas rígidas (no polarizables)

al nivel de Born–Oppenheimer (como las consideraremos en esta Tesis), es válida la siguiente aditividad de potenciales de pares:

$$U(\vec{X}^N) = \frac{1}{2} \sum_{i,j}^m \sum_{\alpha_i \neq \beta_j}^{N_i N_j} U_{ij}(\vec{X}_{\alpha_i}, \vec{X}_{\beta_j}) \quad (2.1)$$

donde  $U_{ij}(\vec{X}_{\alpha_i}, \vec{X}_{\beta_j})$  denota el potencial de interacción entre las moléculas  $\alpha$  de tipo  $i$  y  $\beta$  de tipo  $j$ .

Si consideramos que el sistema está en contacto con un reservorio de calor a temperatura  $T$  y se encuentra en un volumen  $V$  (densidades  $\rho_i = N_i/V$ ), la medida de probabilidad básica es la medida de Gibbs canónica [8, 9]:

$$dP(\vec{X}^N) = \frac{e^{-\frac{U(\vec{X}^N)}{k_B T}}}{Z(V, T, N_1, N_2, \dots, N_m)} d\vec{X}^N \quad (2.2)$$

donde  $k_B$  es la constante de Boltzmann y la función de partición configuracional canónica de la mezcla  $Z(V, T, N_1, N_2, \dots, N_m)$  es la constante de normalización:

$$Z(V, T, N_1, N_2, \dots, N_m) = \int_{V^N} e^{-\frac{U(\vec{X}^N)}{k_B T}} d\vec{X}^N \quad (2.3)$$

La relación con la termodinámica viene dada por el potencial termodinámico en función de sus variables características. Para el conjunto canónico es la energía libre de Helmholtz:

$$A(V, T, N_1, N_2, \dots, N_m) = -k_B T \ln \left[ \frac{Z(V, T, N_1, N_2, \dots, N_m)}{\prod_{i=1}^m (N_i! \Lambda_i^{3N_i})} \right] \quad (2.4)$$

donde  $\Lambda_i = \left( \frac{2\pi \hbar^2}{m_i k_B T} \right)^{1/2}$  es la longitud de onda térmica de de Broglie para partículas de la especie  $i$ . Aquí  $m_i$  indica la masa de cada una de las partículas de tipo  $i$ , y  $\hbar = h/2\pi$  está definida en términos de la constante de Planck  $h$ .

Para sistema aditivos como los que consideramos aquí, una forma alternativa de relacionar la termodinámica con el comportamiento a nivel microscópico del sistema, es a través de la función de distribución de pares que, en el sistema canónico, se define:

$$\rho_{ij}^{(2)}(\vec{X}_1, \vec{X}_2) = \int_{V^N} P(\vec{X}^N) \left[ \sum_{\alpha_i}^{N_i} \sum_{\beta_j}^{N_j} \delta(\vec{X}_{\alpha_i} - \vec{X}_1) \delta(\vec{X}_{\beta_j} - \vec{X}_2) \right] d\vec{X}^N \quad (2.5)$$

La relación de las funciones de distribución de pares  $\rho_{ij}^{(2)}(\vec{X}_1, \vec{X}_2)$  con la termodinámica se materializa en las siguientes expresiones para la energía interna  $E$ , la presión  $P$  y la compresibilidad isotérmica  $\chi_T$ :

$$E = E_C + \frac{1}{128\pi^4 V} \sum_{ij}^m \iint_V U_{ij}(\vec{X}_1, \vec{X}_2) \rho_{ij}^{(2)}(\vec{X}_1, \vec{X}_2) d\vec{X}_1 d\vec{X}_2 \quad (2.6)$$

$$P = k_B T \sum_{i=1}^m \rho_i - \frac{1}{384\pi^4 V} \sum_{ij}^m \iint_V [(\vec{r}_2 - \vec{r}_1) \cdot \vec{\nabla}_{(\vec{r}_2 - \vec{r}_1)} U_{ij}(\vec{X}_1, \vec{X}_2)] \rho_{ij}^{(2)}(\vec{X}_1, \vec{X}_2) d\vec{X}_1 d\vec{X}_2 \quad (2.7)$$

$$\chi_T = \frac{1}{k_B T \sum_{i=1}^m \rho_i} + \frac{1}{64\pi^4 k_B T V} \sum_{ij}^m \frac{1}{\rho_i \rho_j} \iint_V [\rho_{ij}^{(2)}(\vec{X}_1, \vec{X}_2) - \rho_i^{(1)}(\vec{X}_1) \rho_j^{(1)}(\vec{X}_2)] d\vec{X}_1 d\vec{X}_2 \quad (2.8)$$

En la ec. (2.6),  $E_C$  denota la parte cinética de energía interna, mientras que en (2.8) usamos la función de distribución de una partícula:

$$\rho_i^{(1)}(\vec{X}_1) = \int_{V^N} P(\vec{X}^N) \left[ \sum_{\alpha_i}^{N_i} \delta(\vec{X}_{\alpha_i} - \vec{X}_1) \right] d\vec{X}^N \quad (2.9)$$

que para un sistema homogéneo e isotrópico es constante:  $\rho_i^{(1)}(\vec{X}_1) = \rho_i / 8\pi^2$ .

## 2.2.1 ECUACIONES INTEGRALES Y TEORÍAS PERTURBATIVAS PARA LAS FUNCIONES DE CORRELACIÓN.

Las funciones de distribución de pares  $\rho_{ij}^{(2)}(\vec{X}_1, \vec{X}_2)$  ocupan un lugar especial en la mecánica estadística de sistemas continuos. Por un lado, miden la densidad de probabilidad de encontrar a dos moléculas, una de tipo  $i$  y otra de tipo  $j$  en las configuraciones  $\vec{X}_1$  y  $\vec{X}_2$ , respectivamente. De esta forma, brindan información a nivel microscópico sobre cómo se disponen las moléculas entre sí. Por otro lado, permiten obtener las funciones termodinámicas del sistema de manera relativamente sencilla.

Como comentamos anteriormente, estas funciones pueden obtenerse a partir de experiencias de difracción de rayos X o de neutrones, ya que se

relacionan con los factores de estructura parciales de la intensidad de difracción, mediante una transformada de Fourier.

Además, las funciones  $\rho_{ij}^{(2)}(\vec{X}_1, \vec{X}_2)$  se pueden calcular a partir de los potenciales de interacción  $U_{ij}(\vec{X}_1, \vec{X}_2)$  ( $i, j = 1, 2, \dots, m$ ). Esto puede hacerse mediante SM o dentro de la misma teoría mecánico–estadística, que es lo que más nos interesa aquí.

Las teorías mecánico–estadísticas que apuntan a obtener las funciones de distribución a partir de los potenciales de pares pueden clasificarse, en general, en dos grupos: ecuaciones integrales y teorías perturbativas.

### 2.2.1.1 Ecuaciones integrales.

Muchas de las ecuaciones integrales se basan en el sistema de relaciones de Ornstein–Zernike (OZ), que vinculan a las funciones de correlación directa  $c_{ij}(\vec{X}_1, \vec{X}_2)$  y total  $h_{ij}(\vec{X}_1, \vec{X}_2)$ :

$$h_{ij}(\vec{X}_1, \vec{X}_2) = c_{ij}(\vec{X}_1, \vec{X}_2) + \frac{1}{8\pi^2} \sum_{k=1}^m \rho_k \int_V c_{ik}(\vec{X}_1, \vec{X}_3) h_{kj}(\vec{X}_3, \vec{X}_2) d\vec{X}_3$$

( $i, j = 1, 2, \dots, m$ ) (2.10)

Las funciones de correlación total se definen como:

$$h_{ij}(\vec{X}_1, \vec{X}_2) = g_{ij}(\vec{X}_1, \vec{X}_2) - 1$$

(2.11)

donde  $g_{ij}(\vec{X}_1, \vec{X}_2)$  es la función de correlación entre pares  $ij$  definida por,

$$\rho_{ij}^{(2)}(\vec{X}_1, \vec{X}_2) = \rho_i^{(1)}(\vec{X}_1) \rho_j^{(1)}(\vec{X}_2) g_{ij}(\vec{X}_1, \vec{X}_2)$$

(2.12)

En la relación de Ornstein–Zernike se interpreta que la correlación total  $h_{ij}(\vec{X}_1, \vec{X}_2)$  entre dos partículas 1 y 2 (de tipo  $i$  y  $j$ ), puede descomponerse en un término  $c_{ij}(\vec{X}_1, \vec{X}_2)$  que da la correlación directa entre las ellas y otro término que da la correlación indirecta entre ambas, a través de terceras partículas del sistema que actúan como intermediarias.

Puede demostrarse [10] que  $c_{ij}(\vec{X}_1, \vec{X}_2)$  y  $h_{ij}(\vec{X}_1, \vec{X}_2)$  verifican la siguiente relación exacta:

$$c_{ij}(\vec{X}_1, \vec{X}_2) = -\frac{1}{k_B T} U_{ij}(\vec{X}_1, \vec{X}_2) + h_{ij}(\vec{X}_1, \vec{X}_2) - \ln[h_{ij}(\vec{X}_1, \vec{X}_2) + 1] + B_{ij}(\vec{X}_1, \vec{X}_2) \quad (2.13)$$

donde  $B_{ij}(\vec{X}_1, \vec{X}_2)$  es la suma de una serie de integrales que dependen de los potenciales interpartículas, que si bien resultan generalmente difíciles de calcular, están bien identificadas.

Reemplazando estas últimas expresiones en las relaciones de Ornstein–Zernike, se obtiene un sistema de ecuaciones integrales exacto para las  $h_{ij}(\vec{X}_1, \vec{X}_2)$  ( $i, j = 1, 2, \dots, m$ ).

Si en la relación exacta (2.13) ignoramos el término  $B_{ij}(\vec{X}_1, \vec{X}_2)$ , al reemplazar en las relaciones de OZ, obtenemos un sistema de ecuaciones integrales aproximado para las  $h_{ij}(\vec{X}_1, \vec{X}_2)$  conocido como hipercadena (HNC). Si en cambio nos quedamos con el primer sumando solamente, es decir, si tomamos  $c_{ij}(\vec{X}_1, \vec{X}_2) = -\frac{1}{k_B T} U_{ij}(\vec{X}_1, \vec{X}_2)$  tenemos la aproximación de Debye–Hückel. Puesto que para grandes separaciones entre las moléculas, las funciones de correlación directa van asintóticamente como los potenciales de pares:

$$\lim_{r \rightarrow \infty} c_{ij}(\vec{X}_1, \vec{X}_2) = -\frac{1}{k_B T} U_{ij}(\vec{X}_1, \vec{X}_2) \quad (2.14)$$

podemos decir que la aproximación de Debye–Hückel consiste en tomar como válido el valor asintótico para todo el rango de  $r$ .

Otra aproximación (relacionada con esta) es la conocida como aproximación esférica media (MSA) [11], que se aplica a potenciales con una parte repulsiva del tipo esferas duras, es decir, a potenciales de pares de la forma:

$$U_{ij}(\vec{X}_1, \vec{X}_2) = \begin{cases} \infty & r < \sigma_{ij} \\ U'_{ij}(\vec{X}_1, \vec{X}_2) & r > \sigma_{ij} \end{cases} \quad (2.15)$$

donde  $\sigma_{ij} = (\sigma_i + \sigma_j)/2$  siendo  $\sigma_i$  el diámetro de las esferas de tipo  $i$ . En este caso la impenetrabilidad de las esferas a distancias menores que  $\sigma_{ij}$  implica, dado su significado, que

$$h_{ij}(\vec{X}_1, \vec{X}_2) = -1 \quad r < \sigma_{ij} \quad (2.16)$$

La MSA aproxima  $c_{ij}(\vec{X}_1, \vec{X}_2)$  a toda distancia mayor que  $\sigma_{ij}$  según:

$$c_{ij}(\vec{X}_1, \vec{X}_2) = -\frac{1}{k_B T} U_{ij}(\vec{X}_1, \vec{X}_2) \quad r > \sigma_{ij} \quad (2.17)$$

Las relaciones (2.16) y (2.17) conjuntamente con las relaciones de OZ (2.10), definen un sistema de ecuaciones integrales mixtas, en el sentido de que las incógnitas son  $c_{ij}(\vec{X}_1, \vec{X}_2)$  para  $r < \sigma_{ij}$  y  $h_{ij}(\vec{X}_1, \vec{X}_2)$  para  $r > \sigma_{ij}$ .

Otra relación aproximada para cerrar a las relaciones de OZ, es la propuesta por Percus y Yevick [12]:

$$c_{ij}(\vec{X}_1, \vec{X}_2) = \left[ 1 - e^{\frac{1}{k_B T} U_{ij}(\vec{X}_1, \vec{X}_2)} \right] \left[ h_{ij}(\vec{X}_1, \vec{X}_2) + 1 \right] \quad (2.18)$$

Podemos notar que para un sistema de esferas duras (sin ningún tipo de atracción o repulsión adicional), es decir para un sistema cuyos potenciales de pares son:

$$U_{ij}(\vec{X}_1, \vec{X}_2) = \begin{cases} \infty & r_{12} < \sigma_{ij} \\ 0 & r_{12} > \sigma_{ij} \end{cases} \quad (2.19)$$

las aproximaciones MSA y Percus-Yevick coinciden.

### 2.2.1.2 Teorías perturbativas.

Las teorías perturbativas se basan esencialmente en la división del potencial de pares en una parte de referencia que denotaremos con el superíndice "(0)", y la perturbación que indicaremos con el superíndice "(1)":

$$U_{ij}(\vec{X}_1, \vec{X}_2) = U_{ij}^{(0)}(\vec{X}_1, \vec{X}_2) + U_{ij}^{(1)}(\vec{X}_1, \vec{X}_2; \gamma) \quad (2.20)$$

donde  $\gamma$  puede ser un parámetro de interacción (carga, momento dipolar, etc.) o un simple parámetro de orden que se hace 1 al final de los cálculos. Indicaremos con  $h_{ij}^{(0)}(\vec{X}_1, \vec{X}_2) = g_{ij}^{(0)}(\vec{X}_1, \vec{X}_2) - 1$  a las funciones de correlación de referencia correspondientes a los potenciales  $U_{ij}^{(0)}(\vec{X}_1, \vec{X}_2)$ .

De acuerdo a la elección que se haga del potencial de referencia y del parámetro de orden, se tienen diferentes teorías perturbativas. Como ejemplo, consideremos brevemente la teoría conocida como ORPA+B<sub>2</sub> (optimized random phase approximation + B<sub>2</sub>) debida a Andersen y Chandler [13]. Partiendo de la

expresión (2.4) para la energía libre de Helmholtz y reemplazando en la expresión (2.3) para la función de partición configuracional, el potencial total por su forma aditiva (2.1) puede escribirse:

$$-\frac{A}{k_B TV} = -\frac{A_0}{k_B TV} + \frac{A_\lambda}{k_B TV} + \frac{A_{anillos}}{k_B TV} + B_2 + \dots \quad (2.21)$$

donde  $A_0$  (la parte ideal de la energía libre),  $A_\lambda$  (llamada aproximación de altas temperaturas),  $A_{anillos}$  y  $B_2$  vienen dados por:

$$\frac{A_0}{k_B TV} = \frac{1}{V} \ln \prod_{i=1}^m [N_i! \Lambda_i^{3N}] \quad (2.22)$$

$$\frac{A_\lambda}{k_B TV} = -\frac{1}{128\pi^4 k_B T} \sum_{ij}^{N_i N_j} \rho_i \rho_j \int_V g_{ij}^{(0)}(\vec{X}_1, \vec{X}_2) U_{ij}^{(1)}(\vec{X}_1, \vec{X}_2) d\vec{X}_1 d\vec{X}_2 \quad (2.23)$$

$$\frac{A_{anillos}}{k_B TV} = -\frac{1}{64\pi^4 k_B T} \sum_{n=2}^{\infty} \frac{1}{n} \sum_{ij}^{N_i N_j} \rho_i \rho_j \iint_V c_{ij}^{(n-1)}(\vec{X}_1, \vec{X}_2) U_{ij}^{(1)}(\vec{X}_1, \vec{X}_2) d\vec{X}_1 d\vec{X}_2 \quad (2.24)$$

$$B_2 = \frac{1}{256\pi^4} \sum_{ij}^{N_i N_j} \rho_i \rho_j \iint_V h_{ij}^{(0)}(\vec{X}_1, \vec{X}_2) [c_{ij}(\vec{X}_1, \vec{X}_2)]^2 d\vec{X}_1 d\vec{X}_2 \quad (2.25)$$

$$+ \frac{1}{128\pi^4} \sum_{ij}^{N_i N_j} \rho_i \rho_j \iint_V g^{(0)}(\vec{X}_1, \vec{X}_2) \sum_{n=3}^{\infty} \frac{1}{n!} [c_{ij}(\vec{X}_1, \vec{X}_2)]^n d\vec{X}_1 d\vec{X}_2$$

En la ecuación (2.24) consideramos:

$$\rho_i \rho_j c_{ij}^{(n)}(\vec{X}_1, \vec{X}_2) = \left(-\frac{1}{k_B T}\right)^n \times \sum_{kl\dots s}^{N_k N_l \dots N_s} \iiint_V \dots \int F_{ik}^{(0)}(\vec{X}_1, \vec{X}_3) U_{kl}^{(1)}(\vec{X}_3, \vec{X}_4) F_{lm}^{(0)}(\vec{X}_4, \vec{X}_5) \dots U_{rs}^{(1)}(\vec{X}_{2n-1}, \vec{X}_{2n}) F_{sj}^{(0)}(\vec{X}_{2n}, \vec{X}_2) d\vec{X}_3 d\vec{X}_4 \dots d\vec{X}_{2n} \quad (2.26)$$

donde  $F_{ij}^{(0)}(\vec{X}_1, \vec{X}_2)$  son los llamados hipervértices de referencia definidos por:

$$F_{ij}^{(0)}(\vec{X}_1, \vec{X}_2) = \rho_i \delta(\vec{X}_1, \vec{X}_2) + \rho_i \rho_j h_{ij}^{(0)}(\vec{X}_1, \vec{X}_2) \quad (2.27)$$

y  $\delta$  simboliza a la delta de Dirac. En la ec. (2.25) consideramos la suma:

$$c_{ij}(\vec{X}_1, \vec{X}_2) = \sum_{n=1}^{\infty} c_{ij}^{(n)}(\vec{X}_1, \vec{X}_2) \quad (2.28)$$

denominada *potencial renormalizado* o *cadena de enlaces*.

A continuación consideramos la correspondiente expresión perturbativa de las funciones de correlación  $g_{ij}(\vec{X}_1, \vec{X}_2) = h_{ij}(\vec{X}_1, \vec{X}_2) + 1$ . Para eso, debemos tener

presente que la función de distribución de pares puede escribirse como la derivada funcional:

$$\rho_{ij}^{(2)}(\vec{X}_1, \vec{X}_2) = 2 \frac{\delta[A/V]}{\delta U_{ij}^{(1)}(\vec{X}_1, \vec{X}_2)} \quad (2.29)$$

Luego, de las ecs. (2.21), (2.25) y (2.28) y efectuando la derivación funcional indicada en (2.29), obtenemos:

$$g_{ij}(\vec{X}_1, \vec{X}_2) = g_{ij}^{(0)}(\vec{X}_1, \vec{X}_2) + C_{ij}(\vec{X}_1, \vec{X}_2) + h_{ij}^{(0)}(\vec{X}_1, \vec{X}_2) C_{ik}(\vec{X}_1, \vec{X}_2) + \dots \quad (2.30)$$

que puede ponerse:

$$g_{ij}(\vec{X}_1, \vec{X}_2) \approx g_{ij}^{(0)}(\vec{X}_1, \vec{X}_2) e^{C_{ij}(\vec{X}_1, \vec{X}_2)} \quad (2.31)$$

### 2.2.2 TEORÍA DE PRATT Y CHANDLER.

Finalizamos este capítulo, describiendo brevemente la teoría que desarrollaron Pratt y Chandler [2] para estudiar soluciones diluidas de solutos apolares en agua.

En su versión más simple, todas las moléculas que forman al sistema son consideradas esféricas, y la interacción entre cada par de moléculas es modelada por potenciales centrales que dependen sólo de la distancia entre ellas.

La teoría trabaja con las funciones de correlación de pares *soluto-agua*  $h_{sa}(r)$ , *soluto-soluto*  $h_{ss}(r)$  y *agua-agua*  $h_{aa}(r)$ , dependientes sólo de la distancia  $r = |\vec{r} - \vec{r}'|$  entre moléculas.

Las correlaciones  $h_{sa}(r)$  y  $h_{ss}(r)$  se obtienen de las relaciones mecánico-estadísticas del tipo OZ, mientras que la función  $h_{aa}(r)$  es extraída directamente de experimentos de difracción con agua pura. De ésta forma se evitan las complicaciones que aparecen al introducir una descripción teórica de agua líquida, puesto que hasta el momento, ninguna resulta ser completamente satisfactoria.

De esta manera, se adopta una estrategia *semiempírica* en la cual las interacciones entre moléculas de soluto, y entre éstas y el agua son consideradas explícitamente. Las correspondientes a agua pura quedan implícitas en la función de distribución radial obtenida experimentalmente.

Luego se elige como modelo de las interacciones dependientes sólo de las distancias entre las moléculas que componen al sistema formado por la mezcla binaria (soluto apolar (s) y agua (a)), al potencial de Lennard–Jones (figura 2.2a):

$$U_{sM}(r) = 4\epsilon_{sM} \left[ \left( \frac{\sigma_{sM}}{r} \right)^{12} - \left( \frac{\sigma_{sM}}{r} \right)^6 \right] \quad (2.32)$$

donde  $M$  representa a una molécula de agua o de soluto ( $M = a$  o  $M = s$ ).

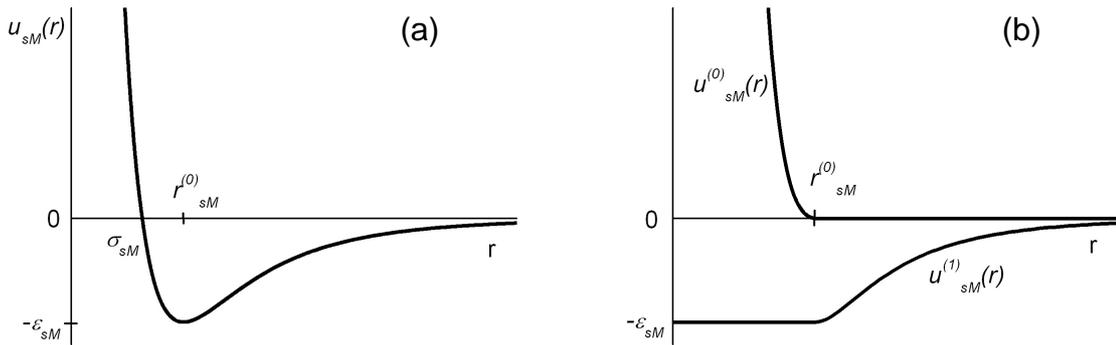


figura 2.2

Estos potenciales son separados (figura 2.2b) en una parte exclusivamente repulsiva  $U_{sM}^{(0)}(r)$  y otra puramente atractiva  $U_{sM}^{(1)}(r)$ , de modo tal que el potencial total sea igual a la suma de ambos,

$$U_{sM}^{(0)}(r) = \begin{cases} \epsilon_{sM} + U_{sM}(r) & r < r_{sM}^{(0)} \\ 0 & r > r_{sM}^{(0)} \end{cases} \quad (2.33a)$$

$$U_{sM}^{(1)}(r) = \begin{cases} -\epsilon_{sM} & r < r_{sM}^{(0)} \\ U_{sM}(r) & r > r_{sM}^{(0)} \end{cases} \quad (2.33b)$$

En esta teoría, se utilizan ecuaciones integrales basadas en relaciones del tipo Ornstein–Zernike para mezclas, con el objeto de calcular las funciones de correlación de referencia. Las funciones de correlación radial de referencia  $g_{sM}^{(0)}(r)$  corresponden al sistema en el cual se desprecia el término atractivo, esto es, las moléculas de soluto y de agua interactúan a través de los potenciales  $U_{sM}(r) = U_{sM}^{(0)}(r)$  definidos en la ec. (2.33a).

Una vez determinada la función de referencia, de la ec. (2.31) tenemos:

$$g_{sM}(r) \approx g_{sM}^{(0)}(r) e^{C_{sM}(r)} \quad (2.34)$$

donde  $C_{sM}(r)$  es el potencial renormalizado dependiente de  $U_{sM}^{(1)}(r)$ .

Las funciones  $g_{sM}^{(0)}(r) = h_{sM}^{(0)}(r) + 1$  pueden calcularse en forma aproximada utilizando un sistema de ecuaciones integrales determinado por las relaciones del tipo de OZ (2.10). Para el caso de sistemas infinitamente diluidos ( $\rho_s \approx 0$ ):

$$h_{sa}^{(0)}(r) = c_{sa}^{(0)}(r) + \rho_a \int c_{sa}^{(0)}(|\bar{r} - \bar{r}'|) h_{aa}^{\text{exp}}(r') d\bar{r}' \quad (2.35a)$$

$$h_{ss}^{(0)}(r) = c_{ss}^{(0)}(r) + \rho_a \int c_{sa}^{(0)}(|\bar{r} - \bar{r}'|) h_{sa}^{(0)}(r') d\bar{r}' \quad (2.35b)$$

Este sistema se cierra mediante las condiciones aproximadas:

$$h_{sa}^{(0)}(r) = -1, \quad r < \sigma_{sa} \quad (2.36a)$$

$$h_{ss}^{(0)}(r) = -1, \quad r < \sigma_{ss} \quad (2.36b)$$

y

$$c_{sa}^{(0)}(r) = 0, \quad r > \sigma_{sa} \quad (2.37a)$$

$$c_{ss}^{(0)}(r) = 0, \quad r > \sigma_{ss} \quad (2.37b)$$

Para la función de correlación  $h_{aa}^{\text{exp}}(r)$  en la ec. (2.35a) se usa la transformada de Fourier del factor de estructura parcial *oxígeno-oxígeno* obtenido a partir de experimentos de difracción con agua pura.

En consecuencia, se obtiene un sistema de ecuaciones integrales cuyas incógnitas son  $c_{sa}^{(0)}(r)$  y  $c_{ss}^{(0)}(r)$  dentro de la región de impenetrabilidad, y  $h_{sa}^{(0)}(r)$  y  $h_{ss}^{(0)}(r)$  afuera. Una vez determinadas, podemos introducir las funciones de correlación de pares  $g_{sa}^{(0)}(r) = h_{sa}^{(0)}(r) + 1$  y  $g_{ss}^{(0)}(r) = h_{ss}^{(0)}(r) + 1$  en las expresiones perturbativas (2.34) o alguna otra equivalente, para obtener  $g_{sa}(r)$  y  $g_{ss}(r)$ .

En los próximos capítulos aplicaremos este esquema (con algunas variantes) a aminoácidos en agua, resolviendo el sistema de ecuaciones integrales involucrado.

## 2.3 REFERENCIAS DEL CAPÍTULO.

- 1 W.F. van Gunsteren, S.R. Billeter, A.A. Eising, W.R.P. Scott, and I.G. Tironi, *Biomolecular Simulation: The Gromos96 Manual and User Guide* (Zürich: vdf Hochschulverlag AG an der ETH Zürich, 1996).
- 2 L.R. Pratt and D. Chandler, *J. Chem. Phys.*, **67** 3683 (1977).
- 3 N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.N. Teller and E. Teller, *J. Chem. Phys.*, **21**, 1087-1092 (1953).
- 4 B.J. Alder and T.E. Wainwright, *In I. Prigogine, editor, Proc. of the Symp. on statistical mechanical theory of transport processes*, 97-131. (Interscience, Wiley, New York, 1958).
- 5 M.P. Allen and D.J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, 1987).
- 6 D. Frenkel, and B. Smit, *Understanding Molecular Simulation. From Algorithms to Applications* (Academic Press, San Diego, 1996). J.R. Sadus, *Molecular Simulation of Fluids, Theory, Algorithms and Object-Oriented* (ELSEVIER Science B.V, 1999).
- 7 J.P. Rickaert, G. Cicotti and H.J.C. Berendsen, *J. Comput. Phys.*, **23**, 327 (1977).
- 8 J.P. Hansen and I.R. McDonald, *Theory of Simple Liquids*, (Academic Press, London, 1986).
- 9 A. Ben-Naim, *Water and Aqueous Solutions* (Plenum Press, New York, 1974).
- 10 J.M.J. Van Leeuwen, J. Groeneveld and J. De Boer, *Physica*, **25**, 792 (1959).
- 11 J.L. Lebowitz and J.K. Percus, *Phys. Rev.*, **144**, 251 (1966).
- 12 J.K. Percus and G.J. Yevick, *Phys. Rev.*, **110**, 1 (1958).
- 13 H.C. Andersen and D. Chandler, *J. Chem. Phys.*, **57**, 1918 (1972).

### **3. UN RESIDUO DE AMINOÁCIDO APOLAR EN AGUA: HIDRATACIÓN HIDROFÓBICA.**

Como sabemos, las condiciones del medio cumplen un rol fundamental en la estabilidad de la estructura nativa. Cuando se modifican las condiciones fisiológicas normales, las proteínas sencillamente adquieren conformaciones diferentes. Esto se observa, muy claramente, en los experimentos donde se induce la desnaturalización de una proteína y su posterior renaturalización, justamente a través de cambios introducidos al medio [1].

Considerando que los aminoácidos son los eslabones fundamentales con los que se construyen todas las cadenas poliméricas que conocemos como proteínas, estudiar sus propiedades termodinámicas en solución acuosa, y determinar cómo se estructura el solvente alrededor de los átomos de los diferentes residuos, puede aportar información importante para comprender mejor algunos aspectos del problema del plegamiento y estabilidad del estado nativo.

En este capítulo abordamos estos tópicos en relación con los aminoácidos apolares. En la primera parte, presentamos algunos resultados obtenidos mediante simulación con Dinámica Molecular. Luego, introducimos un modelo teórico con el cual estudiamos las correlaciones residuo-agua.

#### **3.1 SIMULACIÓN MEDIANTE DINÁMICA MOLECULAR.**

El sistema en estudio estará compuesto, en todos los casos, por una molécula de soluto (el aminoácido completo o solamente la cadena lateral del mismo) y 511 moléculas de solvente (agua o metanol). Algunos detalles de las simulaciones, y los valores asignados a los parámetros de entrada, son recopilados en el *Apéndice B*.

##### **3.1.1 ENERGÍA LIBRE DE TRANSFERENCIA.**

Los aminoácidos pueden clasificarse de acuerdo a su polaridad en tres grupos: apolares, polares y cargados (Tabla 1.1). Aquí estudiaremos a los aminoácidos *apolares* o *no polares*. Dentro de este grupo, incluimos a siete de los

veinte aminoácidos que existen en la naturaleza: *alanina* (A), *valina* (V), *prolina* (P), *isoleucina* (I), *metionina* (M), *leucina* (L) y *fenilalanina* (F) [2].

De las observaciones realizadas sobre proteínas reales se encuentra que los aminoácidos apolares muestran una tendencia a situarse en el interior de las mismas, mientras que los pertenecientes a los otros dos grupos se hallan generalmente expuestos al solvente (agua).

Este hecho condujo a W. Kauzmann [3] a proponer a la interacción hidrofóbica como “la” fuerza responsable del colapso hacia la estructura compacta que las proteínas adquieren en el estado nativo. Si bien el origen preciso de la energía libre asociada con la interacción hidrofóbica sigue siendo controversial [4, 5, 6], a bajas temperaturas, y a temperaturas fisiológicas, se observa que la variación de energía libre es fundamentalmente de origen entrópico [7].

Luego del trabajo de Kauzmann, muchos investigadores intentaron describir aspectos del plegamiento de proteínas cuantitativamente en términos de la *hidrofobicidad* relativa, o del carácter hidrofóbico de los distintos residuos de aminoácidos. Esto despertó un gran interés por construir una escala de hidrofobicidad para los aminoácidos, que permitiera relacionar algunos aspectos de la estructura o función de las proteínas [8].

Si la interacción entre los aminoácidos apolares y cierto solvente fuera favorable con respecto al agua, esperaríamos que en dicho solvente la proteína se despliegue y adopte una conformación más abierta que la nativa. Por esta razón, es interesante estudiar la capacidad que poseen los distintos solventes para inducir la desnaturalización de proteínas [9], y determinar la preferencia que cada aminoácido tiene a interactuar con estos solventes en relación al agua.

Aparecieron así, muchas escalas de hidrofobicidad [10, 11], algunas de las cuales reproducimos en la Tabla 3.1. En general, observamos una dispersión muy grande en los valores obtenidos por los distintos autores. Entre ellas existe una diferencia tan importante, que en algunos casos, llegan a presentar signos opuestos para la variación de energía libre, lo que conduce a interpretaciones antagónicas con respecto a la hidrofobicidad de los residuos.

David Eisenberg y colaboradores [11], proponen una escala de consenso construida a partir de un promedio ponderado de escalas experimentales

existentes en la literatura. En la Tabla 3.1, la escala de Nozaki y Tanford (NT) modificada por Segrest y Feldman (SF) [8, 9], está basada sobre la  $\Delta G_t$  que se obtiene al transferir los residuos desde etanol a agua. Janin [12] y Chothia [13] obtienen sus escalas, calculando  $\Delta G_t$  al transferir el residuo desde el interior de la proteína a la superficie de la misma. La escala de Wolfenden [14] y colaboradores, considera la transferencia de las cadenas laterales de los aminoácidos desde el vapor de agua al agua en estado líquido.

En la Tabla 3.1, mostramos también la escala semiempírica de von Heijne y Blomberg [15]. Estos autores consideran el cambio de energía libre producido al transferir los residuos desde un medio no acuoso a uno acuoso.

Una revisión sobre el debate acerca de qué valores resultarían más (o menos) convenientes utilizar en el problema del plegamiento de proteínas, puede encontrarse en el trabajo de Lazaridis y colaboradores [16].

Tabla 3.1 : Escala de Hidrofobicidad.

Residuo	$\Delta G_{M \rightarrow W}$ [Kcal mol <sup>-1</sup> ]					
	Nozaki– Tanford <sup>[8,9]</sup>	von Heijne– Blomberg <sup>[15]</sup>	Janin <sup>[12]</sup>	Chothia <sup>[13]</sup>	Wolfenden <sup>[14]</sup>	Consenso <sup>[11]</sup>
ala	1.0	2.9	0.3	-0.29	1.94	0.25
pro	1.5	1.1	-0.3	-0.90	-	-0.07
met	2.5	2.1	0.4	-0.24	-1.48	0.26
val	3.0	3.9	0.6	0.09	1.99	0.54
leu	3.5	4.2	0.5	-0.12	2.28	0.53
ile	5.0	4.4	0.7	0.24	2.15	0.73
phe	5.0	5.2	0.5	0.00	-0.76	0.61

Considerando el interés que sin dudas tiene contar con una escala de hidrofobicidad “confiable”, calculamos la diferencia de energía libre ( $\Delta G_t$ ) que se obtiene al transferir el residuo de un aminoácido desde un solvente no polar (metanol) a agua, haciendo una simulación mediante Dinámica Molecular (DM) de este sistema.

Cuando decimos confiable nos referimos a la posibilidad que ofrece la simulación de controlar las variables del sistema en estudio, puesto que, en general, las diferencias encontradas entre las escalas experimentales puede atribuirse a las condiciones en las cuales se realizaron esas mediciones, los solventes utilizados en cada caso, etc.

Estos cálculos son muy importantes en el estudio de la estabilidad de la estructura nativa, ya que se ha sugerido que el etanol puede utilizarse como modelo para representar el interior de una proteína. Por lo tanto,  $\Delta G_t$  entre etanol y agua representaría una medida de lo desfavorable que resulta para un residuo de aminoácido apolar, la interacción con el agua.

Experimentalmente se observa que la variación de energía libre entre la cadena lateral de un aminoácido apolar y el solvente, difiere muy poco cuando se usa metanol, etanol, butanol o acetona [9].

De acuerdo con la propuesta de Tanford [9, 17], la  $\Delta G_t$  de un aminoácido desde solvente orgánico a agua, permite establecer una escala de hidrofobicidad, y a partir de ella, estudiar las contribuciones que los residuos encerrados en el interior de la proteína hacen a la estabilidad de la misma. Ahora bien, para que esta equivalencia sea válida ( $\Delta G_t \equiv$  hidrofobicidad) se deben verificar dos condiciones: (1)  $\Delta G_t$  debe favorecer al medio no acuoso, y (2) debe ser independiente del solvente no polar utilizado [9]. Si bien estas condiciones sólo se cumplen con los aminoácidos apolares [5], en muchos de los trabajos que siguieron a NT, se consideró válida esta equivalencia (equivocadamente) para todos los aminoácidos, incluyendo a los de tipo polar. A pesar de que NT dicen claramente que una escala de hidrofobicidad puede construirse, basada en el cálculo de la  $\Delta G_t$ , sólo para aminoácidos apolares, desafortunadamente en una tabla de su trabajo (tabla IX de [9]) incluyen también aminoácidos polares, lo que pudo ser el origen de la confusión posterior.

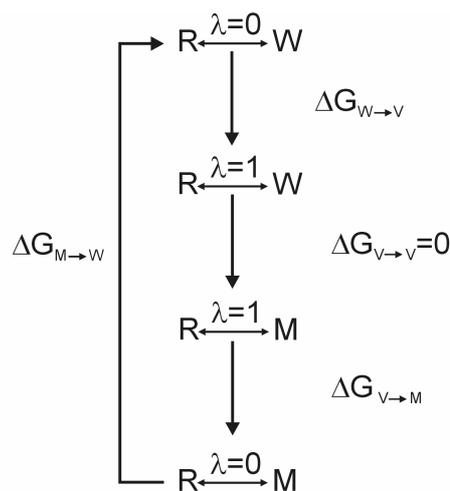
A continuación presentamos la escala de hidrofobicidad que obtuvimos mediante simulación para los siete aminoácidos apolares (alanina, valina, prolina, isoleucina, metionina, leucina, fenilalanina), y señalamos algunas de sus implicancias.

Luego mostramos un gráfico donde se observa la relación lineal encontrada originalmente por Cyrus Chothia, conjuntamente con la surgida de simulación, y discutimos sobre el valor de pendiente obtenido.

### 3.1.1.1 Escala de hidrofobicidad.

Para ser soluble en agua, una molécula debe ocupar un cierto volumen, y por tanto debe romper la estructura del agua al menos dentro de ese volumen. El volumen que ocupa dicha molécula en solución, refleja no sólo su volumen atómico, sino también, cualquier cambio que se produzca al reacomodar las moléculas de agua alrededor de ella.

La solubilidad de una molécula apolar en agua, depende de lo desfavorable que resulte crear una cavidad en relación con lo favorable de las interacciones que posea con las moléculas de agua que la rodean. Una medida de esta propiedad está dada por la hidrofobicidad que presenta cada molécula.



$$\Delta G_{M \rightarrow W} + \Delta G_{W \rightarrow V} + \Delta G_{V \rightarrow V} + \Delta G_{V \rightarrow M} = 0$$

figura 3.1

Para establecer nuestra escala, llevamos a cabo el cálculo de la variación de energía libre que se produce al transferir a cada uno de los residuos de aminoácidos apolares, desde el solvente no polar (metanol) al agua ( $\Delta G_{M \rightarrow W}$ ). Para esto utilizamos el paquete de programas Gromos96 [18], y consideramos un ciclo termodinámico como el de la figura 3.1.

En este esquema,  $\Delta G_{W \rightarrow V}$  representa la diferencia de energía libre que se produce al transferir el residuo que inicialmente está disuelto en agua (W) formando parte del aminoácido, al vacío (V). Luego,  $\Delta G_{M \rightarrow W} = -\Delta G_{W \rightarrow V} - \Delta G_{V \rightarrow M}$ .

Para calcular la diferencia de energía libre  $\Delta G_{W \rightarrow V}$  ( $\Delta G_{V \rightarrow M}$ ), hay dos técnicas que son las comúnmente más usadas: integración termodinámica (IT) y crecimiento lento (CL) [19, 20] (ver *Apéndice A*).

En la técnica de IT, la interacción soluto–solvente es controlada por un parámetro  $\lambda$  que toma valores entre 0 (interacción completa) y 1 (sin interacción). Se realizan simulaciones para valores fijos de  $\lambda$ , y se calcula la variación de energía libre integrando en  $\lambda$ , mediante algún algoritmo de integración estándar.

Por otro lado, en el método de CL, el sistema es *cuasiestáticamente* transformado desde un estado en el cual la interacción es completa ( $\lambda = 0$ ), a otro en el cual los átomos del residuo y del solvente no interactúan ( $\lambda = 1$ ) [20, 21].

Evaluamos  $\Delta G_{M \rightarrow W}$  para el residuo de alanina usando ambos métodos. En el caso de la IT, fue necesario previamente determinar la ineficiencia estadística, de modo de asegurar que los puntos tomados para los cálculos fueran independientes [22]. Encontramos que para garantizar la independencia estadística de los datos, es suficiente tomar una muestra cada 0.2 ps (100 pasos de simulación).

Para la integración numérica sobre  $\lambda$  utilizamos el método de Gauss–Lobatto [23]. Consideramos 3, 5, 7 y 9 puntos de integración (valores de  $\lambda \in [0,1]$ ) y observamos que para 7 o más puntos, el resultado estaba dentro del error estadístico. Mediante IT con 9 puntos de integración, obtuvimos para el residuo de alanina un valor de energía  $\Delta G_{M \rightarrow W} = 0.55 \text{ Kcal mol}^{-1}$ . En este método los errores no pueden ser determinados con precisión, y sólo puede darse una cota superior indicativa [19, 24]. En nuestro caso son del mismo orden de magnitud que los valores medidos.

En cuanto a CL, realizamos simulaciones en ambos sentidos, esto es, partimos de una configuración inicial de equilibrio con todas las interacciones presentes ( $\lambda = 0$ ), y hacemos evolucionar al sistema lentamente hasta un estado final donde los átomos pertenecientes al residuo del aminoácido en estudio, cesan

por completo su interacción con los demás átomos del sistema ( $\lambda = 1$ ). Luego, tomamos ésta última configuración correspondiente al estado sin interacción ( $\lambda = 1$ ) como la inicial, y vamos “encendiéndolas” lentamente hasta un estado en el cual todas las interacciones están nuevamente presentes ( $\lambda = 0$ ).

Esto, que podríamos interpretar como una “histéresis” definida por el “apagado” y posterior “encendido” de las interacciones de los átomos del residuo, puede tomarse como una medida aproximada de la incerteza en la variación de energía libre calculada.

Para el residuo de alanina, obtenemos  $\Delta G_{V \rightarrow M} = (-0.68 \pm 0.01) \text{Kcal mol}^{-1}$  en metanol, y  $\Delta G_{W \rightarrow V} = (0.07 \pm 0.02) \text{Kcal mol}^{-1}$  en agua. Finalmente, expresamos la energía libre de transferencia ( $\Delta G_{M \rightarrow W} = -\Delta G_{W \rightarrow V} - \Delta G_{V \rightarrow M}$ ) para el residuo de alanina como:  $\Delta G_{M \rightarrow W} = (0.61 \pm 0.03) \text{Kcal mol}^{-1}$ .

El método de CL resultó ser más eficiente en cuanto a tiempo y memoria para calcular  $\Delta G_{M \rightarrow W}$ , y además nos permitió una mejor estimación del error que el de IT. Teniendo en cuenta que ambos métodos dan resultados muy similares (al menos para la alanina), para los demás residuos apolares decidimos utilizar el método de CL.

Tabla 3.2: Escala de hidrofobicidad obtenida mediante simulación.

aminoácido apolar	$\Delta G_{M \rightarrow W} [\text{Kcal mol}^{-1}]$	$E_r(\%)$
ala	( 0.61 $\pm$ 0.03 )	5 %
val	( 1.7 $\pm$ 0.2 )	12 %
pro	( 2.0 $\pm$ 0.2 )	10 %
ile	( 3.3 $\pm$ 0.5 )	17 %
met	( 3.5 $\pm$ 0.2 )	6 %
leu	( 3.5 $\pm$ 0.6 )	18 %
phe	( 4.4 $\pm$ 0.1 )	2 %

En todos los casos las corridas se hicieron con crecimiento lento, tanto para agua como para metanol. El factor  $\lambda$  se hace variar de 0 a 1 en 500 ps de simulación. Esto significa que, para el cálculo de  $\Delta G_t$  en cada uno de los solventes, se necesitan 1000 ps (500 ps para el apagado, y 500 ps más para el posterior encendido de las interacciones). Por tanto, el valor final para cada aminoácido, se obtiene luego de 2000 ps (2 ns) de simulación.

Desde luego, cuanto mayor es  $\Delta G_{M \rightarrow W}$ , mayor es la hidrofobicidad. En la Tabla 3.2, presentamos la escala que obtenemos mediante simulación, en orden de hidrofobicidad creciente para los siete aminoácidos no polares considerados.

De acuerdo a NT [9], en el caso de los aminoácidos apolares, la  $\Delta G_t$  no varía sustancialmente si se utiliza metanol, etanol o cualquier otro solvente orgánico. Por el solvente orgánico utilizado, podría esperarse que nuestros resultados de la Tabla 3.2 muestren un mejor acuerdo con los datos de NT (columna 2 de Tabla 3.1) que con el resto. Sin embargo, más allá de algunas coincidencias como en el caso de *leucina*, ni los valores propiamente dichos, ni el orden que guardan los residuos en relación con su energía, acuerdan como se esperaba que lo hicieran.

No obstante, el orden mostrado por los residuos en la Tabla 3.2 resulta razonable si consideramos el número de átomos que integra cada una de las cadenas laterales de los diferentes aminoácidos, y el área expuesta al agua por cada uno de ellos (ver apartado siguiente 3.1.1.2).

Además, tengamos en cuenta que el cálculo de las energías presentadas en la Tabla 3.2, se llevó a cabo en condiciones diferentes a las descritas por NT. En nuestros cálculos (ver *Apéndice B*), consideramos un sistema compuesto por 1 aminoácido más 511 moléculas de solvente, y determinamos la variación en la energía libre que se produce al apagar o encender las interacciones de los átomos de la cadena lateral con los demás átomos del sistema. Mientras que las energías libres de transferencias obtenidas por NT resultan de restar a los valores medidos para cada aminoácido completo, el correspondiente a la glicina (gly).

Sería interesante quizás, contar con medias experimentales más recientes y, teniendo en cuenta el control de los diferentes parámetros que la DM permite, repetir la simulación en iguales condiciones para una comparación más directa.

### **3.1.1.2 Relación lineal de Chothia.**

En un conocido artículo, C. Chothia [25] encontró una relación lineal entre la variación de energía libre de transferencia (hidrofobicidad), y el área superficial accesible a la molécula de agua para algunos residuos de aminoácidos.

El área superficial accesible de un grupo atómico se puede definir como: el área descrita por el centro de una molécula de agua que rueda sobre la superficie de van der Waals que envuelve al grupo, sin penetrar ningún átomo de la molécula. Este concepto fue introducido por Lee y Richards (LR) [26] para describir cuantitativamente la relación entre la proteína y el solvente.

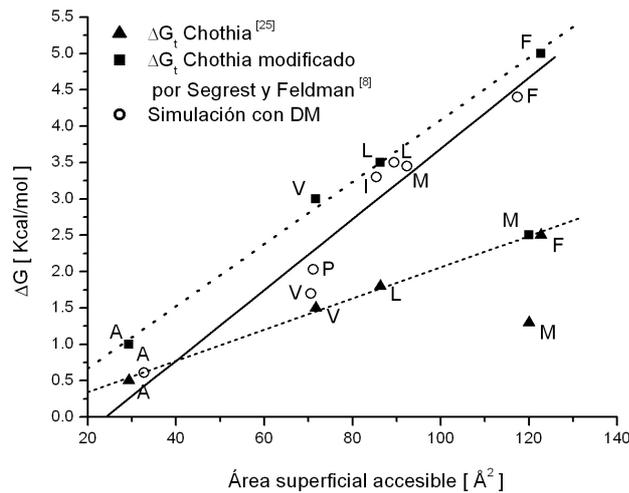


figura 3.2

Chothia [25] tomó los valores de energía libre del trabajo de NT [9], y los graficó en función del área accesible calculada según la definición de LR. Considerando cuatro aminoácidos apolares (A, V, L, F) obtuvo una recta cuya pendiente es igual a  $22 \text{ cal mol}^{-1} \text{ \AA}^{-2}$  y coeficiente de correlación lineal 0.998. En su trabajo, el residuo metionina (M) cae (conjuntamente con otros residuos polares) sobre otra recta de mayor pendiente.

Si tenemos en cuenta la modificación que SF hicieron a los valores de hidrofobicidad de NT para el mismo conjunto de residuos, obtenemos una recta con un valor de pendiente igual a  $42.7 \text{ cal mol}^{-1} \text{ \AA}^{-2}$  y coeficiente de correlación lineal 0.998.

Ahora bien, podemos utilizar la existencia de una relación lineal entre hidrofobicidad y área superficial accesible como test de los valores presentados en la Tabla 3.2, independientemente del valor que obtengamos para la pendiente de la recta. Con este objetivo, calculamos el área superficial accesible al agua utilizando una versión mejorada de lo hecho por LR, como es el algoritmo de Connolly [27].

En la figura 3.2 mostramos con línea llena a la recta que mejor ajusta nuestros datos de simulación. El valor de la pendiente es  $48.6 \text{ cal mol}^{-1} \text{ \AA}^{-2}$ , y su coeficiente de correlación lineal 0.967.

Si tenemos en cuenta que Sharp y colaboradores [28], comparando las rectas obtenidas con diferentes métodos de medición, concluyen que un valor apropiado para la pendiente debería estar entre 46 y 47  $\text{cal mol}^{-1} \text{ \AA}^{-2}$ , vemos que nuestro resultado se encuentra muy próximo a su propuesta. Más aún, si en nuestros cálculos consideramos a la recta que mejor ajusta a los mismos cuatro residuos que utilizó Chothia (A, V, L, F), obtenemos una pendiente igual a 47  $\text{cal mol}^{-1} \text{ \AA}^{-2}$  y coeficiente 0.974.

Por otra parte, observamos que al agregar tres aminoácidos más (P, I, M), el valor de la pendiente prácticamente no muestra un cambio significativo. Al mismo tiempo, gráficos semejantes publicados por otros autores [5, 25], no contienen a todos estos aminoácidos, por lo cual creemos que éste resultado puede ser un interesante aporte.

### **3.1.2 SOLVATACIÓN.**

Las condiciones del medio, y en particular, la interacción de cada aminoácido con las moléculas de su entorno, tienen una participación decisiva en la estabilidad de la estructura nativa de las proteínas en solución. Por esta razón, es interesante estudiar cómo se organizan las moléculas de agua alrededor de cada residuo de aminoácido.

En este apartado, nos ocuparemos de algunos aspectos de la estructura que adquiere el agua al estar en contacto con los aminoácidos apolares, bajo condiciones normales de presión y temperatura.

#### **3.1.2.1 Correlación soluto-solvente.**

La función de distribución radial  $g(r)$  soluto-solvente brinda información sobre la densidad (angularmente promediada) de moléculas de solvente alrededor de un átomo de la molécula de soluto, en función de la distancia al mismo, y relativa a la densidad media del solvente en el seno del líquido. Por eso, para el estudio de propiedades estructurales es muy importante conocerla.

La simulación numérica nos da la posibilidad de controlar las variables del sistema y obtener medidas de propiedades que experimentalmente son muy complicadas. Un ejemplo de esto lo constituye justamente, el cálculo de la función de distribución radial de las moléculas de solvente alrededor de cada uno de los átomos de los aminoácidos, en estado líquido.

Mediante cristalografía de alta resolución, se obtiene información respecto de cómo se ordena el solvente alrededor de un determinado aminoácido [29]. Sin embargo, el estado del solvente en esos experimentos difiere considerablemente del observado en el interior de la célula.

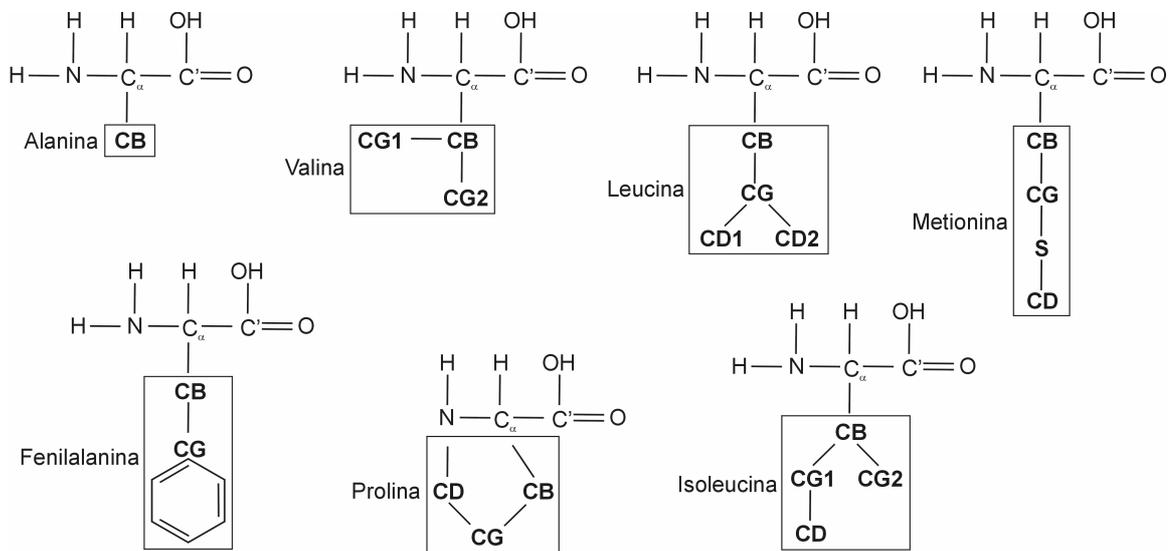


figura 3.3

A continuación, estudiamos la solvatación de los aminoácidos en agua (metanol), analizando las funciones de distribución radial de los átomos de oxígeno del agua (metanol), alrededor de los grupos de átomos de los residuos de cada uno de los siete aminoácidos apolares [30].

En la figura 3.3 mostramos a todos los aminoácidos apolares, a fin de facilitar el análisis y la visualización de la ubicación de cada átomo en los diferentes residuos. En la figura 3.4 graficamos las funciones de distribución radial en agua y metanol superpuestas, para favorecer la comparación de la estructura de solvatación en ambos solventes.

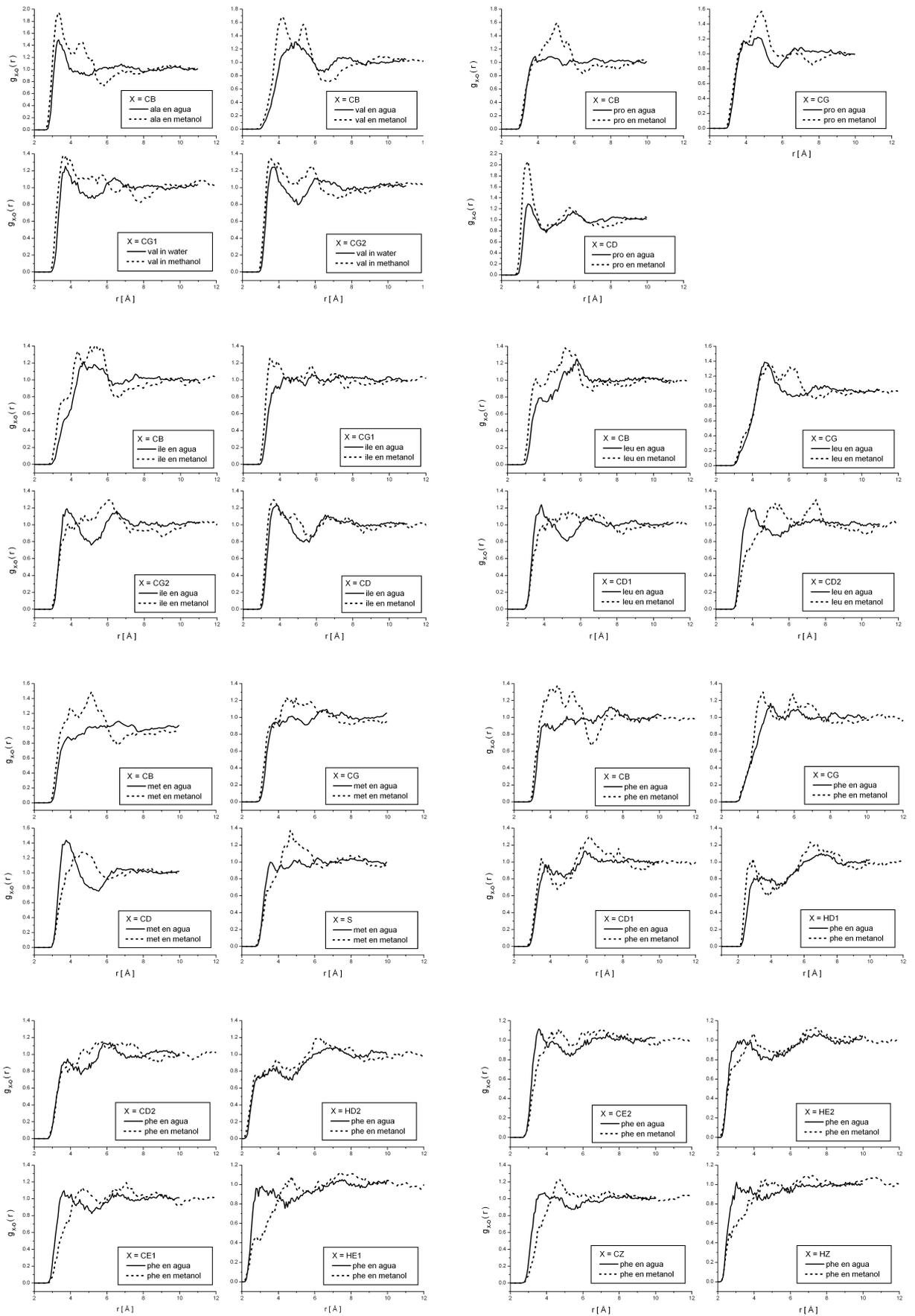


figura 3.4

Realizando una observación global de las curvas de la figura 3.4, podemos señalar que, en general, las funciones de distribución son más estructuradas, con picos y valles más pronunciados para metanol que para agua, lo que indicaría una mayor solvatación en el primer solvente.

Además, en algunos casos, nuevos picos corresponden a nuevas capas de solvatación, y éstas aparecen cuando el solvente es metanol. Esto indicaría que, la presencia del grupo  $\text{CH}_3$  en la molécula de metanol (en lugar de uno de los H en la molécula de agua), promueve una estructura “más rica”.

Las formas de las funciones de distribución radial, son el resultado de una competencia entre diversas fuerzas (atractivas y repulsivas) y de efectos estéricos. Un estudio detallado de las causas del comportamiento particular de cada curva requiere de información complementaria.

Sin embargo, es importante mencionar que las curvas contienen información muy interesante y nada fácil de obtener a partir de métodos experimentales. Por ejemplo, mirando al residuo de leucina en la figura 3.3, podría suponerse que las moléculas de ambos solventes estudiados se estructurarían de la misma forma alrededor de los átomos CD1 y CD2 (ambos grupos metilo) de la cadena lateral.

Sin embargo, las funciones de distribución radial de estos átomos, no muestran la equivalencia o simetría estructural esperada. Algo semejante ocurre para el residuo de valina, con los dos átomos CG1 y CG2 de la cadena lateral. Es evidente que depende de la conformación tridimensional de la molécula, y del entorno (los demás átomos del aminoácido) que cada átomo tiene en solución.

En todas estas simulaciones utilizamos el criterio del *zwitterion* por ser el que predomina en solución acuosa. Esto significa que el aminoácido tiene una carga positiva en su “cabeza” (grupo amino  $\text{NH}_3^+$ ) y una carga negativa en la “cola” (grupo carboxilo  $\text{COO}^-$ ). Sin embargo, la utilización de este criterio no produce una gran perturbación sobre los resultados, ya que cuando realizamos simulaciones con la forma no iónica (cabeza  $\text{NH}_2$  y cola  $\text{COOH}$ ), en iguales condiciones, ni las funciones de correlación radial, ni las energías libre de transferencia presentaron variaciones significativas.

Por otro lado, nos preguntamos qué incidencia tendrán los átomos de la cadena principal (CP) sobre la estructura del agua. En la figura 3.5 mostramos

sólo dos ejemplos, los que consideramos más interesantes de analizar. Con líneas de puntos representamos a las curvas correspondientes al *centro geométrico* de la cadena lateral cuando los demás átomos del aminoácido están presentes en la simulación. Las curvas llenas corresponden a la función que se obtiene cuando sólo se simula a la molécula que representa al residuo, y nuevamente tomamos como posición del átomo del residuo al centro geométrico del mismo.

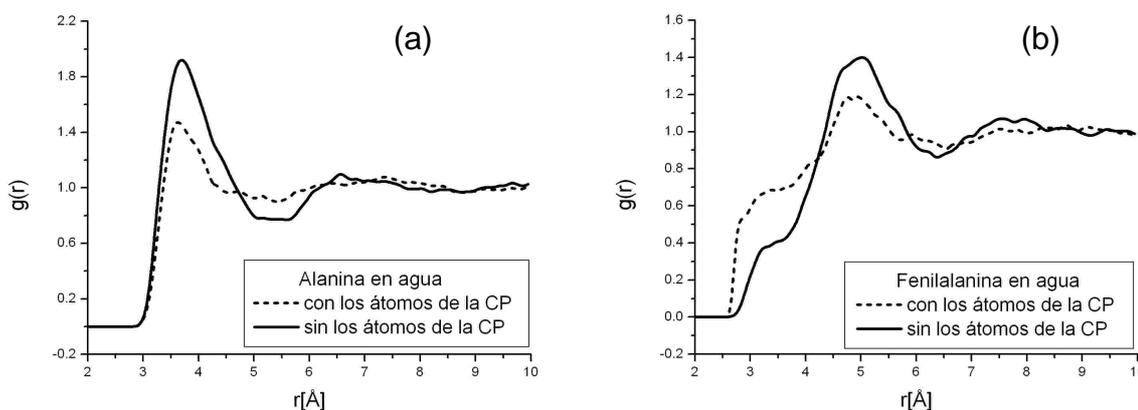


figura 3.5

Para alanina, comparamos la función de distribución radial correspondiente al residuo CB ( $\text{CH}_3$ ) con los átomos de la CP presentes (línea de puntos) y sin ellos (línea llena). Observamos que la posición del pico principal y del primer mínimo es aproximadamente la misma, si bien ambos son más pronunciados para el residuo solo que en presencia de los átomos de la CP.

Debido al espacio que ocupan los átomos de la CP y que impiden que allí puedan ubicarse moléculas de agua, es razonable que en la gráfica aparezca una diferencia de altura del pico. Esto estaría indicando una mayor hidratación del residuo de alanina cuando la CP no está presente, y una distancia a la primer capa de hidratación similar.

Para fenilalanina, notemos la desviación que sufren las curvas alrededor de los (3–4) Å, que podemos relacionar con la forma que presenta el residuo, más cercana a un “disco” que a una “esfera”. En este caso observamos una mayor probabilidad de encontrar moléculas de agua ubicadas a distancias cercanas al centro del anillo cuando la CP está presente, respecto de cuando no lo está.

El pico más importante que muestran las curvas se encuentra a unos 5 Å, este es aproximadamente el valor del “diámetro” del anillo bencénico. Como sucede con alanina, el pico principal es más importante cuando simulamos al residuo sin los átomos de la CP presentes.

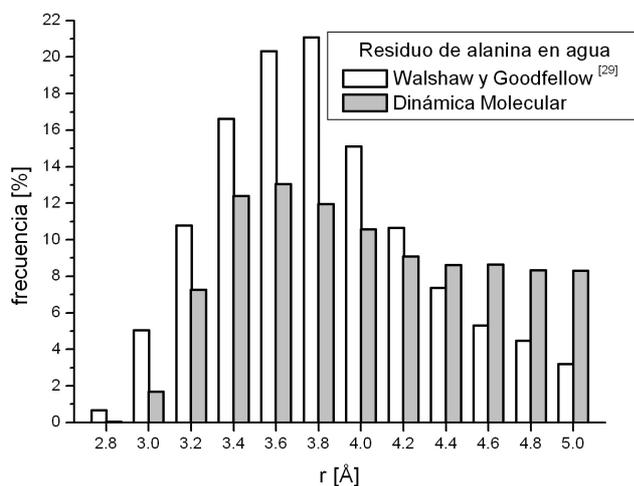


figura 3.6

Otro resultado también muy interesante surge de comparar los datos de simulación, con los obtenidos a partir de mediciones experimentales. Si observamos las posiciones de los máximos y mínimos, éstas están en buen acuerdo con la correspondiente distribución de las moléculas de agua alrededor de las cadenas laterales apolares de estructuras de proteínas bien refinadas obtenidas de datos cristalográficos por Walshaw y Goodfellow [29]. Quizás el ejemplo más claro de esto corresponda a alanina (figura 3.6).

### 3.1.2.2 Puentes de hidrógeno.

Información complementaria muy útil para el análisis de la estructura que presenta el agua alrededor de un residuo de aminoácido apolar, surge del cálculo del número medio de puentes de hidrógeno que forman las moléculas de agua entre sí, cuando el aminoácido está presente respecto de cuando no lo está.

En primer lugar, calculamos mediante DM, el número de moléculas de agua que están formando 1, 2, 3, 4 y 5 puentes de hidrógeno  $n_w$ , en relación al número total de moléculas de agua en el sistema  $N_w$  (figura 3.7). Este cálculo, además de servirnos como control de los resultados que arroja el algoritmo [31], nos permite

graficar la variación relativa que ocurre cuando introducimos en el agua a cada uno de los aminoácidos apolares (figura 3.8).

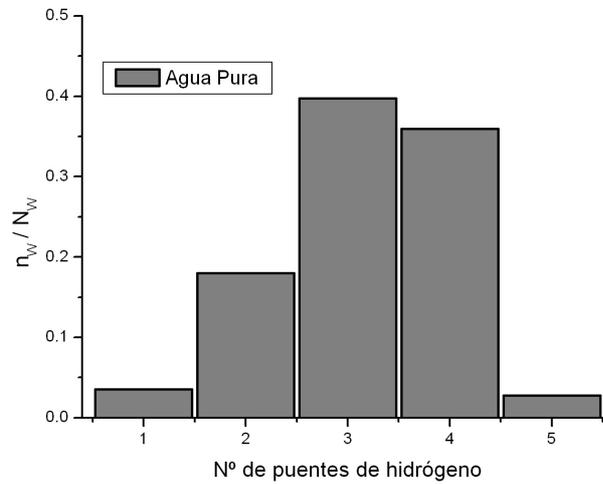


figura 3.7

En la figura 3.8, mostramos dos gráficos. El primero de ellos corresponde a la variación relativa que hay con respecto al agua pura en el número de moléculas de agua que forman 1, 2, 3, 4 y 5 puentes de hidrógeno, cuando introducimos en el sistema un aminoácido apolar completo, esto es, el residuo con la CP.

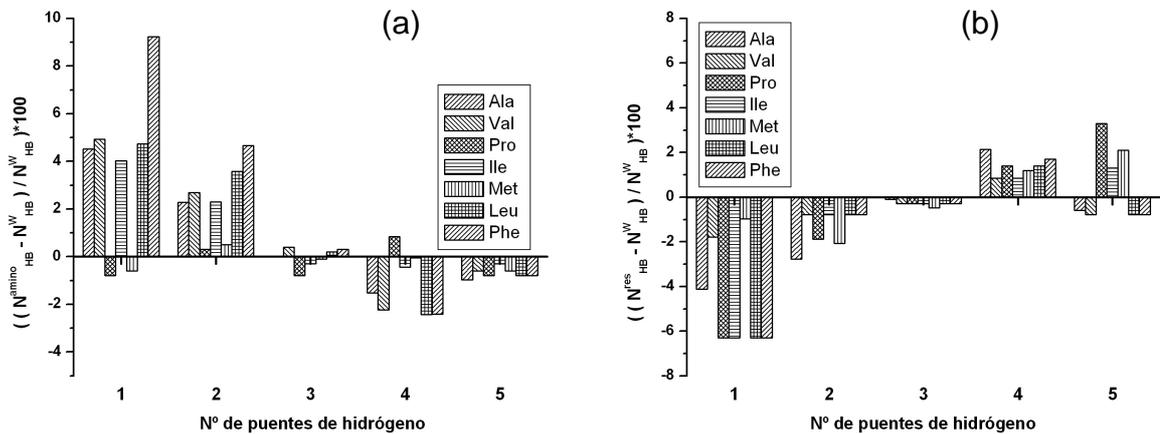


figura 3.8

En el segundo de los gráficos (figura 3.8b), mostramos los resultados que obtenemos cuando introducimos a la molécula correspondiente al residuo sin los

átomos de la CP. Ordenamos a los aminoácidos en forma creciente con la hidrofobicidad, según lo obtenido y presentado en la Tabla 3.2.

Observando ambos gráficos, podemos decir que en ningún caso esa variación supera al 10%. Además, si bien es difícil analizar lo que ocurre a partir del estudio de cada aminoácido en particular, surge a simple vista que el comportamiento de los puentes de hidrógeno que forma el agua, cuando los átomos de la CP están presentes respecto de cuando no lo están, es el opuesto.

En el primer caso, con la CP presente, se produce un cambio en la estructura de puentes de hidrógeno, ya que aumenta el número de moléculas que forman 1 y 2 puentes, y disminuye el número de moléculas que forman 4 y 5. Podría decirse que la presencia del aminoácido completo, esto es, residuo más CP, “destruye” o “rompe” en parte la estructura del agua pura.

Por otra parte, en el segundo de los gráficos vemos que la tendencia es exactamente la opuesta. Pareciera ser que el hecho de que la CP no esté presente, ocasiona en el agua una desviación de la estructura de puentes de hidrógeno donde se favorecen la formación de 4 y 5 en detrimento de 1 y 2. En ambos casos, el número de moléculas de agua que forman 3 puentes de hidrógeno se mantiene aproximadamente constante (variación  $\lesssim 1\%$ ).

Este resultado, conjuntamente con lo observado para la función de distribución radial (figura 3.5), nos indicaría que el pequeño aumento que ocurre en la formación de 4 y 5 puentes de hidrógeno cuando la CP no está presente, debe estar sucediendo en la capa de agua en contacto con el residuo. Esto parece apoyar la idea clásica sobre la hidratación hidrofóbica (en éstas condiciones de temperatura y presión) donde se estaría formando una estructura de “clatrato” alrededor del residuo apolar cuando lo simulamos solo. Sin embargo, cuando consideramos a todo el aminoácido en la simulación, pareciera que los átomos de la CP no permiten que se forme dicha estructura, quizás porque al utilizar el criterio del zwitterion se pierde (al menos en parte) el carácter apolar de la molécula de soluto.

## 3.2 TEORÍA MECÁNICO-ESTADÍSTICA.

### 3.2.1 APLICACIÓN DE LA TEORÍA DE PRATT Y CHANDLER.

El estudio de soluciones muy diluidas de solutos no polares en agua tiene gran interés por varias razones. En primer lugar, estas soluciones presentan algunas propiedades “anómalas” en relación con otros solventes no acuosos, lo que constituye un desafío muy atractivo.

Por otra parte, el estudio de soluciones acuosas muy diluidas, visto como un sistema que se desvía levemente de agua pura, brinda información respecto del agua en sí misma.

Si observamos el desarrollo teórico del estudio de soluciones acuosas diluidas, vemos que estuvo siempre muy ligado al avance conseguido en las teorías de agua pura. Los progresos en este campo casi siempre estuvieron acompañados de los progresos en las teorías de agua pura. Como señalamos antes, no contamos aún con una teoría de agua pura lo suficientemente precisa como para dar cuenta de todas sus peculiaridades. Por eso en este capítulo, tomamos la teoría semiempírica propuesta por Pratt y Chandler para el tratamiento del efecto hidrofóbico [32] que comentamos en el final del capítulo anterior, y la aplicamos a sistemas formados por residuos apolares de aminoácidos disueltos en agua, a dilución infinita.

#### 3.2.1.1 Definiciones y ecuaciones del modelo.

El sistema que vamos a estudiar está compuesto por una molécula de residuo apolar disuelta en agua a dilución infinita, 300 °K de temperatura y presión de 1 atm. En este modelo, tanto las moléculas de residuo de aminoácido como las de agua, son consideradas esferas que interactúan entre sí a través de un potencial  $u(|\vec{r} - \vec{r}'|)$  que depende solamente de la distancia entre ellas.

En adelante, usaremos a la letra “ $\rho$ ” para designar a la molécula de residuo apolar y “ $\alpha$ ” para la de agua. Expresamos a la función de distribución radial residuo-agua ( $\rho\alpha$ ) por,

$$g_{\rho\alpha}(r) = y_{\rho\alpha}(r) \exp[-u_{\rho\alpha}(r)/k_B T] \quad (3.1)$$

donde  $y_{\rho\alpha}(r)$  es la función de distribución radial de cavidad [33].

Como vimos en el apartado (2.2.2) del capítulo anterior, en esta teoría no utilizamos ningún modelo para representar a la interacción de las moléculas de agua pura entre sí. Las interacciones entre moléculas de agua ( $\alpha\alpha$ ) son tratadas en forma exacta, esto es, utilizamos la función de distribución radial  $g_{\alpha\alpha}(r)$  *oxígeno–oxígeno* obtenida en forma experimental.

Teniendo en cuenta que trabajamos en el límite de dilución infinita y que la solubilidad de moléculas apolares en agua es muy baja, es razonable considerar a la función  $g_{\alpha\alpha}(r)$  prácticamente igual a la de agua pura. Sin embargo, para estudiar las propiedades asociadas con los residuos apolares disueltos en agua, debemos introducir un potencial que represente a las interacciones entre una molécula de residuo y una molécula de agua.

En la figura 3.9a repetimos (ver figura 2.2a) la gráfica del potencial de pares desnudo del tipo Lennard–Jones (LJ) elegido para la interacción residuo–agua. En la gráfica indicamos con  $r_{\rho\alpha}^{(0)}(r)$  a la posición del mínimo de  $u_{\rho\alpha}(r)$  y con  $-\varepsilon_{\rho\alpha}$  al valor que toma el potencial  $u_{\rho\alpha}(r)$  en el punto  $r_{\rho\alpha}^{(0)}(r)$ . Tomamos como valor de referencia para la energía  $u_{\rho\alpha}(r \rightarrow \infty) = 0$ .

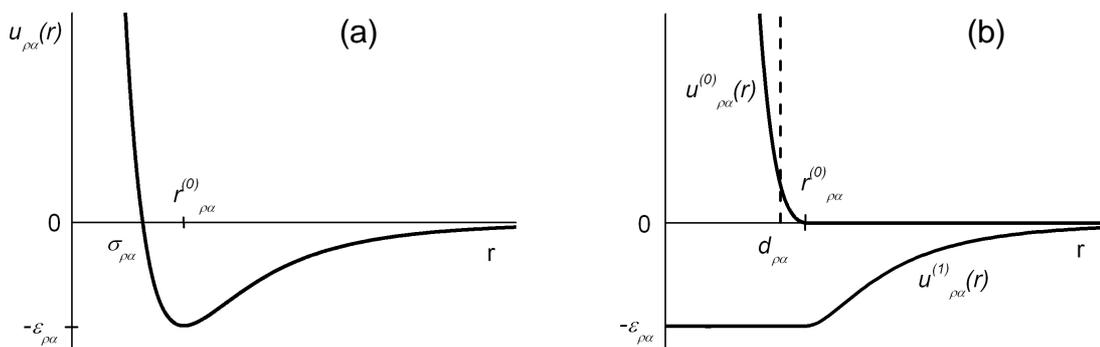


figura 3.9

Descomponemos al potencial desnudo  $u_{\rho\alpha}(r)$  en dos partes (ecs. 2.33): una repulsiva de corto alcance  $u_{\rho\alpha}^{(0)}(r)$  y otra atractiva de largo alcance  $u_{\rho\alpha}^{(1)}(r)$  (figura 3.9b).

Si tenemos en cuenta que la parte repulsiva del potencial es de corto alcance y varía abruptamente como función de la distancia de separación entre partículas, entonces podemos suponer que tiene una gran incidencia en el ordenamiento de las moléculas. Matemáticamente esto significa que,

$$g_{\rho\alpha}(r) \approx g_{\rho\alpha}^{(0)}(r) \quad (3.2)$$

donde  $g_{\rho\alpha}^{(0)}(r)$  es la función de distribución radial para el sistema de referencia, esto es, la función de distribución radial para el hipotético fluido en el cual  $u_{\rho\alpha}^{(1)}(r) = 0$ . Entonces escribimos,

$$g_{\rho\alpha}(r; u_{\rho\alpha}(r)) \approx g_{\rho\alpha}(r; u_{\rho\alpha}^{(0)}(r)) \equiv g_{\rho\alpha}^{(0)}(r) \quad (3.3)$$

Con el superíndice “(0)” designamos al sistema de referencia, esto es, al fluido en el cual las moléculas apolares interactúan con el agua a través de  $u_{\rho\alpha}(r) = u_{\rho\alpha}^{(0)}(r)$ .

Estamos interesados en describir las propiedades estructurales asociadas con el sistema de referencia, esto es, una molécula esférica apolar disuelta en agua líquida, e interactuando con ésta según (2.33a). Este problema está íntimamente relacionado con la Mecánica Estadística de *esferas duras* (ED) disueltas en agua. En este caso, la molécula de soluto es representada por una esfera impenetrable, y el potencial  $u_{\rho\alpha}(r) = u_{\rho\alpha}^{ED}(r)$  estará expresado por

$$u_{\rho\alpha}^{ED}(r) = \begin{cases} \infty & r < d_{\rho\alpha} \\ 0 & r \geq d_{\rho\alpha} \end{cases} \quad (3.4)$$

donde  $d_{\rho\alpha}$  simboliza al diámetro de la esfera dura (figura 3.9b), y cumple con la siguiente condición,

$$\int d\vec{r} y_{\rho\alpha}^{ED}(r) \left\{ \exp[-u_{\rho\alpha}^{(0)}(r)/k_B T] - \exp[-u_{\rho\alpha}^{ED}(r)/k_B T] \right\} = 0 \quad (3.5)$$

Luego, con buena aproximación, podemos calcular a la función de distribución radial  $g_{\rho\alpha}^{(0)}(r)$  para dos moléculas que interactúan con el potencial repulsivo (2.33a) mediante,

$$g_{\rho\alpha}^{(0)}(r) = y_{\rho\alpha}^{ED}(r) \exp[-u_{\rho\alpha}^{(0)}(r)/k_B T] \quad (3.6)$$

donde  $y_{\rho\alpha}^{ED}(r)$  es la función de distribución de cavidad para esferas duras [32].

### 3.2.1.2 Cálculo de la función de correlación radial.

A continuación, presentamos los elementos teóricos básicos para la determinación de la función de distribución radial. Primero, calculamos a la función de cavidad radial para esferas duras apolares en agua a dilución infinita. Luego, utilizamos la ecuación (3.6) para determinar a la función  $g_{\rho\alpha}^{(0)}(r)$  para el potencial de referencia.

Para comenzar, restamos la contribución del gas ideal a la función de distribución y definimos a la función de correlación radial de pares,

$$h_{\rho\alpha}(r) = g_{\rho\alpha}(r) - 1 \quad (3.7)$$

Luego (ver apartado 2.2.2) introducimos a la función  $c_{\rho\alpha}(r)$  a través de una ecuación del tipo Ornstein–Zernike (OZ) [33],

$$h_{\rho\alpha}(r) = c_{\rho\alpha}(r) + \rho_{\alpha} \int d\vec{r}' c_{\rho\alpha}(|\vec{r} - \vec{r}'|) h_{\alpha\alpha}^{\text{exp}}(r') \quad (3.8)$$

donde  $h_{\alpha\alpha}^{\text{exp}}(r) = g_{\alpha\alpha}^{\text{exp}}(r) - 1$  es la función de correlación radial para el agua líquida pura, determinada en forma experimental [34, 35, 36].

Es importante señalar que las funciones que aparecen en la ecuación (3.8), dependen sólo de la distancia de separación entre moléculas, pero no de las orientaciones. Por esta razón, la ecuación (3.8) es diferente de la expresión usual de OZ (ec. 2.10).

Más allá de esto, el significado físico de esas funciones de correlación radial coincide con las ideas que OZ emplearon para escribir su expresión. Una molécula de soluto y una molécula de agua están correlacionadas en forma directa (el primer término del segundo miembro de (3.8)), o en forma indirecta a través de las otras moléculas de agua que actúan como intermediarias (la integral del segundo miembro). La interpretación literal de estas ideas para el sistema de esferas duras sugiere,

$$c_{\rho\alpha}^{\text{ED}}(r) = 0 \quad r > d_{\rho\alpha} \quad (3.9)$$

que se combina con la condición de borde siguiente,

$$g_{\rho\alpha}^{\text{ED}}(r) = h_{\rho\alpha}^{\text{ED}}(r) + 1 = 0 \quad r < d_{\rho\alpha} \quad (3.10)$$

A continuación, determinamos a la función  $c_{\rho\alpha}^{ED}(r)$  que satisface la condición (3.9). Una vez hecho esto, reemplazamos en la ecuación (3.8), e integramos para obtener  $h_{\rho\alpha}^{ED}(r)$  cuando  $r > d_{\rho\alpha}$ .

Para determinar a la función de correlación directa  $c_{\rho\alpha}^{ED}(r)$ , usamos un procedimiento variacional. Definimos un funcional  $I[c_{\rho\alpha}^{ED}(r)]$  del modo siguiente:

$$I[c_{\rho\alpha}^{ED}(r)] = \int d\vec{r} c_{\rho\alpha}^{ED}(r) + \int d\vec{r} [c_{\rho\alpha}^{ED}(r)]^2 + \frac{\rho_\alpha}{(2\pi)^3} \int d\vec{k} [\hat{c}_{\rho\alpha}^{ED}(k)]^2 \hat{h}_{\alpha\alpha}^{\text{exp}}(k) \quad (3.11)$$

donde indicamos con el símbolo “^” a la *función transformada*,

$$\hat{h}_{\alpha\alpha}^{\text{exp}}(k) = \int d\vec{r} \exp[-i \vec{k} \vec{r}] h_{\alpha\alpha}^{\text{exp}}(r) \quad (3.12)$$

Luego, la derivada funcional de  $I[c_{\rho\alpha}^{ED}(r)]$  es:

$$\frac{\delta I}{\delta c_{\rho\alpha}^{ED}(r)} = g_{\rho\alpha}^{ED}(r) \quad (3.13)$$

De la ecuación anterior y (3.10), obtenemos

$$\frac{\delta I}{\delta c_{\rho\alpha}^{ED}(r)} = 0 \quad r < d_{\rho\alpha} \quad (3.14)$$

Resolver esta ecuación variacional es equivalente a resolver las ecuaciones (3.8, 3.9 y 3.10). Proponemos para la función  $c_{\rho\alpha}^{ED}(r)$  el siguiente desarrollo,

$$c_{\rho\alpha}^{ED}(r) = \sum_{n=0}^{n_{\max}} c_n \left[ \frac{r - d_{\rho\alpha}}{d_{\rho\alpha}} \right]^n \quad r < d_{\rho\alpha} \quad (3.15)$$

Luego, la ecuación (3.14) se reduce a un conjunto de ecuaciones algebraicas lineales acopladas,

$$\frac{\partial I}{\partial c_n} = 0 \quad n = 1, 2, \dots, n_{\max} \quad (3.16)$$

Una vez determinados los coeficientes de la ecuación (3.15), queda determinada la función  $c_{\rho\alpha}^{ED}(r)$ . Reemplazamos esta función en la ecuación (3.8) y calculamos  $h_{\rho\alpha}^{ED}(r)$  para  $r > d_{\rho\alpha}$ .

La función de cavidad radial puede aproximarse por,

$$y_{\rho\alpha}^{ED}(r) \approx \begin{cases} -c_{\rho\alpha}^{ED}(r) & r < d_{\rho\alpha} \\ g_{\rho\alpha}^{ED}(r) & r > d_{\rho\alpha} \end{cases} \quad (3.17)$$

Finalmente reemplazamos la expresión anterior en la ecuación (3.6), y obtenemos la función de distribución radial residuo-agua para el potencial repulsivo de referencia  $u_{\rho\alpha}^{(0)}(r)$ .

En la figura 3.10 comparamos las gráficas de tres funciones de distribución radial residuo-agua para el caso de alanina (A). Representamos con líneas de trazos a la función de distribución radial para el potencial  $u_{\rho\alpha}^{ED}(r)$  y con línea llena a la función calculada mediante la ecuación (3.6), a partir de la función de distribución de cavidad determinada según (3.17).

Los círculos vacíos corresponden a la función de distribución radial residuo de alanina-agua ( $A\alpha$ ) obtenida de simulación mediante DM, sin los átomos de la CP presentes, a presión de 1 atm y 300 °K de temperatura. Para representar al agua en la simulación utilizamos el modelo SPC/E [37].

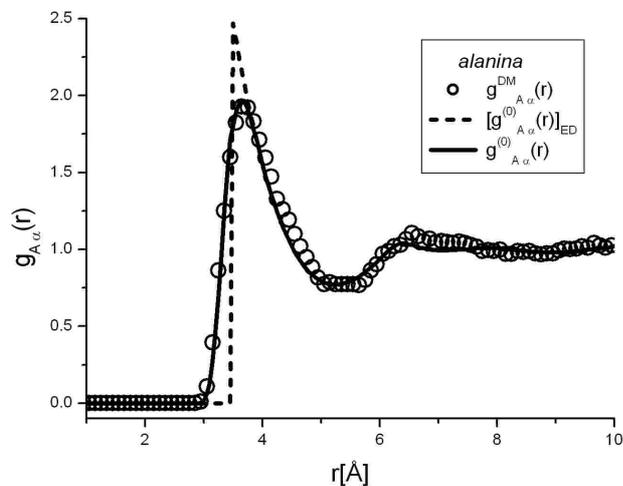


figura 3.10

Como vemos en la figura 3.10, el acuerdo exhibido por las funciones de distribución obtenidas en forma teórica y mediante simulación para el residuo de alanina es muy bueno. Sin embargo, no encontramos un acuerdo tan bueno como éste para todos los residuos apolares.

Sería muy tedioso mostrar a todas las gráficas correspondientes a todos los residuos apolares. En lugar de hacer esto, vamos a presentar sólo dos casos más

a modo de ejemplo (figura 3.11). Ya vimos al residuo que presenta el “mejor” acuerdo, ahora procedemos a mostrar un caso “intermedio” (leucina), y al “peor” de todos (fenilalanina).

Para leucina (L), las curvas surgidas de DM y del modelo teórico, presentan un acuerdo aceptable en cuanto a la posición del pico y del primer mínimo, aún cuando no coincidan totalmente.

En cambio, para fenilalanina (F), encontramos la mayor diferencia entre las curvas de DM y del modelo. Si bien las posiciones del pico principal y del primer mínimo son razonables, en la gráfica de la función de distribución surgida de simulación observamos una primera “meseta” pequeña alrededor de 3–4 Å.

Como comentamos anteriormente, el residuo de fenilalanina posee un anillo bencénico que le confiere una forma más parecida a un “disco” que a una “esfera” (figura 3.3). Esta primera “meseta” entonces, sería una medida de la probabilidad de que las moléculas de agua se sitúen cercanas a la superficie plana del disco, mientras que la posición del pico más alto estaría relacionada con el diámetro del mismo.

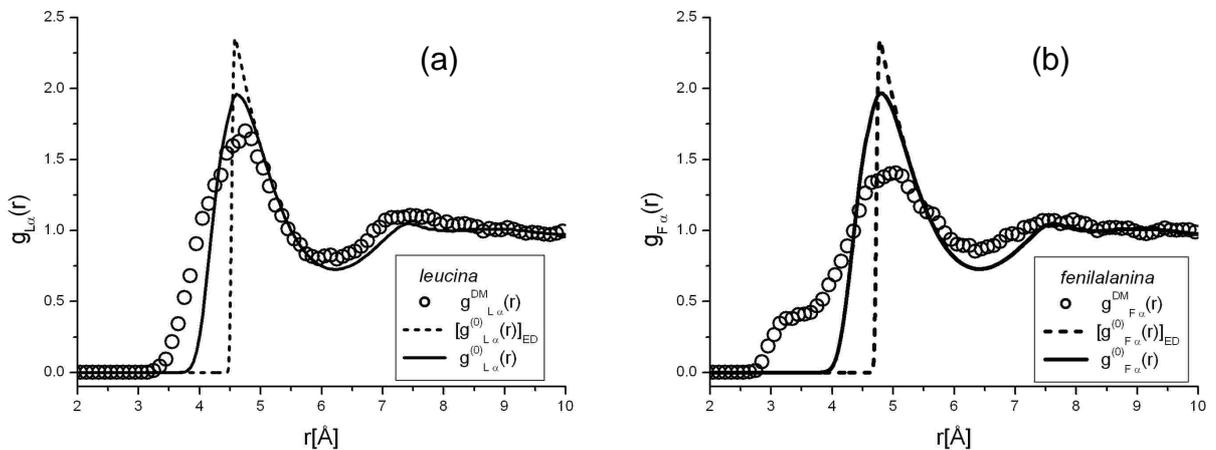


figura 3.11

Puesto que en nuestro modelo consideramos a los residuos como esferas, es lógico que la curva llena que corresponde a la teoría sea incapaz de reproducir este hecho, y que su pico esté ubicado en la posición determinada por el diámetro del anillo.

Las curvas de círculos abiertos, fueron obtenidas simulando a la molécula del residuo sola, esto es, sin los átomos de la CP presente. Luego, calculamos a la función de distribución radial tomando las distancias entre el punto que ocupa el centro de geometría de la molécula de residuo y los oxígenos de las moléculas de agua.

Hasta aquí, despreciamos el término atractivo de  $u_{\rho\alpha}(r)$  y obtuvimos la función de distribución radial para el potencial de referencia. Ahora, podemos dar un paso más y considerar en forma perturbativa la incidencia del término atractivo  $u_{\rho\alpha}^{(1)}(r)$  del potencial. De este modo determinamos una nueva  $g_{\rho\alpha}(r)$  a partir del resultado anterior.

Los detalles de éste cálculo serán presentados en el próximo capítulo, aquí sólo adelantamos que en él hacemos uso de la teoría de Andersen, Chandler y Weeks [38], conocida como la aproximación “EXP”.

Una vez más, presentamos sólo una gráfica a modo de ejemplo (alanina), aunque debemos aclarar que el comportamiento de los demás residuos es semejante en cuanto a la variación relativa que sufre la función de distribución radial ( $\leq 10\%$ ) respecto de su valor de referencia (gráfica con círculos abiertos pequeños en la figura 3.12).

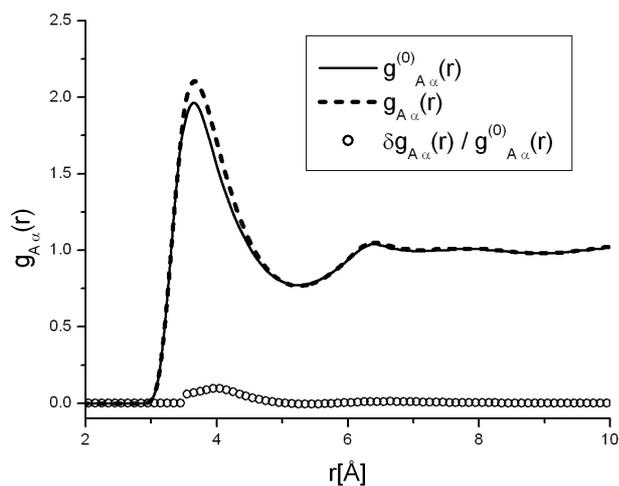


figura 3.12

Para finalizar este apartado, hacemos algunos comentarios acerca de la función de distribución radial utilizada para el agua pura y los parámetros  $\{\varepsilon_\alpha, \sigma_\alpha\}$  empleados para caracterizarla en los cálculos.

Disponemos de tres funciones de distribución radial para agua pura obtenidas en forma experimental. Una del año 1971 determinada por Narten y Levy [34], otra de 1986 por Soper y Phillips [35] y la más reciente por Soper en 1994 [36].

Además de estas funciones, contamos con una cuarta que fue obtenida por nosotros mediante una Dinámica Molecular con 512 moléculas de agua, en las mismas condiciones de presión y temperatura que los sistemas anteriores, utilizando el modelo SPC/E ya mencionado.

De las cuatro alternativas posibles, decidimos utilizar una de las funciones de distribución radial oxígeno–oxígeno obtenida en forma experimental, en lugar de aquella surgida de una simulación cuyo modelo fue ajustado para coincidir con ésta. Luego, de las funciones experimentales posibles, optamos por la más reciente de las tres, esto es, la función de distribución radial oxígeno–oxígeno obtenida por Soper en 1994.

En cuanto a la elección de los parámetros asociados al agua pura, creímos conveniente elegir  $\left\{ \sigma_\alpha = 3.166 \text{ \AA}, \varepsilon_\alpha / k_B = 78.2 \text{ }^\circ\text{K} \right\}$  [37], correspondientes al modelo SPC/E que utilizamos en la DM.

### 3.2.2 NÚMERO DE COORDINACIÓN.

En nuestros cálculos teóricos, supusimos a los residuos apolares interactuando con el agua a través de un potencial del tipo Lennard–Jones (2.33). Para determinar a las funciones de distribución radial presentadas en los apartados anteriores, necesitamos introducir cuatro parámetros de entrada  $\{\varepsilon_\alpha, \sigma_\alpha, \varepsilon_\rho, \sigma_\rho\}$  que caractericen al residuo apolar y al agua.

Luego, utilizamos las relaciones:

$$\begin{aligned} \sigma_{\rho\alpha} &= (\sigma_\rho + \sigma_\alpha) / 2 \\ \varepsilon_{\rho\alpha} &= \sqrt{\varepsilon_\rho \varepsilon_\alpha} \end{aligned} \quad (3.18)$$

para determinar los parámetros que intervienen en el potencial.

Para reducir al mínimo el número de parámetros a ajustar en forma arbitraria, decidimos tomar los valores que surgen del modelo SPC/E para el agua pura, y el  $\sigma_\rho$  surgido de mediciones experimentales realizadas por otros autores [6] (Tabla 3.3).

Por último, asignamos un valor al parámetro  $\varepsilon_\rho$  con el siguiente criterio: primero obtenemos la función de distribución  $g_{\rho\alpha}(r)$  a partir de simulación mediante DM para un dado residuo apolar. A continuación, calculamos la  $g_{\rho\alpha}(r)$  que surge del modelo teórico expuesto en el apartado anterior y ajustamos el valor del parámetro  $\varepsilon_\rho$  hasta que el número de coordinación ( $N_C$ ) obtenido de esta última función difiere en menos de un 1% del valor encontrado para la anterior. Para calcular el número de coordinación  $N_C$  usamos la expresión usual,

$$N_C = 4\pi\rho_\alpha \int_0^{r^{\min}} dr r^2 g_{\rho\alpha}(r) \quad (3.19)$$

El cálculo de  $N_C$  no sólo brinda un criterio con el cual determinar un valor razonable para el parámetro  $\varepsilon_\rho$ , también provee información acerca de la hidratación de cada residuo. En la Tabla 3.3 detallamos los valores obtenidos en cada caso.

Tabla 3.3 Parámetros y número de coordinación para los residuos apolares.

Residuos	$\sigma_\rho$ [Å]	$\varepsilon_\rho/k_B$ [°K]	$N_C$
ala (A)	3.794	148	18–19
pro (P)	4.932	120	24–25
val (V)	5.324	120	25–26
met (M)	5.818	110	27–28
ile (I)	5.887	90	28–29
leu (L)	5.887	80	29–30
phe (F)	6.277	70	31–32

Observemos que a medida que aumenta el diámetro del residuo disminuye el valor del parámetro  $\varepsilon_\rho$ . Por otro lado, vemos que  $N_C$  aumenta cuando aumenta

el tamaño del residuo, lo cual resulta razonable si tenemos presente que  $N_C$  es una medida del número de moléculas de agua que se hallan en contacto con cada residuo.

Los valores de  $N_C$  encontrados para los residuos apolares son bastante más grandes que el correspondiente a agua pura ( $N_C = 4 - 5$ ), donde la estructura es aproximadamente tetraédrica.

Dejamos por un momento de lado a las funciones de distribución radial  $g_{\rho\alpha}(r)$  que arroja el modelo teórico, y nos concentramos en las funciones obtenidas mediante DM para los residuos apolares. Vemos que todas presentan un pico principal bastante ancho, le sigue un mínimo algo menor que uno, y finalmente, la función  $g_{\rho\alpha}(r)$  se hace prácticamente constante e igual a uno. Entonces, podemos decir que existe una capa de hidratación que cuenta con un número  $N_C$  de moléculas, a partir de la cual tenemos agua pura en estado líquido. De esta forma, el valor de  $N_C$  nos brinda una medida aproximada del número de moléculas de agua que estarían involucradas en la variación de la estructura de puentes de hidrógeno que veíamos en el apartado 3.1.2.2.

### 3.3 REFERENCIAS DEL CAPÍTULO.

- 1 C.B. Anfinsen, *Science*, **181**, 223 (1973).
- 2 C. Branden and J. Tooze, *Introduction to Protein Structure* (Garland Publishing, Inc., New York, 1991).
- 3 W. Kauzmann, *Adv. Protein Chem.*, **14**, 1 (1959).
- 4 W. Blokzijl and J.B.F.N. Engberts, *Angew. Chem. Int. Ed. Engl.*, **32**, 1545 (1993).
- 5 P.A. Karplus, *Prot. Sci.*, **6**, 1302 (1997).
- 6 T.E. Creighton, *Proteins. Structures and Molecular Properties* (W.H. Freeman and Company, New York, 1993).
- 7 T. Lazaridis, *Acc. Chem. Res.*, **12**, 931 (2001).
- 8 J.P. Segrest and R.J. Feldman, *J. Mol. Biol.*, **87**, 853 (1974).
- 9 Y. Nozaki and C. Tanford, *J. Biol. Chem.*, **246**, 2211 (1971).

- 10 J.L. Cornette, K.B. Cease, H. Margalit, J.L. Spouge, J.A. Berzofsky and C. DeLisi, *J. Mol. Biol.*, **195**, 659 (1987). Sitios web: (<http://www.lsbu.ac.uk/water/phobic.html>, <http://solon.cma.univie.ac.at/~neum/software/protein/aminoacids.html> ).
- 11 D. Eisenberg, R.M. Weiss, T.C. Terwilliger and W. Wilcox, *Faraday Symp. Chem. Soc.*, 17/8 (1982).
- 12 J. Janin, *Nature*, **277**, 491 (1978).
- 13 C. Chothia, *J. Mol. Biol.*, **105**, 1 (1976).
- 14 R. Wolfenden, L. Anderson, P.M. Cullis and C.C.B. Southgate, *Biochemistry*, **20**, 849 (1981).
- 15 G. von Heijne and C. Blomberg, *Eur. J. Biochem.*, **97**, 175 (1979).
- 16 T. Lazaridis, G. Archontis and M. Karplus, *Adv. Prot. Chem.*, **14**, 1 (1995).
- 17 C. Tanford, *J. Am. Chem. Soc.*, **84**, 4240 (1962).
- 18 W.F. van Gunsteren, S.R. Billeter, A.A. Eising, W.R.P. Scott, and I.G. Tironi, *Biomolecular Simulation: The Gromos96 Manual and User Guide* (Zürich: vdf Hochschulverlag AG an der ETH Zürich, 1996).
- 19 X. Daura, P.H. Hünenberger, A. E. Mark, E. Querol, F.X. Avilés and W.F. van Gunsteren, *J. Am. Chem. Soc.*, **118**, 6285 (1996).
- 20 W.F. van Gunsteren and P.K. Weiner, *Computer Simulation of Biomolecular System*, (ESCOM Science Publishers, Leiden, 1989).
- 21 H.J.C. Berendsen, J.P.M. Postma and W.F. van Gunsteren, *Molecular Dynamics and Protein Structure*, edited by J. Hermans (Western Springs, IL: Policrystal Book Service, 1985).
- 22 M.P. Allen, and D.J. Tildesley, *Computer Simulation of Liquids*, (Oxford University Press, 1987).
- 23 M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions*, (Dover, New York, 1965).
- 24 J.P. Ryckaert, G. Ciccotti and H.J.C. Berendsen, *J. Comput. Phys.*, **23**, 327 (1997).
- 25 C. Chothia, *Nature*, **248**, 338 (1974).
- 26 B. Lee and F.M. Richards, *J. Mol. Biol.*, **55**, 379 (1971).
- 27 M.L. Connolly, *Science*, **221**, 709 (1983). M.L. Connolly, *J. Appl. Crystallogr.*, **16**, 548 (1983).

- 28 K.A. Sharp, A. Nicholls, R. Friedman and B. Honig, *Science*, **252**, 106 (1991).
- 29 J. Walshaw and J.M. Goodfellow, *J. Mol. Biol.*, **231**, 392 (1993).
- 30 D.G. Renzi, C.M. Carlevaro, C. Stoico and F. Vericat, *Molecular Physics*, **99**, 913 (2001).
- 31 C.M. Carlevaro, E.R. Caffarena and J.R. Grigera, *Intl. J. Boil. Macromol.*, **23**, 149 (1998).
- 32 L.R. Pratt and D. Chandler, *J. Chem. Phys.*, **67**, 3683 (1977).
- 33 J.P. Hansen and I.R. McDonald, *Theory of Simple Liquids* (Academic Press, New York, 1986).
- 34 A.H. Narten and H.A. Levy, *J. Chem. Phys.*, **55**, 2263 (1971).
- 35 A.K. Soper and M.G. Phillips, *J. Chem. Phys.*, **107**, 47 (1986).
- 36 A.K. Soper, *J. Chem. Phys.*, **101**, 6888 (1994).
- 37 H.J.C. Berendsen, J.R. Grigera and T.P. Straatsma, *J. Phys. Chem.*, **91**, 6269 (1987).
- 38 C. Andersen, D. Chandler and J.D. Weeks, *Adv. Chem. Phys.*, **34**, 105 (1979).

## 4. DOS RESIDUOS APOLARES DISUELTOS EN AGUA: INTERACCIÓN HIDROFÓBICA.

Comenzamos extendiendo el modelo mecánico–estadístico presentado en los capítulos anteriores (apartados 2.2.2 y 3.2), y lo aplicamos a un sistema formado por dos residuos de aminoácidos apolares disueltos en agua.

En primer lugar, calculamos las funciones de correlación correspondientes al sistema de referencia, esto es, cuando consideramos a las moléculas de residuo interactuando entre sí y con el agua, a través de un potencial repulsivo.

A continuación, determinamos en forma perturbativa el efecto producido por la acción de interacciones atractivas sobre las funciones de distribución radial residuo–residuo, y obtenemos los correspondientes potenciales efectivos de interacción de pares de aminoácidos en agua como función de la distancia.

Finalmente, cotejamos algunos de los resultados surgidos de este modelo con los calculados por otros autores mediante simulación.

### 4.1 TEORÍA MECÁNICO–ESTADÍSTICA.

Retomamos los cálculos realizados en el capítulo anterior para tratar a la hidratación hidrofóbica, y extendemos la teoría para calcular la función de distribución radial residuo–residuo [1], y el potencial de fuerza media asociado a ella. El potencial de fuerza media  $W_{\rho\rho'}(r)$  representa una medida de la interacción inducida por el agua sobre las dos moléculas de residuo, y lo calculamos mediante:

$$W_{\rho\rho'}(r) = -k_B T \ln[g_{\rho\rho'}(r)] \quad (4.1)$$

donde  $g_{\rho\rho'}(r)$  es la función de distribución radial de dos solutos esféricos ( $\rho$  y  $\rho'$ ) disueltos en agua.

En nuestro esquema, necesitaremos conocer también a la función de distribución radial de agua líquida pura. Como antes, para determinar las correlaciones que involucran residuos de aminoácidos apolares, elegimos un procedimiento semiempírico que introduce el comportamiento del agua real a

través de la función de distribución radial oxígeno–oxígeno obtenida en forma experimental [2]. La idea de trabajar con un método semiempírico para soluciones acuosas, es bien conocida y ha sido usada por muchos otros investigadores[3].

#### 4.1.1 TEORÍA DEL EFECTO HIDROFÓBICO.

El sistema que consideramos, está compuesto por dos moléculas de residuo apolar esféricas disueltas en agua, que interactúan entre sí a través de un potencial  $u(|\vec{r} - \vec{r}'|)$  dependiente de la distancia entre los centros de cada residuo, o entre el centro de un residuo y el oxígeno de la molécula de agua.

En el capítulo anterior, presentamos muchas de las ecuaciones que siguen. Sin embargo, preferimos repetir a algunas de ellas para mostrar el acoplamiento que existe entre las interacciones *residuo–agua* y *residuo–residuo*. Escribimos a las funciones de distribución radial mediante [4]:

$$g_{\rho M}(r) = y_{\rho M}(r) \exp[-u_{\rho M}(r)/k_B T] \quad (4.2)$$

donde designamos con la letra  $M$  tanto a una molécula de residuo como a una molécula de agua ( $M = \rho'$  o  $M = \alpha$ ).

El potencial de interacción directa entre pares de moléculas elegido, es del tipo Lennard–Jones (ec. 2.32). Al igual que en el capítulo 3, descomponemos al potencial completo  $u_{\rho M}(r)$  en dos partes [4, 5]: una repulsiva de corto alcance  $u_{\rho M}^{(0)}(r)$ , y otra atractiva de largo alcance  $u_{\rho M}^{(1)}(r)$ , o sea,  $u_{\rho M}(r) = u_{\rho M}^{(0)}(r) + u_{\rho M}^{(1)}(r)$  donde,

$$u_{\rho M}^{(0)}(r) = \begin{cases} \varepsilon_{\rho M} + u_{\rho M}(r) & r < r_{\rho M}^{(0)} \\ 0 & r > r_{\rho M}^{(0)} \end{cases} \quad (4.3a)$$

$$u_{\rho M}^{(1)}(r) = \begin{cases} -\varepsilon_{\rho M} & r < r_{\rho M}^{(0)} \\ u_{\rho M}(r) & r > r_{\rho M}^{(0)} \end{cases} \quad (4.3b)$$

Definimos al sistema de referencia como aquel en el cual se desprecia la contribución del término atractivo  $u_{\rho M}^{(1)}(r)$  en el potencial total  $u_{\rho M}(r)$ ,

$$g_{\rho\rho'}(r; u_{\rho\rho'}, u_{\rho\alpha}, u_{\rho'\alpha}) \approx g_{\rho\rho'}(r; u_{\rho\rho'}^{(0)}, u_{\rho\alpha}^{(0)}, u_{\rho'\alpha}^{(0)}) \equiv g_{\rho\rho'}^{(0)}(r) \quad (4.4a)$$

$$g_{\rho\alpha}(r; u_{\rho\alpha}) \approx g_{\rho\alpha}(r; u_{\rho\alpha}^{(0)}) \equiv g_{\rho\alpha}^{(0)}(r) \quad (4.4b)$$

Más adelante, determinaremos en forma perturbativa las contribuciones que los términos atractivos  $u_{\rho M}^{(1)}(r)$  hacen al potencial de fuerza media  $W_{\rho\rho}(r)$ . En principio, tomamos  $u_{\rho M}(r) = u_{\rho M}^{(0)}(r)$  y definimos a  $\sigma_{\rho M}$  y  $\varepsilon_{\rho M}$ , a través de las relaciones usuales entre los parámetros,

$$\begin{aligned} \sigma_{\rho M} &= (\sigma_{\rho} + \sigma_M)/2 \\ \varepsilon_{\rho M} &= \sqrt{\varepsilon_{\rho} \varepsilon_M} \end{aligned} \quad (4.5)$$

Para estudiar las propiedades estructurales asociadas al sistema de referencia formado por moléculas apolares disueltas en agua, primero representamos a la molécula de soluto por una esfera dura (ED), cuyo potencial de interacción  $u_{\rho M}^{ED}(r)$  viene dado por,

$$u_{\rho M}^{ED}(r) = \begin{cases} \infty & r < d_{\rho M} \\ 0 & r \geq d_{\rho M} \end{cases} \quad (4.6)$$

El parámetro  $d_{\rho M}$  representa al diámetro de la esfera impenetrable, y se determina a partir de la condición:

$$\int d\vec{r} y_{\rho M}^{ED}(r) \left\{ \exp[-u_{\rho M}^{(0)}(r)/k_B T] - \exp[-u_{\rho M}^{ED}(r)/k_B T] \right\} = 0 \quad (4.7)$$

Luego, calculamos a la función de distribución radial  $g_{\rho M}^{(0)}(r)$  para dos esferas que interactúan con el potencial repulsivo (4.3a) mediante,

$$g_{\rho M}^{(0)}(r) = y_{\rho M}^{ED}(r) \exp[-u_{\rho M}^{(0)}(r)/k_B T] \quad (4.8)$$

donde  $y_{\rho M}^{ED}(r)$  es la función de distribución de cavidad para esferas duras [4].

#### 4.1.1.1 Correlaciones para el sistema de referencia.

Como lo hicimos en el capítulo 3, restamos la contribución del gas ideal a las funciones de distribución y definimos las funciones de correlación radial de pares,

$$h_{\rho M}^{(0)}(r) = g_{\rho M}^{(0)}(r) - 1 \quad (4.9)$$

Luego introducimos las funciones  $c_{\rho M}^{(0)}(r)$  conjuntamente con las ecuaciones del tipo Ornstein–Zernike [4],

$$h_{\rho\alpha}^{(0)}(r) = c_{\rho\alpha}^{(0)}(r) + \rho_{\alpha} \int d\vec{r}' c_{\rho\alpha}^{(0)}(|\vec{r} - \vec{r}'|) h_{\alpha\alpha}^{\text{exp}}(r') \quad (4.10a)$$

$$h_{\rho\rho'}^{(0)}(r) = c_{\rho\rho'}^{(0)}(r) + \rho_{\alpha} \int d\vec{r}' c_{\rho\alpha}^{(0)}(|\vec{r} - \vec{r}'|) h_{\rho'\alpha}^{(0)}(r') \quad (4.10b)$$

donde  $h_{\alpha\alpha}^{\text{exp}}(r) = g_{\alpha\alpha}^{\text{exp}}(r) - 1$  es la función de correlación de pares para el agua líquida pura, obtenida en forma experimental [2].

A diferencia de lo que ocurre en las expresiones originales de Ornstein–Zernike (OZ) (ver apartado 2.2.1.1), las funciones de correlación que aparecen en las ecuaciones (4.10a) y (4.10b) no dependen de las orientaciones. Estas funciones representarían el “análogo” esférico de las definidas por OZ (ec. (2.10)).

Las ecuaciones (4.10) conjuntamente con las relaciones,

$$c_{\rho\alpha}^{\text{ED}}(r) = 0 \quad r > d_{\rho\alpha} \quad (4.11a)$$

$$c_{\rho\rho'}^{\text{ED}}(r) = 0 \quad r > d_{\rho\rho'} \quad (4.11b)$$

y las condiciones de borde,

$$g_{\rho\alpha}^{\text{ED}}(r) = h_{\rho\alpha}^{\text{ED}}(r) + 1 = 0 \quad r < d_{\rho\alpha} \quad (4.11c)$$

$$g_{\rho\rho'}^{\text{ED}}(r) = h_{\rho\rho'}^{\text{ED}}(r) + 1 = 0 \quad r < d_{\rho\rho'} \quad (4.11d)$$

forman un sistema de ecuaciones que debemos resolver para determinar a la función  $c_{\rho\alpha}^{\text{ED}}(r)$ . Para esto, utilizamos el procedimiento variacional descrito en el capítulo 3 (apartado 3.2.1.2).

Luego, insertamos  $c_{\rho\alpha}^{\text{ED}}(r)$  en la ecuación (4.10a) de modo que satisfaga la condición (4.11c), e integramos para obtener  $h_{\rho\alpha}^{\text{ED}}(r)$  cuando  $r > d_{\rho\alpha}$ . Finalmente, de las ecuaciones (4.10b), (4.11b), y (4.11d) obtenemos,

$$g_{\rho\rho'}^{\text{ED}}(r) = \begin{cases} 0 & r < d_{\rho\rho'} \\ 1 + \frac{\rho_{\alpha}}{(2\pi)^3} \int d\vec{k} e^{i\vec{k}\cdot\vec{r}} \hat{c}_{\rho\alpha}^{\text{ED}}(\vec{k}) \hat{c}_{\rho'\alpha}^{\text{ED}}(\vec{k}) [1 + \rho_{\alpha} \hat{h}_{\alpha\alpha}^{\text{exp}}(\vec{k})] & r > d_{\rho\rho'} \end{cases} \quad (4.12)$$

donde usamos al símbolo “^” para indicar a la función transformada según Fourier (ec. (3.12)).

Si aproximamos a la función de cavidad radial por,

$$y_{\rho M}^{\text{ED}}(r) \cong \begin{cases} -c_{\rho M}^{\text{ED}}(r) & r < d_{\rho M} \\ g_{\rho M}^{\text{ED}}(r) & r > d_{\rho M} \end{cases} \quad (4.13)$$

y la introducimos en la ecuación (4.8), obtendremos a las funciones de distribución  $g_{\rho\alpha}^{(0)}(r)$  y  $g_{\rho\rho'}^{(0)}(r)$  para el sistema de referencia, esto es, cuando las moléculas interactúan a través de un potencial del tipo (4.3a).

Antes de continuar, debemos hacer un comentario respecto de la aproximación hecha en la ecuación (4.13). Cuando  $r < d_{\rho M}$ , la aproximación es razonablemente buena para  $M = \alpha$ , pero no tanto para  $M = \rho'$ . En realidad, la aproximación es aceptable para valores de  $r$  cercanos a  $d_{\rho\rho'}$ , pero deja de serlo a medida que  $r$  tiende a cero [1].

Sin embargo, esto no afecta a los resultados, ya que sólo usamos a la función de cavidad radial  $y_{\rho\rho'}^{ED}(r)$  para determinar  $g_{\rho\rho'}^{(0)}(r)$  a través de la ecuación (4.8). Si miramos ésta ecuación, vemos que en la región donde la aproximación de la función  $y_{\rho\rho'}^{ED}(r)$  es mala ( $r$  pequeño), el factor exponencial es prácticamente igual a cero ( $u_{\rho\rho'}^{(0)}(r)$  toma valores muy grandes). Mientras que, para aquellos valores de  $r$  en los que el factor exponencial toma valores finitos y “pequeños”, la aproximación  $y_{\rho\rho'}^{ED}(r) \cong -c_{\rho\rho'}^{ED}(r)$  es aceptable [1].

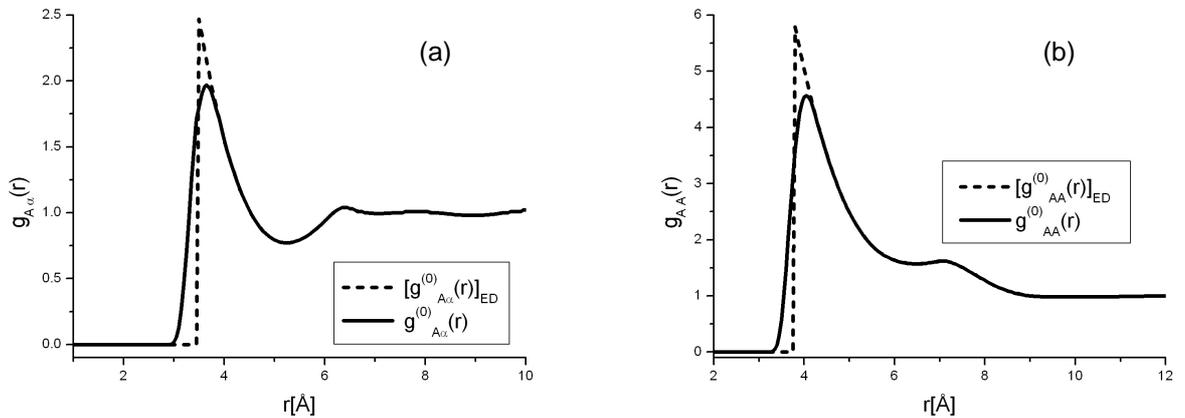


figura 4.1

En la figura 4.1 presentamos las gráficas de las funciones de distribución radial correspondientes al residuo de alanina. Las curvas hechas con líneas de trazo corresponden a las funciones que obtenemos para el potencial de esferas duras (4.6), mientras que las curvas con líneas llenas corresponden a las

funciones de referencia calculadas mediante (4.8), donde los residuos apolares interactúan a través del potencial repulsivo (4.3a).

Como resultado de combinar las ecuaciones (4.8) y (4.13), ambas funciones de distribución radial exhiben un “ablandamiento” del pico principal cualitativamente semejante, aunque cuantitativamente más significativo para la correlación residuo–residuo que residuo–agua.

#### 4.1.1.2 Efecto de las interacciones atractivas.

En la estructura de los líquidos simples, las interacciones atractivas que varían lentamente desempeñan un papel secundario con respecto a las repulsivas. Por esta razón, una estrategia muy común en el estudio de éstos sistemas consiste en despreciar el efecto de las interacciones atractivas, y si el modelo resultara “exitoso” considerando sólo fuerzas repulsivas, incluirlas luego en forma perturbativa.

Estamos interesados en la distribución radial (a dilución infinita) de residuos apolares  $\rho'$  alrededor de uno  $\rho$  central, con todas las interacciones presentes,

$$W_{\rho\rho'}(r) = -k_B T \ln g_{\rho\rho'}(r; \rho_\alpha, T) \quad (4.14)$$

donde  $k_B$  es la constante de Boltzmann,  $\rho_\alpha$  la densidad del agua y  $W_{\rho\rho'}(r)$  es el potencial de fuerza media de pares. Definimos:

$$W_{\rho\rho'}^{(0)}(r) = -k_B T \ln g_{\rho\rho'}^{(0)}(r) \quad (4.15)$$

al potencial de fuerza media cuando despreciamos a las fuerzas atractivas. Si suponemos que la estructura del agua pura contiene todos los efectos de la interacción entre sus moléculas, la contribución de las interacciones atractivas asociadas con los residuos estará en  $\delta W_{\rho\rho'}(r) = W_{\rho\rho'}(r) - W_{\rho\rho'}^{(0)}(r)$ .

Teniendo en cuenta que trabajamos en el límite de dilución infinita, el medio que rodea a las moléculas de residuo es básicamente agua pura. Luego, en este límite, la  $g_{\rho\alpha}^{(0)}(r)$  debería ser prácticamente insensible a la inclusión de fuerzas atractivas. Por lo tanto, si nuestra representación física es correcta, esperamos que la función de correlación residuo–agua cambie muy poco al incluir a las fuerzas atractivas.

Ahora bien, para determinar la energía de solvatación de un par  $\rho\rho'$ , necesitaremos conocer la probabilidad de encontrar una molécula de agua en una región particular alrededor de las dos moléculas de residuo. Por consiguiente, para el cálculo del potencial efectivo de pares deberemos contar con información explícita sobre las correlaciones de tres cuerpos  $\rho\rho'\alpha$ .

Indicaremos con  $g_{\rho\alpha}^{ED}(r_\alpha)$  ( $g_{\rho'\alpha}^{ED}(r_\alpha)$ ) a la función de distribución de moléculas de agua en una capa de radio  $r_\alpha$  alrededor del residuo  $\rho$  ( $\rho'$ ); y con  $g_{\rho\alpha}^{ED}(r_\alpha | r)$  ( $g_{\rho'\alpha}^{ED}(r_\alpha | r)$ ) a la distribución radial de moléculas de agua alrededor de un residuo esférico  $\rho$  ( $\rho'$ ), cuando una segunda molécula de residuo  $\rho'$  ( $\rho$ ) se encuentra a una distancia fija  $r$  de la primera. Así, esta función de distribución representa la correlación entre dos moléculas de residuo apolar cuando están fijas formando un dímero de longitud  $r$ .

La estructura del agua alrededor de un dímero en solución acuosa dependerá, en principio, de las interacciones entre las moléculas de residuo que lo forman, y del agua. Para determinar las correlaciones intermoleculares, debemos introducir funciones de distribución intramoleculares que describan la estructura molecular del soluto. Alteramos la forma de las ecuaciones (4.10) para permitir la presencia de correlaciones intramoleculares, y obtenemos un conjunto de ecuaciones como las siguientes:

$$h_{\chi\alpha}^{(0)}(r) = \omega_{\chi\gamma} * c_{\gamma\alpha}^{(0)} + \rho_\alpha \omega_{\chi\gamma} * c_{\gamma\alpha}^{(0)} * h_{\alpha\alpha}^{\text{exp}}(r) \quad (4.16a)$$

$$h_{\chi\gamma}^{(0)}(r) = \omega_{\chi\delta} * c_{\delta\nu}^{(0)} * \omega_{\nu\gamma}(r) + \rho_\alpha \omega_{\chi\nu} * c_{\nu\alpha}^{(0)} * h_{\gamma\alpha}^{(0)}(r) \quad (4.16b)$$

En estas ecuaciones, los subíndices denotados con letras griegas están asociados con las moléculas de residuo (menos  $\alpha$  que corresponde al agua) y se encuentra implícita la sumatoria sobre subíndices griegos repetidos. El símbolo  $a * b(r)$  indica una integral de convolución:

$$a * b(r) = \int d\vec{r}' a(r') b(|\vec{r}' - \vec{r}|) \quad (4.17)$$

La función  $\omega_{\chi\gamma}(r)$  es la correlación intramolecular para el soluto:

$$\omega_{\chi\gamma}(r) = \delta_{\chi\gamma} \delta(\vec{r}) + (1 - \delta_{\chi\gamma}) \left\langle \delta(\vec{r} - \vec{r}_1^\chi + \vec{r}_1^\gamma) \right\rangle \quad (4.18)$$

donde  $\vec{r}_1^\chi$  denota la posición del sitio  $\chi$  en la molécula de residuo "1" y los corchetes angulares indican el promedio sobre el ensamble, en el equilibrio. Nuevamente usamos las relaciones de clausura

$$\begin{aligned} c_{\rho M}^{ED}(r) &= 0 & r > d_{\rho M} \\ g_{\rho M}^{ED}(r) &= 0 & r < d_{\rho M} \end{aligned} \quad (4.19)$$

donde  $M$  representa a una molécula de residuo ( $\rho'$ ) o de agua ( $\alpha$ ).

Las ecuaciones (4.16) y (4.19) se resuelven numéricamente. El método empleado está basado en la serie diagramática que es equivalente a la solución de esas ecuaciones integrales [6]. Resumimos el algoritmo usado para determinar a las funciones  $g_{\rho\alpha}^{ED}(r_\alpha | r)$  y  $g_{\rho'\alpha}^{ED}(r_\alpha | r)$  en las ecuaciones siguientes,

$$\hat{t}_{\chi\alpha}^{(n)}(k) = \hat{\omega}_{\chi\gamma}(k) \hat{f}_{\gamma\alpha}^{(n)}(k) \left[ 1 + \rho_\alpha \hat{h}_{\alpha\alpha}^{\text{exp}}(k) \right] - \hat{f}_{\chi\alpha}^{(n)}(k) \quad (4.20a)$$

$$f_{\chi\alpha}^{(n)}(r_\alpha) = \tau_{\chi\alpha}^{(n-1)}(r_\alpha) f_{\chi\alpha}^{(1)}(r_\alpha) \quad (4.20b)$$

$$f_{\chi\alpha}^{(1)}(r_\alpha) = \begin{cases} -1 & r_\alpha < d_{\chi\alpha} \\ 0 & r_\alpha > d_{\chi\alpha} \end{cases} \quad (4.20c)$$

$$\tau_{\chi\alpha}^{(0)}(r_\alpha) = 1 \quad (4.20d)$$

$$g_{\chi\alpha}^{ED}(r_\alpha | r) = \left[ 1 + f_{\chi\alpha}^{(1)}(r_\alpha) \right] \lim_{N \rightarrow \infty} \sum_{n=0}^N \tau_{\chi\alpha}^{(n)}(r_\alpha) \quad (4.20e)$$

Calculamos secuencialmente  $\tau_{\chi\alpha}^{(n)}$ , empezando por (4.20c) y (4.20d). Luego empleamos la técnica de transformadas rápidas de Fourier para las relaciones recursivas (4.20b) y (4.20a). La función  $g_{\chi\alpha}^{ED}(r_\alpha | r)$  converge monótonamente en menos de  $10^{-3}$  de su función límite para  $N \leq 70$ .

Luego de calcular a las funciones  $g_{\rho\alpha}^{ED}(r_\alpha | r)$  y  $g_{\rho'\alpha}^{ED}(r_\alpha | r)$  con el algoritmo anterior, determinamos a la función de distribución radial  $g_{\rho\rho'}(r)$  cuando hay fuerzas atractivas presentes a partir de,

$$\begin{aligned} \delta W_{\rho\rho'}(r) &= u_{\rho\rho'}^{(1)}(r) + \rho_\alpha \int d\vec{r}_\alpha \left[ g_{\rho\alpha}^{ED}(r_\alpha | r) - g_{\rho\alpha}^{ED}(r_\alpha) \right] u_{\rho\alpha}^{(1)}(r_\alpha) + \\ &\quad + \rho_\alpha \int d\vec{r}_\alpha \left[ g_{\rho'\alpha}^{ED}(r_\alpha | r) - g_{\rho'\alpha}^{ED}(r_\alpha) \right] u_{\rho'\alpha}^{(1)}(r_\alpha) \quad (4.21) \\ g_{\rho\rho'}(r) &= g_{\rho\rho'}^{(0)}(r) \exp \left[ -\delta W_{\rho\rho'}(r) / k_B T \right] \end{aligned}$$

que es exacta en el límite de  $\rho_\alpha$  pequeña [7].

Por otra parte, utilizamos la teoría de Andersen, Chandler y Weeks [8] conocida como la aproximación “EXP” con un doble propósito, calcular el efecto del término atractivo del potencial (4.3b) sobre la función de distribución radial residuo–agua de referencia, y chequear a la función de distribución residuo–residuo  $g_{\rho\rho'}(r)$  que arroja nuestro modelo. En el límite de dilución infinita (comparar con ec. (2.31)),

$$g_{\rho M}(r) = g_{\rho M}^{(0)}(r) \exp[C_{\rho M}(r)] \quad (4.22)$$

donde el potencial renormalizado  $C_{\rho M}(r)$  se determina resolviendo las siguientes ecuaciones integrales,

$$\begin{aligned} C_{\rho\alpha}(r) &= \Phi_{\rho\alpha}(r) + \rho_\alpha \Phi_{\rho\alpha} * h_{\alpha\alpha}^{\text{exp}}(r) \\ C_{\rho\rho'}(r) &= \Phi_{\rho\rho'}(r) + \rho_\alpha \Phi_{\rho\alpha} * [C_{\rho'\alpha}(r) + 2h_{\rho'\alpha}^{(0)}(r)] \\ \Phi_{\rho M}(r) &= -u_{\rho M}^{(1)}(r)/k_B T \quad r > d_{\rho M} \\ C_{\rho M}(r) &= 0 \quad r < d_{\rho M} \end{aligned} \quad (4.23)$$

nuevamente, los asteriscos indican integrales de convolución.

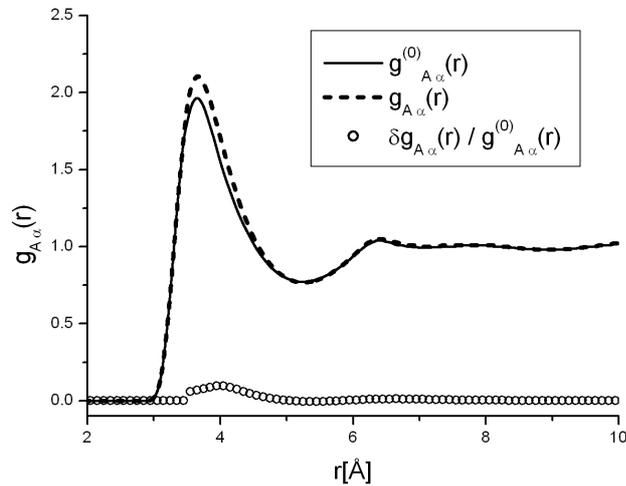


figura 4.2

Antes de entrar directamente en el cálculo y análisis de las funciones de distribución radial residuo–residuo y de los potenciales de fuerza media, verificamos mediante la ecuación (4.22), que efectivamente la función de

distribución radial residuo–agua cambia muy poco al considerar los efectos que introduce el término atractivo del potencial.

En la figura 4.2 repetimos el gráfico mostrado en el capítulo 3 para el residuo de alanina, como un ejemplo de los resultados que obtenemos cuando introducimos el efecto del término atractivo del potencial  $u_{\rho\alpha}^{(1)}(r)$ . En realidad, encontramos que la diferencia relativa entre las funciones de distribución es menor al 10%, para *todos* los residuos apolares (círculos abiertos pequeños).

En la sección siguiente presentamos las gráficas de las funciones de distribución radial residuo–residuo calculadas según (4.21) y las comparamos con las obtenidas según (4.22) a partir de las ecuaciones (4.23), para tres de los siete residuos apolares.

#### 4.1.2 APLICACIÓN DE LA TEORÍA A RESIDUOS APOLARES.

Recordemos que el conjunto de residuos apolares está integrado por siete aminoácidos: *alanina* (A), *valina* (V), *prolina* (P), *isoleucina* (I), *metionina* (M), *leucina* (L) y *fenilalanina* (F) [9].

Si bien aplicamos la teoría anteriormente expuesta a todo el grupo de residuos de aminoácidos apolares, por razones de espacio decidimos no mostrar las gráficas correspondientes a todos los pares de residuos. Únicamente presentaremos aquellos casos que consideramos interesantes analizar.

##### 4.1.2.1 Función de distribución de pares.

Comenzamos analizando los resultados correspondientes al par *alanina–alanina*, y luego exponemos dos casos más: *alanina–valina* y *valina–valina*.

En la figura 4.3, mostramos las gráficas de las funciones  $\delta W_{AA}(r)$  y  $g_{AA}(r)$  calculadas con las ecuaciones (4.21).

En la figura 4.3a observamos la contribución que las interacciones atractivas hacen a la distribución radial de residuos de alanina alrededor de otro residuo idéntico en el origen. Como vemos, la función  $\delta W_{AA}(r)/k_B T$  toma un valor cercano a uno para  $r \approx \sigma_A$  y a partir de allí, disminuye monótonamente a cero.

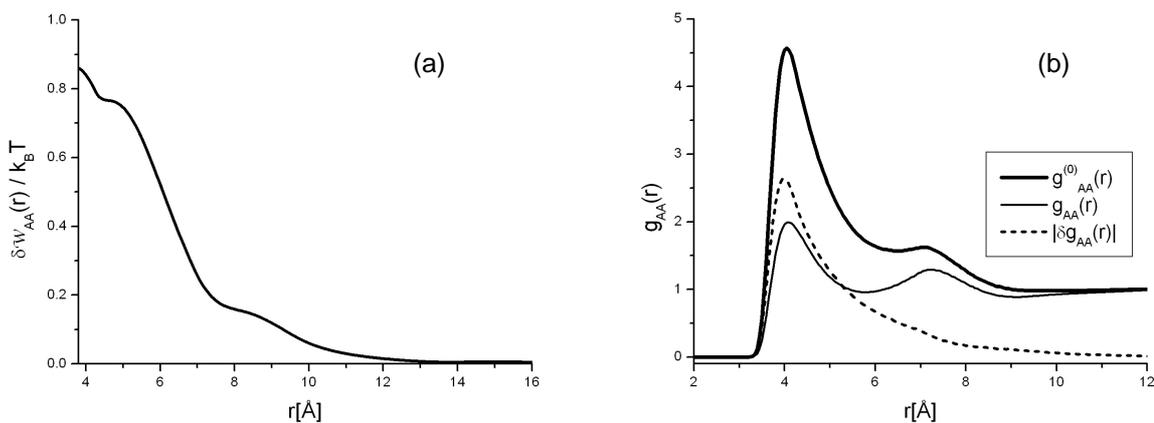


figura 4.3

La línea llena y gruesa de la figura 4.3b corresponde a la función de distribución radial obtenida cuando los residuos interactúan a través de un potencial repulsivo  $u_{pp'}^{(0)}(r)$ . Obtenemos la curva llena y fina, cuando introducimos en forma perturbativa (4.21) la contribución que hacen las interacciones atractivas. La curva con línea de trazos muestra el efecto neto del término atractivo.

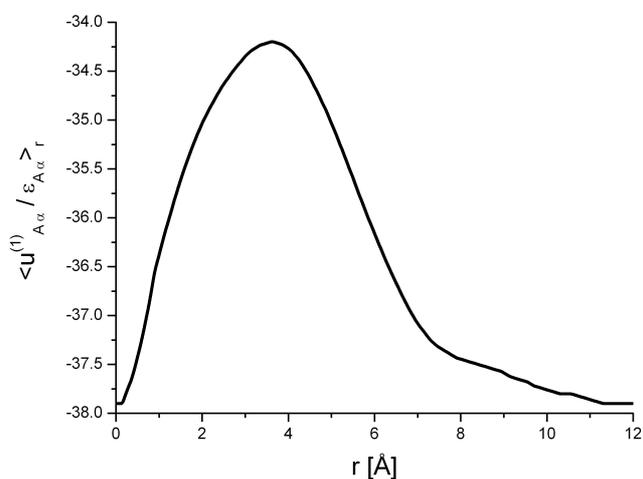


figura 4.4

La energía de solvatación del par AA definida como:

$$\left\langle \frac{u_{A\alpha}^{(1)}}{\epsilon_{A\alpha}} \right\rangle_r = 2\rho_\alpha \int d\vec{r}_\alpha \frac{g_{A\alpha}^{ED}(r_\alpha | r) u_{A\alpha}^{(1)}(r_\alpha)}{\epsilon_{A\alpha}} \quad (4.24)$$

es el término que da la predicción de campo medio en el cálculo de  $\delta W_{AA}(r)$  de la ecuación (4.21). En la figura 4.4 graficamos a la energía de solvatación en función de la distancia  $r$  entre moléculas de residuo de alanina que forman el dímero.

Como dijimos en el apartado anterior, si trabajamos a dilución infinita y consecuentemente la función  $g_{\rho\alpha}^{(0)}(r)$  cambia muy poco al incluir fuerzas atractivas (figura 4.2), entonces podemos esperar que la aproximación EXP prediga resultados comparables a los de nuestra aproximación de campo medio.

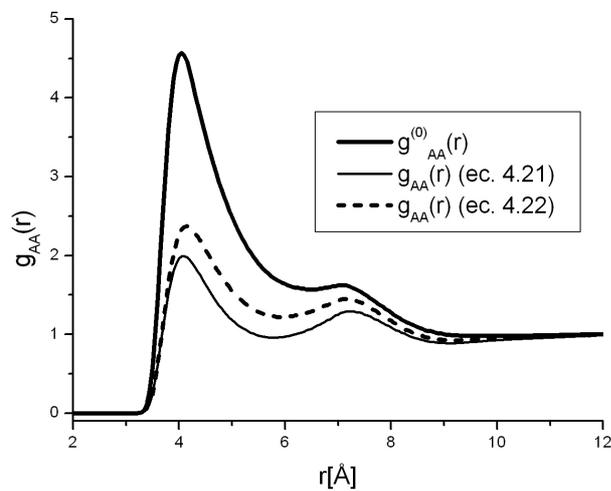


figura 4.5

En la figura 4.5 mostramos a las funciones de distribución radial obtenidas con ambas aproximaciones, graficadas junto a la función de referencia  $g_{AA}^{(0)}(r)$ .

Podemos observar que, si bien existe una diferencia cuantitativa, cualitativamente ambas funciones de distribución muestran el mismo comportamiento. En los dos casos, la inclusión de fuerzas atractivas produce una disminución en la altura del pico principal casi a la mitad de su valor, luego presentan un segundo pico menos pronunciado alrededor de los 7.5 Å ( $\approx (\sigma_A + \sigma_\alpha)$ ) y ambas se hacen prácticamente iguales a uno, para distancias algo mayores.

Es interesante destacar la presencia de un segundo pico, separado del primero por un valor correspondiente al diámetro de una molécula de agua, y la

ausencia de más picos en la función de distribución de pares de residuos, ya que éste es el comportamiento que se requiere experimentalmente [10].

Finalizamos esta sección presentando a las funciones de distribución radial correspondientes a los pares *alanina–valina* y *valina–valina* calculadas según (4.8), (4.21) y (4.22). Para los restantes residuos apolares, las funciones se comportan de un modo semejante al exhibido por éstos.

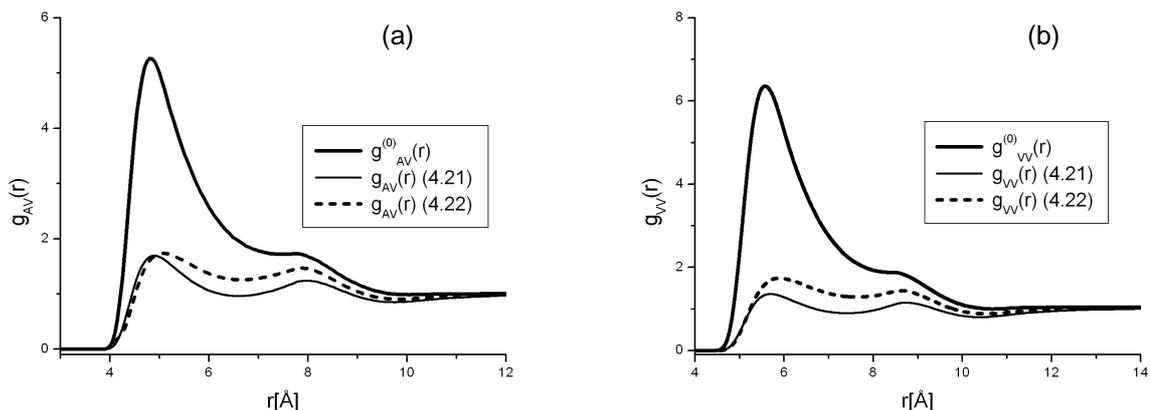


figura 4.6

Como señalamos anteriormente, la función de distribución radial  $g_{\rho\alpha}(r)$  brinda información sobre la densidad relativa de moléculas de agua a una cierta distancia  $r$  de la molécula de residuo  $\rho$ .

En la figura 4.2, observamos que al incluir en los cálculos el término atractivo del potencial, la función de distribución radial aumenta levemente respecto de su valor de referencia  $g_{\rho\alpha}^{(0)}(r)$ . Fundamentalmente, vemos un aumento en la altura del pico principal, y podríamos decir que, como era de esperar, al incluir fuerzas atractivas soluto–solvente se produce una *mayor* hidratación del residuo.

Por otra parte, en el sistema formado por dos residuos de aminoácidos apolares disueltos en agua a dilución infinita, consideramos a la función de distribución radial residuo–residuo  $g_{\rho\rho'}(r)$  como una medida de la probabilidad de encontrar un segundo residuo a cierta distancia  $r$  del primero. En este caso, la acción de fuerzas atractivas entre las moléculas que forman el sistema (4.21), provoca una disminución considerable de la función de distribución radial residuo–

residuo para todo valor de  $r$ , principalmente para distancias cercanas al contacto (figura 4.5).

Si restamos a los potenciales de campo medio la interacción directa entre residuos,

$$\begin{cases} W_{IH}(r) = W_{\rho\rho'}(r) - u_{\rho\rho'}(r) \\ W_{IH}^{(0)}(r) = W_{\rho\rho'}^{(0)}(r) - u_{\rho\rho'}^{(0)}(r) \end{cases} \quad (4.25)$$

donde el superíndice (0) indica *ausencia de fuerzas atractivas*, encontramos que,

$$W_{IH}(r) - W_{IH}^{(0)}(r) \propto \delta W_{\rho\rho'}(r) = -\frac{1}{\beta} \ln \left[ \frac{g_{\rho\rho'}(r)}{g_{\rho\rho'}^{(0)}(r)} \right] > 0 \quad (4.26)$$

De acuerdo con este resultado, la interacción hidrofóbica sería muy fuerte en ausencia de fuerzas atractivas entre los residuos, y se debilita notablemente al incluirlas. En una primera mirada, este hecho es contrario a lo que intuitivamente esperaríamos obtener: ¿por qué al incluir fuerzas atractivas entre los residuos, la probabilidad de encontrarlos en contacto disminuye? Procuraremos dar una interpretación de este resultado.

Al formular la pregunta del párrafo anterior, cuestionábamos el cambio observado en la  $g_{\rho\rho'}(r)$  en relación con la interacción atractiva directa *residuo–residuo* utilizada en el cálculo. A continuación, daremos un argumento que nos permite entender de otro modo, el comportamiento de las funciones de correlación de pares.

Hasta aquí, en nuestro razonamiento omitimos considerar la influencia que tiene la función de distribución radial agua–agua. Ésta función, obtenida en forma experimental, contiene implícitamente *todas* las contribuciones de la interacción *solvente–solvente*. Si recordamos además que los cálculos se llevan a cabo en el límite de dilución infinita, concluimos que la incidencia de la  $g_{\alpha\alpha}^{\text{exp}}(r)$  será muy importante, mucho más que la interacción directa *residuo–residuo*. Es por esto que, para analizar las gráficas 2 y 5, nos concentraremos en la influencia que presentan las dos interacciones más importantes: *residuo–agua* y *agua–agua*.

Desde este punto de vista, podemos pensar que la gran “atracción hidrofóbica” que muestran las moléculas de residuo apolar para el potencial de referencia, es producto de la “preferencia” a pasar al seno del líquido que exhiben

las moléculas de agua, en lugar de permanecer en la superficie de contacto con el residuo. Ésta preferencia estaría gobernada por la competencia que se establece entre las interacciones *residuo–agua* versus *agua–agua*, en la que se impondría claramente ésta última.

Al introducir fuerzas atractivas residuo–agua, la curva correspondiente a la hidratación hidrofóbica muestra sólo un pequeño cambio cuantitativo, como cabe esperar, si recordamos que el sistema en este caso está formado por una única molécula de residuo disuelta en agua (dilución infinita).

En cambio, cuando el sistema está integrado por más de un residuo, la interacción hidrofóbica entre ellos se ve fuertemente afectada. Esto sucedería porque al incluir fuerzas atractivas en la interacción residuo–agua, las moléculas de solvente estarían menos “interesadas” en abandonar la superficie de contacto con el residuo y pasar al seno del líquido, respecto de cuando considerábamos sólo interacciones repulsivas. Por esta razón, observamos que al introducir fuerzas atractivas entre las moléculas del sistema, la “atracción hidrofóbica” entre los residuos apolares disminuye.

Según este razonamiento, la interacción hidrofóbica entre residuos apolares, estaría gobernada, esencialmente, por la “preferencia relativa” que presentan las moléculas de agua a permanecer en contacto con el residuo o pasar al seno del líquido, más que por la interacción directa entre los dos residuos apolares.

Para la explicación anterior, nos apoyamos en las gráficas correspondientes al residuo de alanina. Los mismos argumentos podrían utilizarse con cualquiera de los siete aminoácidos apolares.

#### 4.1.2.2 **Potenciales efectivos de pares.**

Los potenciales efectivos de interacción entre pares de residuos de aminoácidos en función de la distancia, calculados a partir de (4.14) y (4.21), constituyen uno de los resultados más importantes de esta tesis, por sus posibles aplicaciones en el estudio del plegamiento y estabilidad de proteínas.

Como antes, no mostraremos a todas las gráficas posibles (28 en total), sino más bien las de aquellos residuos que permiten tener una visión general de los resultados obtenidos. Elegimos ilustrar lo que ocurre con todo el grupo a través de

tres residuos: *alanina*, *valina* y *fenilalanina*. El de menor tamaño, uno intermedio y el mayor de todo este grupo.

Nuevamente comenzamos con el residuo de alanina. En la figura 4.7 incluimos las gráficas del potencial de interacción entre dos residuos de alanina en el vacío (línea de trazos), y del potencial de fuerza media que contiene los efectos inducidos por el solvente (agua).

En este último caso, observamos la presencia de dos mínimos separados por una barrera, el primero alrededor de los 4 Å correspondiente al par de residuos en contacto, y un segundo mínimo (menos profundo) alrededor de los 7.5 Å, donde los dos residuos de alanina se encontrarían separados por una molécula de agua. Además, vemos que para distancias mayores a 12 Å la interacción entre ambos prácticamente desaparece.

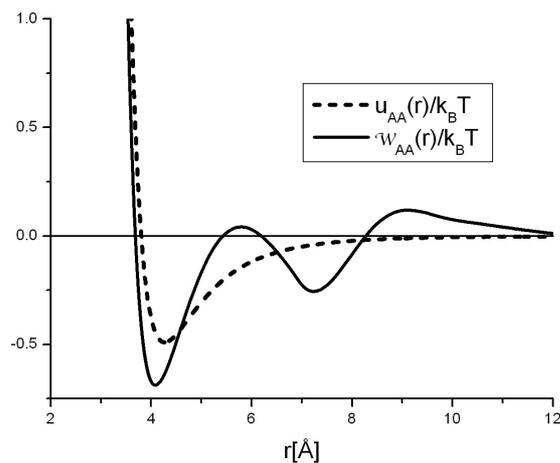


figura 4.7

En la figura 4.8 mostramos a los potenciales efectivos de pares entre: (a) dos residuos iguales, y (b) dos residuos distintos. Estos potenciales presentan características similares para todos los residuos apolares, al menos en términos cualitativos. A distancias menores que el contacto se hacen fuertemente repulsivos, le siguen dos pozos y dos barreras bien definidas y a partir de allí tienden rápidamente a cero.

El primer pozo es el más profundo y se encuentra a una distancia cercana al valor de contacto para los residuos involucrados. El segundo es menos profundo y

se encuentra separado del primero por una distancia del orden del tamaño de una molécula de agua ( $\sigma_\alpha \approx 3.2 \text{ \AA}$ ).

Elegimos a estos tres residuos (del conjunto de siete), con el propósito de estudiar la relación (si existiera alguna) entre el potencial de fuerza media y el tamaño de los residuos.

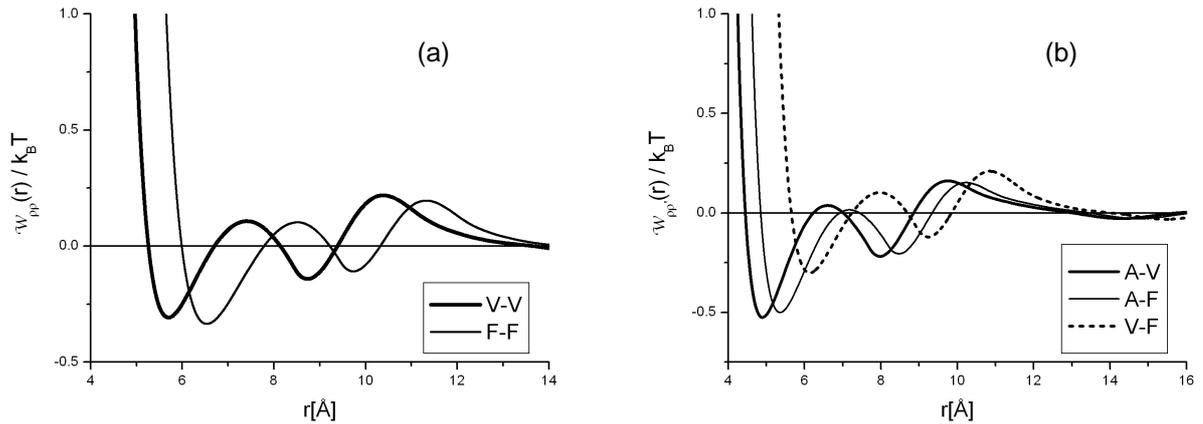


figura 4.8

En estas condiciones, si consideramos la profundidad de los pozos y la altura relativa de las barreras, observamos una mayor interacción efectiva entre dos residuos de alanina (el más pequeño), que entre dos residuos de fenilalanina (el más grande).

Sin embargo, el residuo de valina (también de menor tamaño que el de fenilalanina), se comporta diferente a la alanina. Vemos que, de éstos resultados, no puede afirmarse que exista una relación directa entre la intensidad de la interacción efectiva y el tamaño de los residuos.

## 4.2 COMPARACIÓN CON SIMULACIÓN.

Cerramos este capítulo procurando brindar una referencia más cercana, una prueba de la “racionalidad” de éstos resultados. Con este propósito, buscamos un potencial de interacción entre pares de solutos apolares disueltos en agua, que haya sido determinado como función de la distancia, y por un camino diferente al transitado por nosotros.

Concretamente, el potencial de fuerza media que utilizaremos en la comparación, es el que calcularon W. Young y col. [11] mediante simulación con Dinámica Molecular entre dos moléculas de metano.

Si bien ellos estaban interesados en estudiar la influencia del solvente sobre la interacción *metano–metano*, y utilizaron tres modelos diferentes para representar al agua líquida, nosotros estableceremos nuestra comparación con uno sólo de sus resultados (TIP4P). Elegimos a la curva correspondiente a este modelo de agua porque es el que tiene los parámetros  $\sigma_\alpha$  y  $\varepsilon_\alpha$  más parecidos a los del SPC/E, utilizado en nuestros cálculos.

En la figura 4.9a mostramos a las funciones de distribución radial soluto–agua obtenidas mediante: Dinámica Molecular con Gromos96 (círculos huecos), el modelo mecánico-estadístico desarrollado en este capítulo (línea entera), y Dinámica Molecular determinada por W. Young y col. [11] (línea de trazos).

Si bien el soluto apolar considerado en las dos primeras curvas es el residuo de alanina y en la última metano, los valores asignados a los parámetros  $\sigma_\rho$  y  $\varepsilon_\rho$  son muy similares y permiten establecer una comparación directa entre estos resultados.

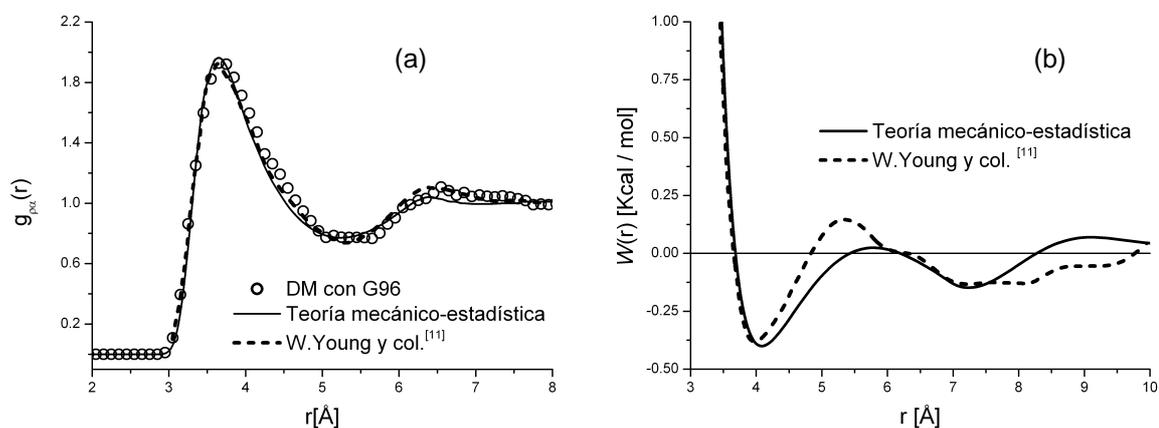


figura 4.9

En la figura 4.9a observamos un muy buen acuerdo entre las tres funciones de distribución radial soluto–agua. En la figura 4.9b graficamos los potenciales de fuerza media correspondientes a alanina–alanina (línea llena) y metano–metano (línea de trazos). El primero de estos resultados es obtenido con la teoría

mecánico–estadística presentada en este capítulo, mientras que el otro surge de una simulación mediante Dinámica Molecular.

Como vemos, ambos potenciales muestran un buen acuerdo respecto de la profundidad y posición de los pozos, y prácticamente se anulan para distancias mayores a los 10 Å. Además, observamos que en ambos casos el mínimo correspondiente al contacto resulta  $0.2 - 0.3 \text{Kcal mol}^{-1}$  menor que el mínimo a solvente separado, coincidiendo con lo calculado por muchos otros autores [11].

### 4.3 REFERENCIAS DEL CAPÍTULO.

- 1 L.R. Pratt and D. Chandler, *J. Chem. Phys.*, **67**, 3683 (1977).
- 2 A.K. Soper, *J. Chem. Phys.*, **101**, 6888 (1994).
- 3 R.A. Pierotti, *J. Phys. Chem.*, **69**, 281(1965); R.A. Pierotti, *Chem. Rev.*, **76**, 717 (1975); F.H. Stillinger, *J. Solution Chem.*, **2**, 141 (1973).
- 4 J.P. Hansen and I.R. McDonald, *Theory of Simple Liquids* (Academic Press, New York, 1986).
- 5 J.D. Weeks, D. Chandler and C. Andersen, *J. Chem. Phys.*, **54**, 5237 (1971).
- 6 L.R. Pratt and D. Chandler, *J. Chem. Phys.*, **73**, 3430 (1980).
- 7 L.R. Pratt and D. Chandler, *J. Chem. Phys.*, **73**, 3434 (1980).
- 8 C. Andersen, D. Chandler and J.D. Weeks, *Adv. Chem. Phys.*, **34**, 105 (1979).
- 9 C. Branden and J. Tooze, *Introduction to Protein Structure* (Garland Publishing, Inc., New York, 1991).
- 10 A. Pertsemlidis, A.K. Soper, J.M. Sorenson and T. Head–Gordon, *Proc. Natl. Acad. Sci. USA*, **96**, 481 (1999).
- 11 W.S. Young and C.L. Brooks III, *J. Chem. Phys.*, **106**, 9265 (1997).

## 5. EXTENSIÓN DE LA TEORÍA MECÁNICO-ESTADÍSTICA.

A continuación, presentamos una extensión de la teoría utilizada en los dos capítulos anteriores para residuos de aminoácidos apolares, y la aplicamos a los residuos polares y cargados.

Comenzamos proponiendo potenciales de interacción directa de pares dependientes solamente de la distancia entre moléculas. Luego los separamos en dos términos, uno repulsivo de corto alcance, y otro atractivo de largo alcance con el que calculamos el efecto producido sobre las funciones de distribución radial residuo–residuo de referencia. Finalmente, obtenemos los correspondientes potenciales efectivos de interacción entre pares de residuos en función de la distancia.

Por último, tomamos los valores de potencial al contacto surgidos de nuestro modelo teórico y los comparamos con la energía efectiva de pares de residuos publicados por otros autores.

### 5.1 AMINOÁCIDOS POLARES Y CARGADOS.

Sabemos que el efecto hidrofóbico tiene un gran protagonismo en el plegamiento y estabilidad de las proteínas en condiciones fisiológicas normales. Sin embargo, no es la única interacción presente.

Kauzmann [1] consideró que las otras interacciones (coulombiana, puente de hidrógeno, etc.) poseían una incidencia menor, prácticamente despreciable. Actualmente, se cree que el plegamiento de proteínas globulares es el resultado de un delicado balance entre un conjunto de fuerzas, algunas de las cuales favorecen el estado colapsado y otras se oponen a él.

Para determinar experimentalmente con precisión la incidencia de cada una de las contribuciones (en el hipotético caso que las conociéramos a todas), necesitaríamos de mediciones termodinámicas muy precisas durante el proceso de *plegado* y *desplegado* de proteínas en solución acuosa diluida, algo que resulta verdaderamente difícil de conseguir.

En este capítulo, nos planteamos obtener alguna información sobre las “fuerzas efectivas” entre pares de residuos, cualquiera sea su polaridad. Para

esto, proponemos potenciales de interacción directa apropiados para cada caso, y los introducimos en la teoría desarrollada anteriormente para estudiar al efecto hidrofóbico.

No tenemos una justificación rigurosa de la validez del modelo utilizado en los capítulos anteriores para el caso de partículas con carga, o momento dipolar. Sin embargo, decidimos avanzar en la determinación de los potenciales de campo medio, y juzgar luego si los resultados obtenidos son razonables o no, a partir del acuerdo (o desacuerdo) que muestren respecto de los valores publicados por otros autores sobre la base de otras ideas y métodos de cálculo.

### 5.1.1 POTENCIALES DE INTERACCIÓN DIRECTA.

Como adelantamos, la idea es aplicar la teoría desarrollada en los capítulos anteriores a los aminoácidos cargados y polares. Para esto, debemos proponer potenciales de interacción directa que sean independientes de las orientaciones de las moléculas. Sin entrar en detalles, digamos que la gráfica para  $u_{\rho M}(r)$  será siempre de la forma mostrada en la figura 5.1a, y en todos los casos podremos separar al potencial en dos términos como hicimos con el potencial de Lennard-Jones (figura 5.1b).

Si los residuos involucrados en la interacción son *apolares*, utilizamos un potencial del tipo Lennard-Jones (12-6) para representar ambas interacciones, *residuo-residuo* y *residuo-agua* (ec. 2.32).

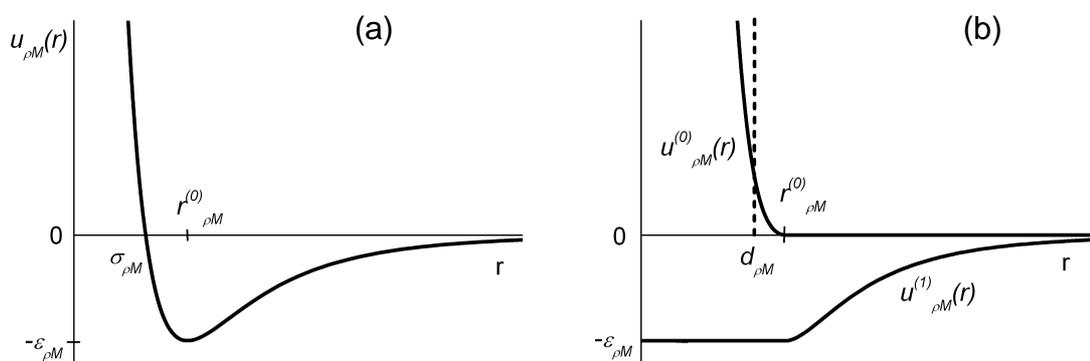


figura 5.1

Para el resto de las interacciones consideraremos un potencial formado por un término repulsivo idéntico al utilizado hasta ahora para residuos apolares, más un término atractivo que elegimos en cada caso y que no depende de las orientaciones sino sólo de la distancia entre moléculas. Así, cuando el sistema se encuentre formado por dos residuos de aminoácidos polares disueltos en agua, consideraremos a los residuos como *esferas blandas* con momento dipolar.

Para representar a la interacción entre las moléculas de residuos y el agua, proponemos un potencial de la forma:

$$u_{\rho M}^{DD}(r) = 4\varepsilon_{\rho M} \left( \frac{\sigma_{\rho M}}{r} \right)^{12} - k_B T \ln \left[ \frac{1}{2} \int_{-1}^1 dz i_0 \left[ \frac{1}{k_B T} \frac{\mu_\rho \mu_M}{4\pi\varepsilon_0 r^3} \sqrt{1+3z^2} \right] \right] \quad (5.1)$$

donde  $i_0(x) = \text{senh}(x)/x$  es la función de Bessel modificada esférica de orden cero. La forma funcional para el término atractivo del potencial  $u_{\rho M}^{DD}(r)$  surge de promediar angularmente la interacción usual *dipolo–dipolo* (DD) [2].

Los parámetros involucrados en el potencial, son detallados en el *Apéndice C* de esta tesis. El valor  $\mu_\rho$  asignado a los residuos polares fue ajustado para conseguir el mejor acuerdo posible entre las funciones de distribución radial residuo–agua obtenida con este modelo y la encontrada con simulación mediante Dinámica Molecular.

La interacción utilizada en el caso de dos residuos cargados es del tipo *ión–ión* (II) entre ellos, y del tipo *ión–dipolo* (ID) entre cada uno de los residuos y el agua. El potencial propuesto para la interacción entre estos residuos es de la forma,

$$u_{\rho\rho'}^{II}(r) = 4\varepsilon_{\rho\rho'} \left( \frac{\sigma_{\rho\rho'}}{r} \right)^{12} - \frac{q_\rho q_{\rho'}}{4\pi\varepsilon_0 r} \quad (5.2)$$

donde  $q_\rho$  ( $q_{\rho'}$ ) denota a la carga efectiva asociada a cada residuo y  $\varepsilon_0$  es la permitividad dieléctrica del vacío.

Para caracterizar la interacción entre un residuo cargado y otro polar o entre un residuo cargado y el agua, necesitamos introducir una función potencial más. Empleamos el mismo procedimiento usado para  $u_{\rho M}^{DD}(r)$ , y promediamos

angularmente al término atractivo para que el potencial resultante dependa sólo de la distancia entre moléculas,

$$u_{\rho M}^{ID}(r) = 4\varepsilon_{\rho M} \left( \frac{\sigma_{\rho M}}{r} \right)^{12} - k_B T \ln \left[ i_0 \left[ \frac{1}{k_B T} \frac{q_{\rho} \mu_M}{4\pi \varepsilon_0 r^2} \right] \right] \quad (5.3)$$

la función  $i_0(x)$  es nuevamente la de Bessel modificada esférica de orden cero.

En el siguiente apartado presentaremos las gráficas de las funciones de distribución radial residuo–agua y residuo–residuo determinadas utilizando las funciones descritas matemáticamente por las ecuaciones: 2.32 (LJ), 5.1 (DD), 5.2 (II) y 5.3 (ID). Para mayor claridad, resumimos en la Tabla 5.1 a las seis interacciones posibles entre pares de residuos.

Tabla 5.1: Potenciales de interacción directa residuo-residuo y residuo-agua.

RESIDUOS	$u_{\rho\rho'}(r)$	$u_{\rho\alpha}(r)$	$u_{\rho'\alpha}(r)$
1 <i>apolar–apolar</i>	LJ	LJ	LJ
2 <i>apolar–cargado</i>	LJ	LJ	ID
3 <i>apolar–polar</i>	LJ	LJ	DD
4 <i>cargado–cargado</i>	II	ID	ID
5 <i>cargado–polar</i>	ID	ID	DD
6 <i>polar–polar</i>	DD	DD	DD

### 5.1.2 FUNCIÓN DE CORRELACIÓN DE PARES.

Como en el capítulo anterior, determinamos primero las funciones de distribución radial para el potencial de referencia  $u_{\rho M}^{(0)}(r)$ , y luego calculamos la corrección introducida por el término atractivo  $u_{\rho M}^{(1)}(r)$ .

En las gráficas siguientes (figura 5.2) mostramos las funciones de distribución residuo–agua obtenidas para dos aminoácidos polares: *serina* (S) y *triptófano* (W). Si bien observamos un muy buen acuerdo entre las funciones correspondientes a la simulación y al modelo teórico para el residuo de serina,

debemos señalar que la función obtenida mediante Dinámica Molecular está tomada con respecto a la posición que ocupa el  $C_\beta$  del residuo, y no con respecto al centro geométrico de la molécula como lo habíamos hecho con los residuos apolares.

Esto se debe a que las moléculas de agua pueden ocupar posiciones muy cercanas a los átomos de los residuos polares (y cargados) durante la simulación. Considerando además que el centro geométrico de los residuos se encuentra fuera del volumen excluido por los átomos que forman la cadena lateral, la función de distribución radial que obtenemos respecto de ese punto, toma valores significativos para  $r=0$ . Por esta razón, elegimos calcular a la función de distribución radial de simulación con respecto a uno de los átomos que forman el residuo. Para el residuo serina elegimos al  $C_\beta$  y para el triptófano al  $C_{\delta_2}$ .

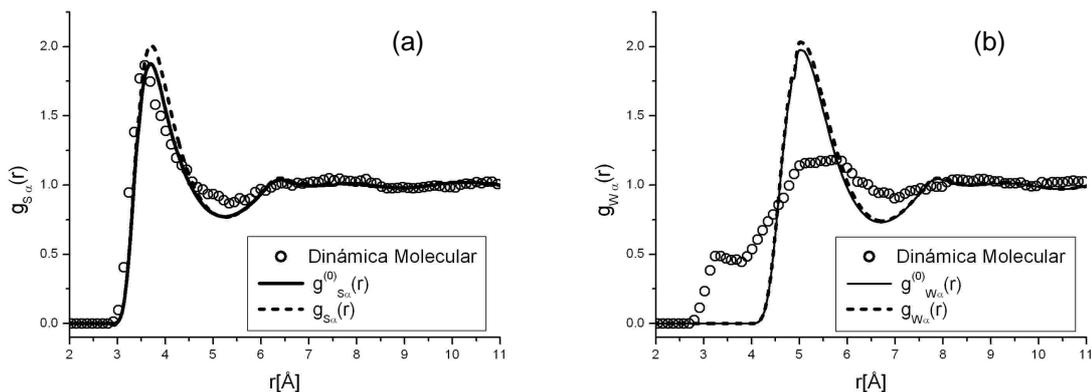


figura 5.2

En el triptófano, como en la fenilalanina, observamos que la función de distribución radial obtenida mediante simulación da cuenta de la forma plana de la cadena lateral, presentando un primer pico alrededor de los  $3.5 \text{ \AA}$  y otro más importante entre los  $5$  y  $6 \text{ \AA}$ . Desde luego, si suponemos esférico al residuo, con la teoría sólo podemos aspirar a reproducir el pico más alto asociado con el valor de  $\sigma_W$ .

Procederemos en forma análoga para determinar a las funciones de distribución radial residuo-agua correspondientes a los residuos cargados. Para el ácido aspártico (D) y el ácido glutámico (E) consideramos al  $C_\gamma$ , para lisina (K)

al  $C_\delta$  y para arginina (R) al  $C_\epsilon$ . Elegimos a estos átomos como referencia porque ocupan posiciones intermedias en la cadena lateral de cada uno de los aminoácidos cargados.

Como ejemplo de las funciones de distribución radial residuo–agua que obtenemos mediante simulación, mostramos dos curvas: una correspondiente a un residuo con carga positiva (figura 5.3a) y otra a uno con carga negativa (figura 5.3b).

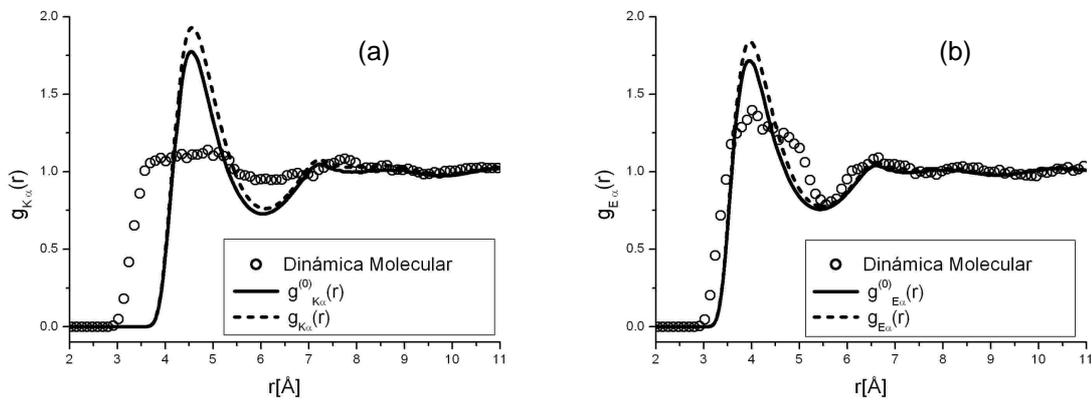


figura 5.3

Las curvas graficadas con líneas llenas y líneas de trazo surgen de la teoría y fueron obtenidas ajustando los parámetros de modo tal que el  $N_c$  de cada una coincida (dentro de un 1%) con la alcanzada mediante simulación.

Observamos que en los residuos con carga negativa el acuerdo es algo mejor, quizás esto se deba al tamaño más pequeño de los mismos. En cambio, los aminoácidos con carga positiva tienen una cadena lateral lineal con una conformación más “alargada”, y por lo tanto, más alejada de lo que consideramos en nuestro modelo como una esfera.

Finalizamos esta sección mostrando las gráficas correspondientes a las correlaciones residuo–residuo para los cuatro aminoácidos anteriores.

Observamos en la figura 5.4a que las curvas correspondientes a los residuos polares se comportan todas cualitativamente del mismo modo. Todas muestran un pequeño pico y luego tienden a uno a partir de los 14 Å, aproximadamente.

En la figura 5.4b graficamos a las funciones correspondientes a los residuos cargados. En este caso las curvas correspondientes a residuos con carga de igual

signo se comportan de un modo semejante independientemente de que sean positivas o negativas. Pero, como era de esperar, vemos una gran diferencia para la función de distribución radial de residuos con cargas opuestas. Si bien en todos los casos estas funciones tienden a uno, ninguna de las tres alcanza ese valor aún para distancias tan grandes como 50 Å .

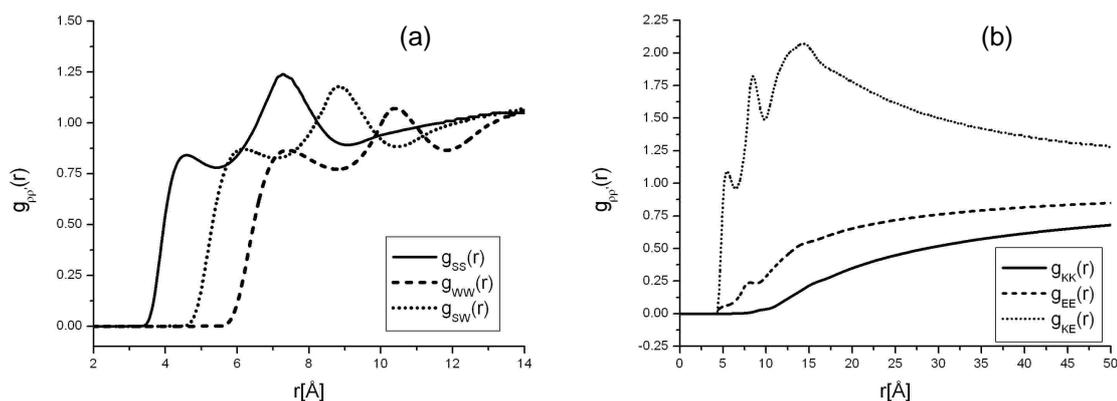


figura 5.4

### 5.1.3 POTENCIALES EFECTIVOS DE PARES.

Como detallamos en el capítulo 1 (Tabla 1.1), podemos clasificar a los veinte aminoácidos encontrados en la naturaleza según su polaridad en: apolares, polares y cargados [3]. A estos últimos a su vez, los subdividimos en cargados positivos y negativos.

El más simple de los 20 aminoácidos es la glicina (gly). Su molécula de residuo está compuesta por un único átomo de hidrógeno. Por este motivo, en muchos trabajos de investigación sobre proteínas, utilizan a este aminoácido para estimar propiedades relacionadas con los demás residuos, tomándolo como representativo de los átomos de la cadena principal.

En nuestro modelo, despreciamos el tamaño del residuo de glicina y consideramos nula su interacción con los demás residuos. Luego, con los 19 aminoácidos restantes obtenemos un conjunto de 190 potenciales de pares.

### 5.1.3.1 Aminoácidos polares.

En este grupo incluimos a ocho aminoácidos: serina (S), treonina (T), cisteína (C), asparagina (N), glutamina (Q), histidina (H), tirosina (Y) y triptófano (W) [3].

No mostraremos a todos los potenciales de pares que obtenemos para este grupo (36 en total), sino sólo algunos de ellos. En particular, nos concentraremos sobre los dos residuos seleccionados anteriormente, serina y triptófano.

Elegimos a la serina por ser el aminoácido polar de más frecuente aparición en proteínas [4], y al triptófano por ser el aminoácido polar donde el modelo teórico presenta su peor resultado.

A continuación, presentamos a los potenciales efectivos de interacción de pares para los dos residuos polares mencionados.

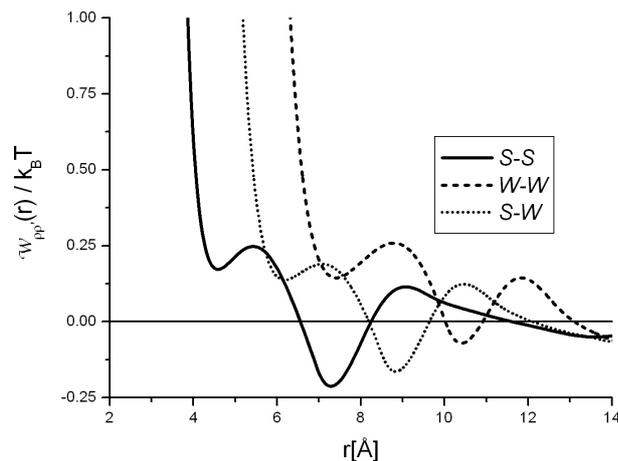


figura 5.5

En la figura 5.5 comparamos a los potenciales efectivos de pares de SS, WW y SW. Estos potenciales exhiben características similares para todos los residuos polares, al menos en términos cualitativos. A distancias menores que el contacto se hacen fuertemente repulsivos, poseen un mínimo (negativo) a una distancia aproximada de  $(\sigma_{pp'} + \sigma_{\alpha})$ , y le sigue una barrera a partir de la cual tienden a cero desde valores negativos.

Las tres curvas anteriores presentan un primer “pozo” ( $r \approx d_{pp'}$ ) que toma en todos los casos un valor positivo. Sin embargo, esto mismo no se repite para

todos los residuos polares. Por ejemplo, para asparagina (figura 5.6) vemos apenas una inflexión de la curva de potencial en  $r \approx d_{NN}$ .

Para todos los pares de residuos, este primer “pozo” (positivo) y el segundo (negativo), se encuentran separados por una distancia algo menor que el tamaño de una molécula de agua. A este segundo mínimo de valor negativo le sigue una barrera de  $(0.23-0.33) k_B T$ , y luego vuelve a hacerse negativo a una distancia aproximada a  $2\sigma_\alpha$  del mínimo anterior. A partir de allí el potencial efectivo tiende a cero tomando valores (en todos los casos) mayores a  $-0.1 k_B T$ .

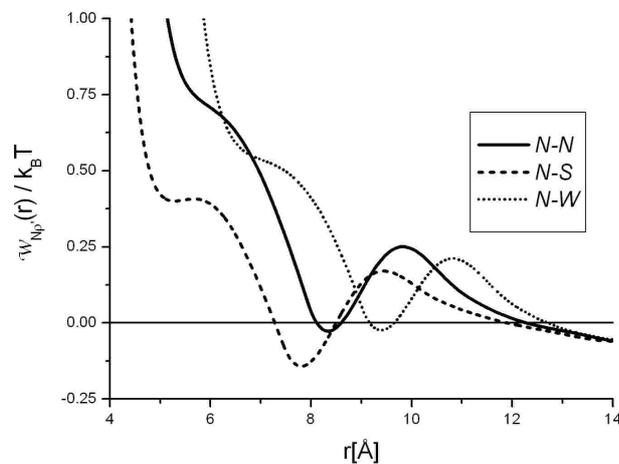


figura 5.6

### 5.1.3.2 Aminoácidos cargados.

Este grupo cuenta con un total de cuatro aminoácidos, dos con carga positiva: lisina (K) y arginina (R), y dos con carga negativa: ácido aspártico (D) y ácido glutámico (E).

En este conjunto, los cuatro aminoácidos cargados tienen un valor similar de frecuencia de aparición en proteínas (5–6%). Por lo tanto, no podemos usar el criterio que tomamos anteriormente para seleccionar los residuos a mostrar.

Considerando que los potenciales efectivos de pares no son tantos en este conjunto (10 en total), aprovechamos para mostrarlos a todos y así dar una idea del grado de diferencias que obtenemos entre potenciales de aminoácidos pertenecientes a un mismo grupo.

En las dos primeras gráficas de la figura 5.7 mostramos las curvas correspondientes a las interacciones efectivas entre residuos con cargas de igual signo. En las gráficas (c) y (d) presentamos a los potenciales efectivos entre residuos con cargas de signos opuestos.

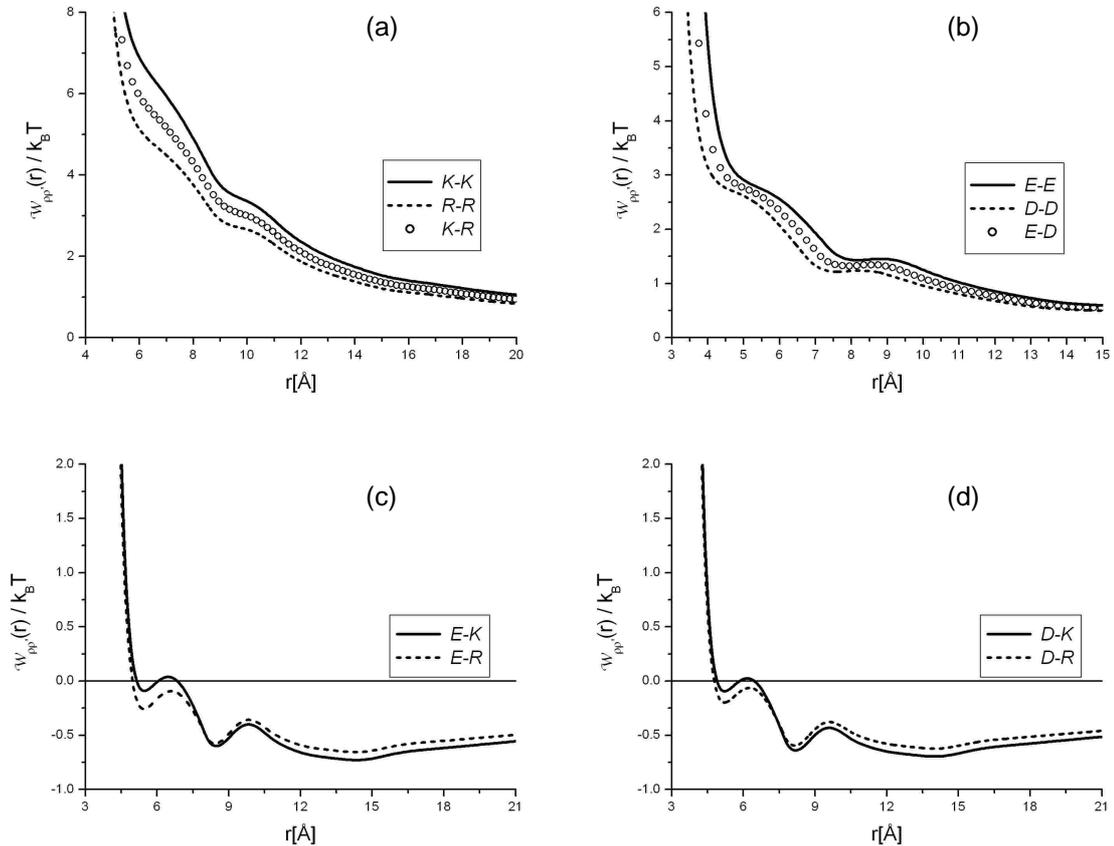


figura 5.7

Las curvas de los potenciales mostrados en la figura 5.7 exhiben un razonable acuerdo cualitativo con lo obtenido por otros autores para sistemas formados por iones disueltos en agua [5, 6]. Algunas de las diferencias con las curvas encontradas por estos autores, se deben a que los valores de carga efectiva que nosotros asignamos a los residuos ajustando el número de coordinación, son un orden de magnitud más pequeño que el utilizado en esos cálculos (*Apéndice C*).

Si bien no hay hasta el momento un total entendimiento de las interacciones mediadas por agua entre iones disueltos [6], las gráficas de la figura 5.7 muestran un comportamiento con características aceptables. Aunque no resulte muy notorio

(observar que las cargas efectivas son menores a  $0.2e$ , y que suponemos una solución infinitamente diluida), las curvas asociadas a residuos con carga negativa (b), exhiben más estructura que las correspondientes a residuos con carga neta positiva (a).

Este efecto estaría relacionado con la distribución de cargas presente en la molécula de agua. Puesto que la molécula de agua posee dos sitios positivamente cargados (los hidrógenos) ubicados en posiciones casi opuestas al único sitio con carga negativa (el oxígeno), es posible que se encuentre formando un *punte* entre dos residuos de carga negativa, algo que no puede suceder entre dos residuos con cargas positivas.

Respecto de las figuras (c) y (d) sólo cabe señalar que fue necesario graficarlas por separado para que pueda apreciarse a cada una de ellas, ya que prácticamente se superponen en un amplio rango de distancias de separación.

### 5.1.3.3 Aminoácidos pertenecientes a grupos diferentes.

Para finalizar este apartado, mostramos algunos potenciales efectivos de interacción entre pares de residuos de aminoácidos pertenecientes a grupos con polaridades diferentes. Para esto seleccionamos tres residuos. Uno apolar (A), uno polar (S) y uno cargado (E).

Por supuesto no mostraremos a los 116 potenciales efectivos de pares que corresponderían a este caso. Ni siquiera podemos afirmar que estos que hemos elegido sean representativos del conjunto. Sólo lo hacemos a modo de ejemplo, y para mostrar que la teoría presentada permite calcular a *todos* los potenciales efectivos de pares.

En la gráfica de la figura 5.8 incluimos tres curvas correspondientes a las interacciones entre residuos de tipo apolar–cargado (AE), apolar–polar (AS) y cargado–polar (ES).

Es interesante notar que la curva correspondiente al potencial de pares alanina–serina es muy similar a la obtenida para alanina–alanina (figura 4.7). Teniendo en cuenta que interactúan en forma directa según un potencial del tipo LJ y la similitud que existe entre los valores de los parámetros  $\sigma_p$  y  $\epsilon_p$ , podemos

decir que en este caso, el efecto del agua no introduce diferencias sustanciales en la interacción efectiva apolar–polar (comparar las filas 1 y 3 de la Tabla 5.1).

Sin embargo, la curva correspondiente a la interacción apolar–cargado (AE) presenta un comportamiento cualitativamente semejante a la anterior, pero no cuantitativamente. Esta variación, que excede a la pequeña diferencia entre los valores de  $\sigma_\rho$  y  $\varepsilon_\rho$ , pone en evidencia el efecto del solvente sobre la interacción efectiva, aún cuando el potencial de interacción directa en este caso vuelva a ser del tipo LJ (comparar las filas 1 y 2 de la Tabla 5.1).

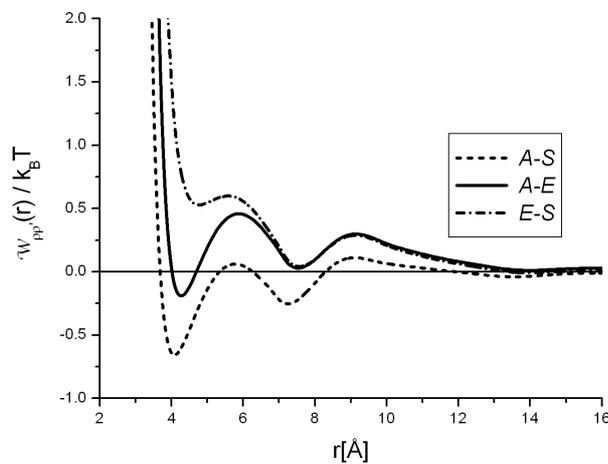


figura 5.8

El potencial efectivo entre un residuo cargado y uno polar (ES) representado por la curva de trazo y punto en la figura 5.8, es un ejemplo del efecto del agua sobre esta interacción (fila 5 de la Tabla 5.1), algo que no habíamos mostrado en ninguna de las gráficas anteriores. Una vez más observamos la presencia de mínimos locales, seguidos por pequeñas barreras y un valor prácticamente nulo para distancias superiores a  $(3 - 4)\sigma_\alpha$ .

## 5.2 POTENCIALES AL CONTACTO.

La hipótesis termodinámica establece que el estado nativo de una proteína es el de más baja energía libre bajo condiciones fisiológicas [7].

Los estudios de plegamiento de proteínas (predicción de estructura, reconocimiento de la conformación nativa, diseño, etc.) generalmente hacen uso

de alguna función de “energía efectiva”. Hay dos tipos diferentes de funciones de “energía efectiva” usadas para éstas aplicaciones. La primera se obtiene de un análisis fundamental de las fuerzas entre las partículas [8]. La segunda, es una función energía derivada de una base de datos de estructuras de proteínas conocidas, mediante algún procedimiento de tipo estadístico [9].

Para muchas aplicaciones, como el reconocimiento de la conformación nativa entre un conjunto de conformaciones posibles, no es necesario que los potenciales estadísticos reproduzcan el verdadero perfil de energías. Simplemente se impone como condición que el estado nativo sea el de más baja energía efectiva en términos relativos, más allá del valor absoluto que se obtenga para la proteína en estudio.

Existen opiniones encontradas acerca del significado físico de estos potenciales estadísticos obtenidos a partir de estructuras de proteínas conocidas [10]. No profundizaremos sobre estas controversias. Más allá de las discusiones, no puede negarse el éxito que muchas de estas “funciones energía” han tenido en la identificación de estructuras nativas [9, 11].

Ahora bien, si estas funciones tuvieran el significado físico que algunos autores les atribuyen, podríamos intentar compararlas con los potenciales efectivos obtenidos mediante la teoría mecánico–estadística desarrollada en esta tesis, y eso es lo que procuraremos hacer a continuación.

### **5.2.1 CORRELACIÓN DE LOS VALORES AL CONTACTO.**

Muchos algoritmos utilizados para reconocimiento de la conformación nativa trabajan sobre un modelo de red, esto es, las partículas que representan a los residuos están unidas formando una cadena y sólo pueden ubicarse en posiciones bien definidas del espacio. Luego, la energía de una dada conformación se calcula sumando las contribuciones entre primeros vecinos no ligados covalentemente. Por esta razón, muchos autores sólo se interesan por los valores de energía al contacto entre residuos.

Por otra parte, los potenciales efectivos que nosotros obtenemos son función de la distancia entre residuos. Por lo tanto, para poder comparar nuestros resultados con los de origen estadístico obtenidos por otros autores, debemos fijar

un criterio para determinar el valor al contacto. Elegimos el valor correspondiente al primer mínimo del potencial efectivo de pares ( $\approx d_{\rho\rho'}$ ), y en aquellos casos donde no hay un mínimo (figura 5.7a y b), consideramos los puntos donde la derivada se hace cero o muy próxima a cero.

Comparamos los valores de energía efectiva al contacto para los 210 pares de residuos con las tablas establecidas por Thomas y Dill (TD), y por Miyazawa y Jeringan (MJ) [11]. Como dijimos más arriba, los potenciales estadísticos son optimizados para cumplir con alguna aplicación como el reconocimiento del plegamiento nativo. Por lo tanto, en lugar de comparar los valores de energía uno a uno en términos absolutos, creemos más apropiado buscar alguna correlación del ordenamiento relativo de los diferentes valores de energía.

Con este fin, consideramos el coeficiente de correlación dependiente de la posición de Spearman ( $r_s$ ) [12]. Este parámetro, presenta varias ventajas con respecto al coeficiente de correlación lineal, el cual se calcula usando los valores propiamente dicho de cada una de las tablas, que desde luego, no coinciden.

El coeficiente de correlación de Spearman compara las posiciones relativas de las energías, reemplazando a cada valor de la tabla por su posición con respecto a los demás.

Llamamos con  $V_i$  a la posición que ocupan los valores de energía  $E_i$  en nuestra tabla (RV), y con  $D_i$  a las posiciones de los valores de energía  $\varepsilon_i$  en una de las otras tablas (MJ o TD). Luego, el coeficiente de correlación se define como [12],

$$r_s = \frac{\sum_i (V_i - \bar{V})(D_i - \bar{D})}{\sqrt{\sum_i (V_i - \bar{V})^2} \sqrt{\sum_i (D_i - \bar{D})^2}} \quad (5.4)$$

Este coeficiente de correlación resulta ser más robusto que el coeficiente de correlación lineal, y menos dependiente de “anomalías” en los datos.

En la Tabla 5.2, resumimos los resultados de la comparación entre energías efectivas determinadas por métodos estadísticos [11] y los valores al contacto que obtenemos con nuestro modelo.

Tabla 5.2: Resultados de la comparación de energías efectivas.

	RV vs. MJ	RV vs. TD
$r_s$	0.484	0.349
NS	$1.05 \times 10^{-13}$	$2.07 \times 10^{-7}$

Para calcular el parámetro  $r_s$  y nivel de significancia (NS) de su desviación de cero, utilizamos la rutina “*spear.f*” en fortran de *Numerical Recipes* [12]. De acuerdo con la teoría  $r_s \in [-1,1]$ . Si el valor obtenido es igual a 1 (-1), se dice que hay una *correlación (anticorrelación)* completa. Un valor del parámetro igual a cero significa que *no existe* ninguna correlación entre las tablas.

Cuando el valor de  $r_s$  difiere de los anteriores, debemos recurrir a la información que nos brinda el nivel de significancia. En un caso como este (Tabla 5.2), un valor muy pequeño de NS ( $\ll 0.01$ ) implica que existe una correlación ( $r_s > 0$ ) significativa entre los valores de energía de las dos tablas comparadas.

### 5.3 REFERENCIAS DEL CAPÍTULO.

- 1 W. Kauzmann, *Adv. Protein Chem.*, **14**, 1 (1959).
- 2 D. Laría and F. Vericat, *Phys. Rev. A*, **43**, 1932 (1991).
- 3 C. Branden and J. Tooze, *Introduction to Protein Structure* (Garland Publishing, Inc., New York, 1991).
- 4 T.E. Creighton, *Proteins. Structures and Molecular Properties* (W.H. Freeman and Company, New York, 1993).
- 5 E. Guàrdia, R. Rey and J.A. Padró, *Chem. Phys.*, **155**, 187-195 (1991).
- 6 N. Bagatella-Flores and P. González-Mozuelos, *J. Chem. Phys.*, **117**, 6133-6141 (2002).
- 7 T.E. Creighton, editor, *Protein Folding* (W.H. Freeman and Company, New York, 1992).
- 8 T. Lazaridis and M. Karplus, *Science*, **278**, 1928-1931 (1997).
- 9 G.M. Crippen, *Biochemistry*, **30**, 4232-4237 (1991). V.N. Maiorov and G.M. Crippen, *J. Mol. Biol.*, **227**, 876-888 (1992). M.J. Sippl, *J. Mol. Biol.*, **213**, 859-883

(1990). M. Hendlich, P. Lackner, S. Weitckus, H. Floeckner, R. Froschauer, K. Gottsbacher, G. Casari and M.J. Sippl, *J. Mol. Biol.*, **216**, 167-180 (1990). S. Miyazawa and R.L. Jernigan, *Macromolecules*, **18**, 534-552 (1985).

10 P.D. Thomas and K.A. Dill, *J. Mol. Biol.*, **257**, 457-469 (1996). A. Ben-Naim, *J. Chem. Phys.*, **107**, 3698-3706 (1997).

11 S. Miyazawa and R.L. Jernigan, *J. Mol. Biol.*, **256**, 623-644 (1996). P.D. Thomas and K.A. Dill, *Proc. Natl. Acad. Sci. USA*, **93**, 11628-11633 (1996).

12 W.H. Press, S.A. Teukolsky, W.T. Vetterling and B.P. Flannery, *Numerical Recipes in Fortran 77. The Art of Scientific Computing, Second Edition, Vol 1* (Syndicate of the University of Cambridge, New York, 1997).

## 6. COMENTARIOS FINALES.

Concluimos este trabajo de tesis haciendo una breve reseña de los resultados obtenidos, y comentando acerca de las posibles líneas de investigación que intentaremos desarrollar a partir de aquí.

Como señalamos en la introducción, el problema que motivó este trabajo, y seguramente ocupará nuestro esfuerzo futuro, está relacionado con el proceso de plegamiento de proteínas globulares y la estabilidad de su estado nativo, en condiciones fisiológicas normales.

Por la complejidad y diversidad de factores que intervienen en este problema, decidimos (como primera aproximación) concentrarnos sobre algunos aspectos termodinámicos y estructurales de las unidades monoméricas que forman a una proteína, es decir, de los aminoácidos. Consideramos entonces, un sistema compuesto por un aminoácido disuelto en agua, y lo abordamos con dos técnicas diferentes. Por un lado, realizamos simulaciones mediante Dinámica Molecular, y por otro, aplicamos una teoría mecánico–estadística de líquidos clásicos.

Las simulaciones mediante Dinámica Molecular nos permitieron cumplir con varios objetivos.

En primer lugar, utilizando los métodos de crecimiento lento e integración termodinámica, determinamos la energía libre de transferencia de los aminoácidos apolares. A partir de este cálculo, establecimos una nueva escala de hidrofobicidad. Notemos que hasta el momento no hay una escala que haya sido universalmente aceptada, más aún, observamos una gran dispersión entre los valores publicados. Esto se debe, fundamentalmente, a las condiciones físicas en las que se llevaron a cabo las distintas mediciones. En este sentido, la simulación brinda la posibilidad de definir claramente al sistema, y ejercer un mayor control sobre sus variables, por lo que consideramos que la escala que obtuvimos constituye una muy buena referencia.

Adicionalmente, calculamos el área superficial accesible a la molécula de agua para cada uno de los residuos de aminoácidos apolares, y las graficamos como función de su hidrofobicidad. Ésta grafica, nos provee un doble test de los

resultados. Por un lado, de acuerdo a lo mostrado por Chothia y corroborado después por otros autores, existe una relación lineal entre hidrofobicidad y área superficial accesible. Por el otro, aunque discutible, el valor de la pendiente que dicha recta debe tener. En ambos aspectos, los resultados obtenidos con Dinámica Molecular exhiben un comportamiento aceptable, tanto cualitativa como cuantitativamente.

En segundo lugar, realizamos simulaciones del sistema en equilibrio, y recolectamos sus configuraciones a fin de estudiar aspectos estructurales del solvente alrededor de cada átomo de la molécula de soluto.

Básicamente trabajamos con tres sistemas diferentes: (a) un aminoácido disuelto en agua, (b) un aminoácido disuelto en metanol, y (c) un residuo de aminoácido disuelto en agua. En todos los casos calculamos la función de correlación radial soluto-solvente y, en (a) y (c), el cambio en la distribución del número de puentes de hidrógeno que forman las moléculas de agua en presencia de una molécula de soluto.

Observamos en general, una estructura más rica (con picos y valles más definidos) en las funciones de distribución radial correspondientes a un aminoácido disuelto en metanol, que en agua. Vemos también cómo cambian las funciones de distribución soluto-agua cuando simulamos sólo al residuo, en lugar del aminoácido completo. En particular, encontramos que para alanina la gráfica conserva la posición de los máximos y mínimos, pero se acentúan cuantitativamente sus valores. Finalmente, notamos un buen acuerdo, al menos cualitativo, con la distribución radial de moléculas de agua alrededor de alanina medida experimentalmente.

Respecto del número medio de puentes de hidrógeno que las moléculas de agua forman en presencia o ausencia de una molécula apolar, encontramos un resultado interesante. Cuando consideramos a los aminoácidos apolares disueltos en agua, obtuvimos un aumento en el número de moléculas que formaban 1 y 2 puentes y una disminución de 4 y 5. Repetimos los mismos cálculos en idénticas condiciones para los residuos, y el resultado fue justamente el opuesto. Si bien las variaciones observadas son pequeñas (recordemos que el sistema está constituido por 1 molécula de soluto y 511 de agua), éstas gráficas muestran

claramente cómo los átomos que forman la cadena principal del aminoácido, afectan a la estructura del agua que rodea al soluto.

El resto del trabajo está dedicado al estudio de la hidratación hidrofóbica y de las interacciones que los residuos de aminoácidos presentan en agua. A partir de aquí, las simulaciones son utilizadas sólo como referencia de las correlaciones residuo–agua que determinamos en forma teórica.

Cuando hablamos de la hidratación hidrofóbica de los aminoácidos, concretamente nos referimos al estudio de la distribución de moléculas de agua alrededor de cada residuo apolar. Dentro de este grupo de residuos, el que mejor se ve representado por el modelo es el de alanina. Observando las curvas de las figuras 3.10 y 3.12, podemos ver que existe un muy buen acuerdo entre las funciones de distribución radial obtenidas mediante Dinámica Molecular y teoría.

Ahora bien, considerando el muy buen acuerdo encontrado para alanina, y lo comentado en el apartado 2.1 (figura 2.1) acerca del rol de test de una teoría que las simulaciones desempeñan, surge naturalmente la idea de aplicar esta teoría mecánico–estadística a los demás residuos apolares. Más aún si tenemos en cuenta que los parámetros utilizados para caracterizar a la molécula de agua en la teoría son los del modelo SPC/E (los mismos de la simulación), y que el  $\sigma_A$  tomado de mediciones experimentales casi coincide con el parámetro de Gromos96.

Efectivamente hacemos esto, y en general conseguimos acuerdos aceptables, aunque podemos observar en la figura 3.11b (para fenilalanina) que el modelo muestra una de sus limitaciones más notorias, relacionada con la forma esférica que se supuso para los residuos.

El cálculo del número de coordinación de cada residuo de aminoácido apolar fue realizado al sólo efecto de determinar un valor para el parámetro  $\varepsilon_p$ . Sin embargo, podría utilizarse como información complementaria en la interpretación de otros resultados.

En el capítulo 4, ampliamos la teoría y estudiamos el otro aspecto del efecto hidrofóbico, es decir, la interacción hidrofóbica. En este caso, calculamos las funciones de distribución radial residuo–residuo, y a partir de ella, el potencial de fuerza media de pares. Con este propósito, consideramos primero un sistema de

referencia en el que solamente actúen fuerzas repulsivas, y luego corregimos perturbativamente los resultados incluyendo el efecto de interacciones atractivas. En la figura 4.5, podemos ver el cambio experimentado por la correlación alanina-alanina al introducir la acción de fuerzas atractivas entre las moléculas. Como test de la teoría mecánico-estadística desarrollada, incluimos en la gráfica la curva que surge de la aproximación "EXP" de Andersen, Chandler y Weeks. Más allá de una pequeña diferencia cuantitativa, observamos que ambas teorías entregan funciones que se comportan en forma semejante.

En la figura 4.7, podemos ver la interacción entre dos residuos de alanina que se encuentran en el vacío, y el cambio notable que ocurre en presencia de agua. Tanto para alanina, como para los demás residuos apolares, los potenciales de fuerza media exhiben el mismo comportamiento. Son fuertemente repulsivos a distancias menores que el contacto, presentan dos pozos y dos barreras bien definidas, y se anulan para grandes distancias.

Por último, extendemos la teoría utilizada para los residuos apolares, a los restantes aminoácidos. Para concretar esta idea, necesitamos proponer potenciales de interacción directa entre residuos (cualquiera sea su polaridad) que dependan de la distancia entre las moléculas pero no de sus orientaciones. Para el ajuste de los parámetros de la teoría, nuevamente nos apoyamos en las funciones de distribución radial residuo-agua surgidas de simulación. Los resultados obtenidos y mostrados en el capítulo 5 de esta tesis muestran un comportamiento aceptable, aunque no contamos con resultados experimentales ni de simulación, que nos permitan establecer una comparación más rigurosa.

Precisamente, nuestro objetivo inmediato estará dedicado a examinar con profundidad éstos potenciales efectivos de pares. En este sentido, pensamos utilizarlos en situaciones relacionadas con ciertos problemas de interés biológico, e implementar simulaciones numéricas que nos permitan estudiar propiedades estructurales de proteínas pequeñas o cadenas cortas de aminoácidos.

Además, proyectamos mejorar su determinación desde el punto de vista teórico, considerando las posiciones de los átomos que forman a cada molécula de residuo, y sus interacciones.

# APÉNDICES

## A. RESUMEN DE GROMOS96.

A continuación describimos en forma abreviada, aquellos elementos del paquete de programas Gromos96, que utilizamos en los cálculos presentados en esta tesis.

La intención de este apéndice, es ubicar a dichos cálculos en un contexto más preciso. Desde luego, los detalles sobre los elementos que aquí se exponen pueden encontrarse en las referencias bibliográficas, y en particular en el manual de Gromos96 [1].

### A.1 ECUACIONES DE MOVIMIENTO.

En un esquema clásico, como el usado en Gromos96, el estado de un sistema queda definido por las posiciones y momentos del conjunto de partículas que lo constituyen. Dentro de la aproximación de Born–Oppenheimer, es posible promediar el movimiento rápido de los electrones, y expresar al Hamiltoniano de un sistema como una función de las variables nucleares.

El Hamiltoniano  $H$  de un sistema de  $N$  partículas es una función de las coordenadas y momentos generalizados  $\vec{q}_i$  y  $\vec{p}_i$ :

$$\begin{aligned}\vec{q} &= (\vec{q}_1, \vec{q}_2, \dots, \vec{q}_N) \\ \vec{p} &= (\vec{p}_1, \vec{p}_2, \dots, \vec{p}_N) \\ H(\vec{q}, \vec{p}) &= K(\vec{p}) + V(\vec{q})\end{aligned}\tag{A.1}$$

La energía cinética  $K(\vec{p})$ , toma la forma:

$$K(\vec{p}) = \sum_{i=1}^N \sum_{\alpha} \frac{p_{i\alpha}^2}{2m_i}\tag{A.2}$$

donde  $m_i$  es la masa molecular, y el índice  $\alpha$  varía sobre las componentes (x, y, z) de la partícula  $i$ -ésima.

La energía potencial  $V(\vec{q})$  contiene información sobre las interacciones. Generalmente es  $H(\vec{q}, \vec{p})$  (o  $V(\vec{q})$ ) el que se establece a la entrada de cualquier programa. Una aproximación muy usada en Simulación Molecular (SM), consiste en dividir a la energía potencial en términos que involucran pares, tripletes, etc. de

moléculas. Para un sistema simple que contiene N moléculas, la energía potencial toma la forma siguiente:

$$V(\vec{r}) = \sum_i v_1(\vec{r}_i) + \sum_i \sum_{j>i} v_2(\vec{r}_i, \vec{r}_j) + \sum_i \sum_{j>i} \sum_{k>j>i} v_3(\vec{r}_i, \vec{r}_j, \vec{r}_k) + \dots \quad (\text{A.3})$$

El primer término en la ecuación anterior  $v_1(\vec{r}_i)$ , representa el efecto de un campo externo sobre el sistema. Los restantes términos representan las interacciones entre moléculas.

El segundo término  $v_2(\vec{r}_i, \vec{r}_j)$ , es el potencial de pares y el más importante. Este potencial depende, en general, sólo de la separación entre las moléculas  $r_{ij} = |\vec{r}_i - \vec{r}_j|$ , y puede escribirse  $v_2(r_{ij})$ .

El término  $v_3(\vec{r}_i, \vec{r}_j, \vec{r}_k)$  involucra tripletes de moléculas, y resulta significativo a densidades de líquidos. En general, los términos correspondientes a cuatro y más cuerpos, son sustancialmente menos significativos que  $v_2(r_{ij})$  y  $v_3(\vec{r}_i, \vec{r}_j, \vec{r}_k)$ .

El término de tres cuerpos en el potencial rara vez se incluye en la simulación debido a que los cálculos de cualquier cantidad que involucre una suma sobre tripletes de moléculas consumirían mucho tiempo.

Afortunadamente, la aproximación de pares da una buena descripción de las propiedades de los líquidos, debido a que el efecto promedio de tres cuerpos puede ser parcialmente incluido definiendo un potencial de pares efectivos. Luego, la ecuación (A.3) toma la forma:

$$V(\vec{r}) \approx \sum_i v_1(\vec{r}_i) + \sum_i \sum_{j>i} v_2^{ef}(r_{ij}) \quad (\text{A.4})$$

Normalmente, el potencial que aparece en las simulaciones, es un potencial efectivo de pares de este tipo. Debido a que este potencial efectivo de pares necesita reproducir datos experimentales, podría depender de la densidad, temperatura, etc., mientras que el verdadero potencial de dos cuerpos  $v_2(r_{ij})$  no lo hace.

En Gromos96, se usan las ecuaciones de movimiento de Newton, para los átomos o núcleos:

$$\frac{d\vec{r}_i(t)}{dt} = \vec{v}_i(t) \quad (\text{A.5})$$

y

$$\frac{d\vec{v}_i(t)}{dt} = \frac{\vec{f}_i(t)}{m_i} \quad (\text{A.6})$$

con

$$\vec{f}_i(t) = -\frac{\partial}{\partial \vec{r}_i} V(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N) \quad (\text{A.7})$$

donde las velocidades atómicas se indican por  $\vec{v}_i$ , las fuerzas sobre los átomos por  $\vec{f}_i$ , y el tiempo por  $t$ . En estas ecuaciones de movimiento,  $\vec{r}_i$  son las coordenadas cartesianas de la partícula  $i$  con masa  $m_i$ .

Debería incluirse en el modelo, la descripción no sólo de la fuerza:

$$\vec{f}_i^{int}(t) = -\frac{\partial V^{int}}{\partial \vec{r}_i} \quad (\text{A.8})$$

debido a las partículas explícitamente simuladas, sino también a las fuerzas externas  $\vec{f}_i^{ext}$  provenientes del medio. Esas fuerzas externas pueden ser representadas en forma aproximada por su efecto promedio  $\vec{f}_i^{media}$ . Para la correspondiente función de energía potencial de fuerza media  $V^{media}$ , se tiene:

$$\vec{f}_i^{media}(t) = -\frac{\partial V^{media}}{\partial \vec{r}_i} \quad (\text{A.9})$$

la cual representa el efecto promedio de los grados de libertad que no son explícitamente tratados en la simulación. En la práctica, los parámetros y la forma de  $V^{int}$  son convenientemente elegidos para dar cuenta de los efectos del medio en forma promediada.

## A.2 ALGORITMOS PARA LA RESOLUCIÓN DE LAS ECUACIONES DE MOVIMIENTO.

La parte que más tiempo consume en casi toda simulación con Dinámica Molecular (DM), es el cálculo de las fuerzas que actúan sobre cada una de las partículas.

Si consideramos un modelo de sistema con interacción de pares, debemos calcular la fuerza resultante que actúa sobre la partícula  $i$  debido a todas las demás partículas vecinas.

Entonces, para un sistema de  $N$  partículas, necesitamos evaluar  $N(N-1)/2$  distancias, lo cual implica que el tiempo necesario para el cálculo de las fuerzas depende del número de partículas como  $N^2$ . Para acelerar la evaluación de las fuerzas, se han desarrollado algunas técnicas muy eficientes que hacen que este cálculo sea función de  $N$  en lugar de  $N^2$ .

En el algoritmo de Verlet [2] y sus modificaciones, las coordenadas moleculares se actualizan a partir de cantidades que se calculan en el mismo paso o en el paso previo. En lo que sigue nos concentramos en uno de los algoritmos de Verlet, denominado “salto de rana”, pues éste es el algoritmo que usa Gromos96.

### A.2.1 ALGORITMO SALTO DE RANA.

En 1967, Verlet demostró que un método simple desarrollado por Störmer, podía usarse con éxito en DM. Luego surgieron muchas modificaciones y mejoras que dieron lugar a una familia de algoritmos de Verlet, que los convirtieron en los más utilizados en simulaciones mediante DM.

En 1970 Hockney [3] propuso uno, denominado *salto de rana* (*leap-frog*) que hace uso de la velocidad a intervalos de tiempos medios:

$$\begin{aligned}\vec{v}(t + \delta t/2) &= \vec{v}(t - \delta t/2) + \vec{a}(t) \delta t \\ \vec{r}(t + \delta t) &= \vec{r}(t) + \vec{v}(t + \delta t/2) \delta t \\ \vec{v}(t) &= [\vec{v}(t + \delta t/2) + \vec{v}(t - \delta t/2)]/2\end{aligned}\tag{A.10}$$

La implementación del algoritmo de salto de rana sigue el mismo procedimiento general que el de Verlet, sus trayectorias coinciden, pero el cálculo de velocidades ahora es una parte integral del algoritmo. Las aceleraciones calculadas se usan para determinar las velocidades en el tiempo medio, y éstas a su vez, se usan para actualizar las posiciones y velocidades.

En este algoritmo, en cada paso temporal debemos almacenar los valores actuales de las posiciones y las velocidades en el tiempo medio anterior. Esos

valores se actualizan a partir del cambio de cada una de las coordenadas. Este proceso se esquematiza en la figura A.1.

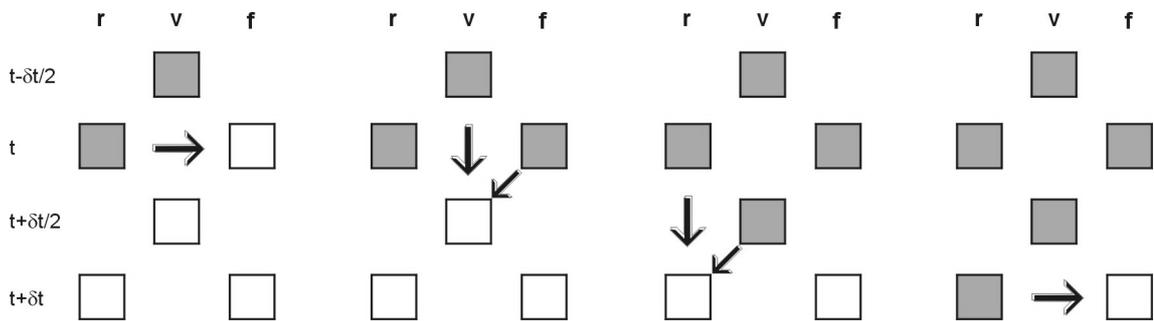


figura A.1

La principal ventaja del algoritmo de salto de rana, es la reducción de la imprecisión numérica por el uso de diferencias entre cantidades más pequeñas. Una pequeña dificultad del algoritmo, es que inicialmente necesitamos conocer la velocidad en el tiempo  $t = -\delta t/2$ , aunque esto se resuelve definiendo:

$$\vec{v}(-\delta t/2) = \vec{v}(0) - \vec{a}(0)(\delta t/2) \quad (\text{A.11})$$

### A.3 CAMPOS DE FUERZAS.

El cálculo del promedio termodinámico de cualquier propiedad, requiere determinar las energías potencial y cinética del sistema. En simulación, la energía potencial del sistema o fuerzas que actúan entre moléculas se calcula comúnmente a partir del potencial intermolecular aditivo de a pares.

El desarrollo de potenciales intermoleculares tiene una larga historia que antecede al advenimiento de los métodos de simulación molecular. Sin embargo, excepto para los casos más simples, el valor predictivo de los diferentes potenciales intermoleculares sólo puede ser evaluado rigurosamente a través de una simulación molecular. A pesar de esto, muy pocos potenciales fueron extensivamente chequeados usando simulación molecular, a excepción de los potenciales de esferas duras (ED) y Lennard-Jones (LJ).

Si bien existen muchos potenciales de pares del tipo “repulsión + dispersión”, en simulación molecular se usan relativamente pocos potenciales diferentes. Además de los potenciales de “esferas duras + término atractivo”, una

representación más apropiada de la interacción intermolecular está dada por el potencial de LJ. Este potencial es de la forma:

$$V_2(r_{ij}) = \varepsilon \left( \left( \frac{m}{n-m} \right) x_{ij}^{-n} - \left( \frac{n}{n-m} \right) x_{ij}^{-m} \right) \quad (\text{A.12})$$

donde  $n$  y  $m$  son constantes,  $x_{ij} = r_{ij}/r_m$  y  $r_m$  es la correspondiente distancia de separación para el mínimo de energía. El diámetro está relacionado con la separación de mínima energía por:

$$\sigma = r_m \left( \frac{m}{n} \right)^{1/n-m} \quad (\text{A.13})$$

La forma más común del potencial de LJ se obtiene cuando  $n=12$  y  $m=6$ :

$$V_2(r_{ij}) = 4\varepsilon \left( \left( \frac{\sigma}{r_{ij}} \right)^{12} - \left( \frac{\sigma}{r_{ij}} \right)^6 \right) \quad (\text{A.14})$$

El potencial de LJ es quizás el más usado en simulaciones moleculares. La virtud más importante de este potencial es que combina una descripción bastante realista de las interacciones moleculares con simplicidad computacional.

Construir un potencial de pares para una molécula es bastante más complicado que un potencial atómico. Para una molécula dada, podemos distinguir entre contribuciones al potencial provenientes de interacciones intra e intermoleculares.

Las interacciones intermoleculares ocurren entre los diferentes átomos de moléculas diferentes. Una molécula experimenta los efectos de interacciones intramoleculares, como consecuencia de las interacciones entre los átomos que la forman, y se las clasifican a su vez como ligadas o no ligadas.

Las interacciones ligadas están restringidas a átomos que están unidos covalentemente en forma directa (interacciones1-2), átomos que están unidos a un átomo en común (interacciones1-3), y átomos con una topología que involucra tres uniones intermedias (interacciones1-4).

Las interacciones no ligadas en general, pueden ser de largo o de corto alcance. Los efectos de inducción, dispersión y electrostáticos representan algunos ejemplos de interacciones de largo alcance. En estos casos, la energía

de interacción es proporcional a alguna potencia de la inversa de la distancia entre átomos.

La interacción electrostática depende de la distribución de cargas en las moléculas, y puede ser tanto atractiva como repulsiva. El efecto de inducción es siempre atractivo y resulta de la distorsión causada por los campos de átomos vecinos. Sin embargo, la contribución más importante es la influencia atractiva de la dispersión proveniente de las fluctuaciones instantáneas, causadas por el movimiento de los electrones.

Las interacciones de corto alcance están caracterizadas por un decaimiento exponencial en la energía de interacción con respecto a la separación intermolecular. A pequeñas separaciones, hay un solapamiento significativo de la función de onda atómica responsable de la repulsión.

El campo de fuerzas o función interacción Gromos96, tiene la forma siguiente:

$$V(\vec{r}) = V^{lig}(\vec{r}) + V^{no\,lig}(\vec{r}) \quad (A.15)$$

con

$$V^{lig}(\vec{r}) = V^{coval}(\vec{r}) + V^{áng}(\vec{r}) + V^{died}(\vec{r}) + V^{tors}(\vec{r}) \quad (A.16)$$

El término en la función de interacción que representa a las interacciones no ligadas, se escribe como

$$V^{no\,lig} = \sum_{\substack{\text{pares no} \\ \text{ligados } (i,j)}} \left\{ V^{LJ}(r_{ij}, C_{12}(i,j), C_6(i,j), R_{cp}, R_{cl}) + V^{CCR}(r_{ij}, q_i, q_j, R_{cp}, R_{cl}, R_{cr}, \epsilon_1, \epsilon_2, \kappa) \right\} \quad (A.17)$$

con

$$V^{LJ}(r_{ij}) = \left( \frac{C_{12}(i,j)}{r_{ij}^6} - C_6(i,j) \right) \frac{1}{r_{ij}^6} \quad (A.18)$$

y

$$V^{CCR}(r_{ij}) = \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1} \left( \frac{1}{r_{ij}} - \frac{C_{cr} r_{ij}^2}{2R_{cr}^3} - \frac{1 - C_{cr}/2}{R_{cr}} \right) \quad (A.19)$$

La sumatoria corre sobre todos los pares de átomos  $i$  y  $j$ .

Sin embargo, hay pares de átomos como los vecinos excluidos, y los no incluidos en la lista de grupos de carga (radio de corte  $R_{cp}$ ) o en interacciones de

largo alcance no ligadas (radio de corte  $R_{cr}$ ), que no son considerados en la sumatoria.

Los átomos  $i$  y  $j$  que están covalentemente unidos, son llamados átomos primeros vecinos y si ellos están covalentemente unidos a un átomo vecino en común, son llamados segundos vecinos.

Debido a la corta distancia  $r_{ij}$  entre primeros o segundos vecinos, la interacción no ligada (A.17) entre estos átomos será muy grande. Entonces, los primeros y segundos vecinos son excluidos de la sumatoria en la fórmula (A.17). Terceros vecinos son excluidos sólo en casos especiales entre átomos que pertenecen o están unidos a anillos aromáticos.

La interacción Coulombiana en (A.17) es de largo alcance, ya que depende inversamente de la distancia  $r_{ij}$ . El tamaño de la interacción Coulombiana a una distancia de corte típica de  $(8-10)\text{Å}$  no es despreciable, aún para cargas atómicas parciales de  $(0.1-0.4)e$ . Esto significa que la interacción Coulombiana sólo puede ser despreciada o aproximada usando un *campo de reacción* continuo, más allá de distancias largas ( $R_{cr} \approx 20-30\text{Å}$ ).

Puesto que el número de átomos vecinos no ligados crece con la distancia proporcionalmente a la tercera potencia del radio de corte, la aplicación de un radio de corte grande sobre las interacciones no ligadas aumenta dramáticamente el tiempo de cálculo requerido.

Esta situación puede aliviarse definiendo *grupos de carga* que convierten a las interacciones electrostáticas en dipolares ( $1/r_{ij}^3$ ). Cuando las cargas atómicas parciales de un grupo de átomos suman exactamente cero, el primer término de la interacción electrostática entre dos grupos de átomos es de carácter dipolar ( $1/r_{ij}^3$ ). La suma de las contribuciones de monopolos de los distintos pares de átomos (a la interacción *grupo-grupo*) será igual a cero.

El *campo de fuerzas* de Gromos96, hace uso de este concepto de grupos de carga. Siempre que sea posible, los átomos que pertenecen a un grupo de carga se eligen de manera que su carga atómica parcial sume cero. En ciertos casos (residuos cargados como los de *Arg* o *Asp*), las cargas atómicas parciales de los

grupos de carga suman  $+e$  o  $-e$ . En las subrutinas de Gromos96, las interacciones no ligadas sólo se calculan entre grupos de carga.

Cuando se usa un radio de corte, debe definirse la distancia entre grupos de carga. La posición de un grupo de carga, se define de modo diferente para grupos de carga que pertenecen al “*soluto*” de las que pertenecen al “*solvente*”. La posición de un grupo de carga para el “*soluto*” se define por su centro geométrico:

$$\vec{R}_{gc} = \sum_{i=1}^{N_{gc}} \frac{\vec{r}_i}{N_{gc}} \quad (\text{A.20})$$

donde  $N_{gc}$  denota al número de átomos que pertenecen al grupo de carga.

La posición de un grupo de carga para el “*solvente*” se toma como la posición del primer átomo de la molécula de solvente. En Gromos96, una molécula de solvente contiene siempre un sólo grupo de carga.

Para evaluar las interacciones no ligadas con suficiente precisión, se usa un radio de corte “*largo*”  $R_{cl}$ . Un valor razonable para este parámetro sería de 15 Å o mayor. Pero resulta que un rango como este es computacionalmente muy costoso si se evalúan todas las interacciones de pares. El número de átomos vecinos dentro de un radio de 15 Å será mayor a 300. Por esta razón, las interacciones se evalúan usando un método de doble rango, y las interacciones electrostáticas más allá de  $R_{cl}$ , se aproximan por medio de un *campo de reacción*.

La interacción no ligada en (A.17) se evalúa en cada paso de simulación usando una lista de pares de grupos de carga que se genera con un radio de corte “*pequeño*”  $R_{cp}$ .

Las interacciones no ligadas de largo alcance, esto es, aquellas entre grupos de carga a una distancia mayor que  $R_{cp}$  y menores que  $R_{cl}$ , se evalúan menos frecuentemente, sólo cada  $n$  pasos de simulación, momento en el cual también se actualiza la lista de pares. Esta lista se mantiene inalterada entre esas actualizaciones. De esta manera, las fuerzas no ligadas de largo alcance pueden ser tenidas en cuenta de un modo aproximado, sin aumentar significativamente el esfuerzo de cálculo.

Las interacciones de van der Waals en la fórmula (A.18) y las interacciones electrostáticas entre cargas de igual signo en la fórmula (A.19), se hacen

infinitamente grandes cuando  $r_{ij}$  tiende a cero. Si se usa una interacción llamada de “carozo blando” de la forma:

$$V^{LJ} = \left( \frac{C_{12}(i, j)}{\alpha_{LJ}(i, j)(\lambda^2 + \mu^2)C_{126}(i, j) + r_{ij}^6} - C_6(i, j) \right) \frac{1}{\alpha_{LJ}(i, j)(\lambda^2 + \mu^2)C_{126}(i, j) + r_{ij}^6} \quad (\text{A.21})$$

se evita la singularidad en  $r_{ij} = 0$ . En Gromos96 esto se usa cuando se trabaja en un cálculo de energía libre, algo que nosotros utilizamos en el capítulo 3 para determinar una escala de hidrofobicidad de residuos de aminoácidos apolares.

Los parámetros de la interacción no ligada de van der Waals  $C_{12}(i, j)$  y  $C_6(i, j)$  en las fórmulas (A.18) y (A.21) dependen del tipo de átomos  $i$  y  $j$ . Básicamente, los parámetros Gromos96 de van der Waals para un par de átomos con códigos enteros  $i$  y  $j$ , son derivados de los parámetros de átomo simple usando las relaciones:

$$\begin{aligned} C_6(i, j) &= C_6^{1/2}(i, i) \cdot C_6^{1/2}(j, j) \\ C_{12}(i, j) &= C_{12}^{1/2}(i, i) \cdot C_{12}^{1/2}(j, j) \\ C_{126}(i, j) &= C_{12}(i, j) / C_6(i, j) \end{aligned} \quad (\text{A.22})$$

El campo de fuerzas Gromos96 no contiene términos especiales en la función de interacción (A.15) que imiten a los puentes de hidrógeno. La capacidad para formar puentes de hidrógeno de las moléculas, es el resultado de un balance entre las interacciones repulsivas y atractivas.

#### A.4 CONDICIONES PERIÓDICAS DE CONTORNO.

Cuando simulamos a un sistema de tamaño finito, debemos decidir cómo tratar a los bordes. Para un líquido o una solución, usualmente minimizamos los efectos de borde o superficie aplicando condiciones periódicas de contorno.

Pese al enorme desarrollo de los equipos de computación, el número de partículas que podemos manejar es pequeño respecto de cualquier sistema real. En una muestra macroscópica, cuyo número de partículas es del orden de  $10^{24}$ , la mayoría de las moléculas están en el “interior” de la muestra. Sin embargo, no ocurre lo mismo en una muestra de tamaño reducido.

Habitualmente, el número de partículas de los sistemas más grandes con los que trabajamos en simulación, es de miles o algunas decenas de miles de ellas. El tamaño del sistema estará limitado por la cantidad de memoria para almacenar datos y, más importante aún, por la velocidad de ejecución del programa.

Lo que hacemos entonces, es poner a los átomos del sistema en una caja cúbica, o en cualquier caja que llene periódicamente el espacio, y la rodeamos con imágenes idénticas de sí misma.

Durante la simulación, cuando las moléculas se mueven en la caja original, sus imágenes periódicas en cada una de las cajas vecinas lo hacen exactamente de la misma manera. Cuando un átomo abandona la caja central por un lado, ingresa a ella con idéntica velocidad por el lado opuesto. No hay paredes en el borde de la caja central, y tampoco moléculas sobre la superficie.

Esta caja simplemente forma un sistema de ejes convenientes para medir las coordenadas de las  $N$  moléculas. Cuando una partícula atraviesa algún borde, sus imágenes lo hacen también a través de los correspondientes bordes, de manera que el número de partículas en la caja central se conserve.

Es importante preguntarse si las propiedades de un sistema periódico infinitamente pequeño, y el sistema macroscópico al cual representa, coinciden. En general, esto dependerá del rango del potencial de interacción intermolecular y del fenómeno bajo estudio.

La experiencia en simulación dice que las condiciones periódicas de contorno tienen un efecto pequeño sobre las propiedades termodinámicas en equilibrio y estructuras de fluidos en la región lejana a una transición de fase, mientras las interacciones sean de corto alcance. Posibles distorsiones por efecto de las condiciones periódicas de contorno se pueden detectar simulando sistemas de tamaños diferente.

El corazón de un programa de DM involucra el cálculo de la energía potencial de una configuración particular, y las fuerzas que actúan sobre las moléculas. Calcular la fuerza resultante sobre una molécula y sus contribuciones a la energía potencial, supone aditividad de pares e incluir todas las interacciones entre ésta y las demás moléculas en la caja de simulación.

En principio, tenemos  $N-1$  términos en la suma. Consideramos a la molécula 1 en el centro de una región que tiene el mismo tamaño y forma que la caja básica de simulación. La molécula 1 interactúa con todas las moléculas cuyo centro cae dentro de esta región, esto es, con las imágenes más cercanas de las otras  $N-1$  moléculas. Esta es la llamada “convención de imagen mínima”.

En la convención de imagen mínima tenemos  $N(N-1)/2$  términos en el cálculo de la energía potencial, debido a la aditividad de pares de las interacciones. Este es un cálculo muy pesado para sistemas de miles de partículas.

Dado que la mayor contribución al potencial y a las fuerzas, proviene de los vecinos cercanos a la molécula de interés, para fuerzas de corto alcance normalmente se aplica un radio de corte esférico. Esto sugiere una distancia de corte  $R_c$ , de modo tal que el potencial de pares  $V(r)$  se anule para  $r > R_c$ .

## A.5 ACOPLAMIENTO A BAÑOS TÉRMICOS Y DE PRESIÓN.

Existen varios métodos para hacer una DM a temperatura constante. Uno muy simple es el llamado acoplamiento débil, en el cual las ecuaciones de movimiento atómicas son modificadas de manera que el resultado neto sobre el sistema sea una relajación de la temperatura hacia el valor  $T_0$  prefijado:

$$\frac{dT(t)}{dt} = \frac{T_0 - T(t)}{\tau_T} \quad (\text{A.23})$$

Considerando un parámetro  $\lambda$  como factor de escala para las velocidades de los átomos en el sistema, obtenemos la expresión:

$$\lambda(t) = \left( 1 + \frac{\delta t}{\tau_T} \left( \frac{T_0}{T(t)} - 1 \right) \right)^{1/2} \quad (\text{A.24})$$

donde el cambio de escala de las velocidades  $\vec{v}_i(t)$  por el factor  $\lambda(t)$  en cada paso de DM, hace que la temperatura  $T(t)$  relaje a  $T_0$ .

El valor de  $\tau_T$  debe elegirse lo bastante chico (acoplamiento fuerte) como para alcanzar la temperatura requerida, pero suficientemente grande (acoplamiento débil) como para evitar disturbios en las propiedades del sistema

por el acoplamiento al baño de temperatura. Un valor razonable para  $\tau_T$  es de aproximadamente 0.1 ps.

Para llevar a cabo una DM a presión constante, se procede en forma similar y el resultado neto sobre el sistema es una relajación de la presión  $P(t)$  hacia el valor de referencia  $P_0$ :

$$\frac{dP(t)}{dt} = \frac{P_0 - P(t)}{\tau_p} \quad (\text{A.25})$$

Multiplicando el volumen y las coordenadas atómicas por un factor de escala  $\mu$ , y resolviendo para el caso isotrópico, obtenemos:

$$\mu(t) = \left( 1 - \frac{\delta t}{\tau_p} (P_0 - P(t)) \right)^{\frac{1}{3}} \quad (\text{A.26})$$

donde el cambio de escala de las coordenadas atómicas  $\vec{r}_i(t)$  y los lados de la caja, conjuntamente con el factor  $\mu(t)$  en cada uno de los pasos de DM, harán que la presión  $P(t)$  relaje hacia el valor  $P_0$ .

La velocidad de relajación está controlada por el tiempo de relajación  $\tau_p$  elegido. Nuevamente, el valor de  $\tau_p$  debe elegirse suficientemente pequeño (acoplamiento fuerte) como para alcanzar la presión promedio deseada, pero suficientemente grande (acoplamiento débil) como para evitar perturbaciones de las propiedades del sistema por el acoplamiento al baño de presión.

Puesto que la definición de la presión depende de la energía cinética, el acoplamiento de presión no debe ser más fuerte que el acoplamiento de temperatura:

$$\tau_p \geq \tau_T \quad (\text{A.27})$$

Valores típicos para  $\tau_p$  son (0.4–0.5) ps.

## A.6 CONFIGURACIÓN Y VELOCIDADES INICIALES.

Una simulación mediante DM comienza con posiciones y velocidades atómicas iniciales. Aunque el resultado debe ser independiente de esto, es

conveniente elegir a la configuración inicial tan representativa como sea posible del sistema molecular en equilibrio.

Para evitar conversiones de la energía de deformación en energía cinética, quitamos las distorsiones en la configuración inicial a través de una corrida corta de minimización de la energía (100 pasos).

Si las velocidades no están disponibles, entonces las tomamos a partir de una distribución de Maxwell de la temperatura:

$$P(v_j) = \frac{\exp(-m_j v_j^2 / (2k_B T_i))}{(2\pi k_B T_i / m_j)^{3/2}} \quad (\text{A.28})$$

donde  $P(v_j)$  es la probabilidad de ocurrencia de la velocidad  $v_j$  para el átomo  $j$ .

Para comenzar la simulación de manera suave, tomamos las velocidades iniciales de una distribución de Maxwell a baja temperatura (50°K) y luego vamos aumentando la temperatura de referencia después de unos cortos períodos de simulación (unos pocos ps) hasta alcanzar la temperatura de referencia, que en nuestros cálculos es 300°K.

Al comienzo de una simulación es conveniente hacer fuerte el acoplamiento a los baños de temperatura y presión para permitir una transferencia rápida del calor y presión en exceso.

Debemos tener en cuenta además, que el período de equilibración de una simulación deberá ser más largo que los tiempos de relajación del sistema, que a su vez, dependen de los tiempos de relajación de las propiedades de interés.

## A.7 CÁLCULO DE ENERGÍAS LIBRES.

En el estudio de la *hidratación hidrofóbica* que presentamos en el capítulo 3, calculamos las variaciones de energía libre (hidrofobicidad) que se producen al transferir un residuo apolar desde un solvente polar (agua) a otro no polar (metanol).

La variación de energía libre entre dos estados de un sistema molecular ( $A$  y  $B$ ) definido sobre un intervalo  $[\lambda_A = 0, \lambda_B = 1]$  sigue simplemente de la definición del cálculo integral elemental:

$$\Delta G = \int_0^1 \frac{\partial G(\lambda)}{\partial \lambda} d\lambda \quad (\text{A.29})$$

Podemos expresar esta integral como un promedio sobre el ensamble:

$$\Delta G = \int_0^1 \left\langle \frac{\partial E(X^N; \lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda \quad (\text{A.30})$$

donde  $E(X^N)$  representa la energía configuracional para un sistema de  $N$  partículas. La evaluación numérica directa de la ecuación anterior en una simulación, integrando entre el estado inicial y final usando alguna relación termodinámica, se llama integración termodinámica [4, 5].

En el caso especial, pero comúnmente encontrado en que  $E(X^N; \lambda)$  es lineal en  $\lambda$ :

$$\begin{aligned} E(X^N; \lambda) &= (1-\lambda)E_0(X^N) + \lambda E_1(X^N) \\ E(X^N; \lambda) &= E_0(X^N) + \lambda \Delta E(X^N) \end{aligned} \quad (\text{A.31})$$

donde

$$\Delta E(X^N) = E_1(X^N) - E_0(X^N) \quad (\text{A.32})$$

entonces (A.30) toma la forma:

$$\Delta G = \int_0^1 \left\langle \Delta E(X^N) \right\rangle_{\lambda} d\lambda \quad (\text{A.33})$$

La metodología para llevar a cabo la integración termodinámica en una simulación molecular es muy directa. Primero realizamos una serie de simulaciones correspondientes a una sucesión de valores de  $\lambda$  en el intervalo  $[0,1]$ , y obtenemos valores discretos para los distintos  $\left\langle \Delta E(X^N) \right\rangle_{\lambda}$ . Luego, realizamos la integración auxiliar sobre  $\lambda$  en forma numérica.

En casos favorables, podemos acelerar el proceso permitiendo a  $\lambda$  variar continuamente en la simulación a través de un proceso introducido por Berendsen y colaboradores llamado crecimiento lento [4, 5]. Sin embargo, estrictamente hablando, durante la simulación el sistema nunca se encuentra en equilibrio, a menos que los cambios en  $\lambda$  sean infinitesimalmente pequeños.

Además, los cambios estructurales que ocurren en el sistema al ir desde el estado inicial al final deben ser alcanzados en la escala de tiempos de

picosegundos. Grandes cambios conformacionales, no tendrían tiempo de suceder y podrían, en casos desfavorables, ocasionar problemas con un crecimiento lento en contraposición a un método discreto donde se realiza una equilibración en cada paso.

## **A.8 CICLO TERMODINÁMICO.**

Para muchos cálculos de energía libre conviene hacer uso de un ciclo termodinámico [5, 6]. Puesto que la energía libre es una función de estados, un cambio en la energía libre será independiente del camino que conecte los estados inicial y final, mientras que el sistema esté en equilibrio y sufra cambios reversibles.

Como a lo largo de un camino cerrado o ciclo, la variación de energía libre es igual a cero, para calcularla podemos elegir entre muchas opciones posibles. Una manera de realizar este cálculo en forma directa, es seguir un camino que conecte dos estados de interés *A* y *B*. Alternativamente, podemos diseñar un ciclo en el cual el camino de *A* a *B* sea sólo una parte, y determinamos su valor calculando la variación de energía libre de los restantes caminos que forman el ciclo. El poder de ésta técnica reside en que podemos definir a los caminos de modo de optimizar la precisión de los cálculos, aún cuando correspondan a procesos que no son válidos para la química.

Elegimos a los estados inicial y final, y al camino que los conecta, de modo tal que el tiempo de relajación del sistema con respecto al cambio en el Hamiltoniano, y el tiempo requerido para muestrear los ensambles, sean ambos mínimos. Esto implica que el camino más directo o químicamente válido no necesariamente será el más eficiente.

En nuestros cálculos con Gromos96, evitamos introducir o remover grados de libertad. Generalmente, cuando cambiamos un átomo real por uno “fantasma” o viceversa, modificamos sólo las interacciones no ligadas. Los términos como la longitud de unión, ángulos de unión y diedros permanecen inalterados.

## A.9 REFERENCIAS.

- 1 W.F. van Gunsteren, S.R. Billeter, A.A. Eising, W.R.P. Scott, and I.G. Tironi, *Biomolecular Simulation: The Gromos96 Manual and User Guide* (Zürich: vdf Hochschulverlag AG an der ETH Zürich, 1996).
- 2 L. Verlet, *Phys. Rev.*, **159**, 98 (1967).
- 3 R.W. Hockney, *Methods Comput. Phys.*, **9**, 136(1970).
- 4 W.F. van Gunsteren and P.K. Weiner, *Computer Simulation of Biomolecular Systems. Theoretical and Experimental Applications* (ESCOM Science Publishers B.V., 1989).
- 5 X. Daura, P.H. Hünenberger, A.E. Mark, E. Querol, F.X. Avilés, and W.F. van Gunsteren, *J. Am. Chem. Soc.*, **118**, 6285 (1996).
- 6 P.A. Bash, U.C. Singh, R. Langridge and P.A. Kollman, *Science*, **236**, 564 (1987).

## B. VALORES DE LOS PARÁMETROS USADOS EN LAS SIMULACIONES CON GROMOS96.

Todas las simulaciones de Dinámica Molecular fueron hechas usando el paquete de programas Gromos96 (G96) [1]. El sistema estuvo compuesto, en todos los casos, por 1 molécula de soluto (aminoácido, o residuo de aminoácido) y 511 moléculas de solvente (agua o metanol).

En el campo de fuerzas de G96, los átomos de hidrógeno alifáticos y los átomos de carbono a los cuales están unidos, son tratados como una sola entidad. Los hidrógenos aromáticos y polares están explícitamente considerados. Los primeros experimentan interacciones electrostáticas y van der Waals, mientras que los hidrógenos polares experimentan solamente interacciones electrostáticas.

Toda la información relacionada con el modelo y parámetros usados para los solventes (agua y metanol), se encuentra en el manual de G96 [1]. Para el agua elegimos el modelo SPC/E [2].

En todos los casos, los parámetros utilizados fueron los siguientes:

- el paso de integración  $\Delta t = 2$  fs,
- para equilibrar termodinámicamente al sistema realizamos una Dinámica Molecular de 100 ps ( $5 \times 10^4$  pasos),
- las longitudes de unión fueron restringidas por medio del algoritmo SHAKE, con una tolerancia geométrica de  $10^{-4}$ ,
- todas las simulaciones se hicieron en el ensamble isotérmico–isobárico a 300 °K de temperatura y presión atmosférica normal (101325 Pa),
- tanto la temperatura como la presión se mantuvieron constantes a través de un acoplamiento débil del sistema a un baño externo, con tiempos de relajación de  $\tau_T = 0.01$  ps y  $\tau_P = 0.05$  ps,
- el sistema estuvo ubicado dentro de una caja rectangular, y se utilizaron condiciones periódicas de contorno para evitar efectos de superficie,
- elegimos un radio de corte aproximadamente igual a la mitad del lado más corto de la caja ( $\approx 10 \text{ \AA}$ ), el cual define el rango de las interacciones no

ligadas. Fuera de la esfera de corte, el solvente es tratado como un continuo, y se calcula un término extra de energía por campo de reacción,

- la permitividad dieléctrica considerada para el modelo SPC/E de agua es  $\epsilon = 70.70$ , y para el metanol  $\epsilon = 32.63$ .

Para determinar a las funciones de distribución radial y el número de puentes de hidrógeno que forman las moléculas de agua, realizamos corridas de 100 ps y recolectamos configuraciones del sistema cada 20 fs (10 pasos de simulación).

El criterio que utilizamos para determinar la formación de un puente de hidrógeno es geométrico. Consideramos que un átomo de oxígeno aceptor y otro donador forman un puente de hidrógeno cuando:

- la distancia entre el primero y un hidrógeno del segundo es menor que  $2.4 \text{ \AA}$ , y
- el ángulo O-H---O es mayor de  $145^\circ$  y menor de  $180^\circ$ .

Para calcular la variación de energía libre de transferencia ( $\Delta G_t$ ) de los residuos, usamos dos técnicas: integración termodinámica (IT) y crecimiento lento (CL) [3, 4].

En la técnica de IT, la interacción soluto–solvente es controlada por un parámetro  $\lambda$  que toma valores entre 0 (interacción completa) y 1 (sin interacción). Se realizan simulaciones para valores fijos de  $\lambda$ , y se calcula la variación de energía libre integrando en  $\lambda$ , mediante algún algoritmo de integración estándar. Previamente es necesario determinar la *ineficiencia estadística*, de modo de asegurar que los puntos tomados para los cálculos sean independientes [5]. Encontramos que para 300 ps, una muestra cada 0.2 ps (100 pasos de simulación) es suficiente para garantizar la independencia estadística de los datos.

Para realizar la integración numérica sobre  $\lambda$  utilizamos el método de Gauss–Lobatto [6]. Hacemos este cálculo considerando 3, 5, 7 y 9 valores de  $\lambda \in [0,1]$  (puntos de integración). Para este cálculo, llevamos al sistema hasta el valor de  $\lambda$  correspondiente mediante CL, y luego de 50 ps de equilibración, corrimos una DM de 200 ps.

Por otro lado, con el método de CL, el sistema es cuasiestáticamente transformado desde un estado en el cual la interacción es completa ( $\lambda = 0$ ), a otro en el cual los átomos del residuo y el solvente no interactúan ( $\lambda = 1$ ) [7].

Realizamos simulaciones en ambos sentidos, esto es, partiendo de una configuración inicial de equilibrio con todas las interacciones presentes ( $\lambda = 0$ ), hacemos evolucionar al sistema lentamente hasta un estado final donde los átomos pertenecientes al residuo del aminoácido en estudio, carecen completamente de toda interacción con los demás átomos del sistema ( $\lambda = 1$ ). Luego, tomamos esta última configuración correspondiente al estado sin interacción de los átomos del residuo ( $\lambda = 1$ ) como la inicial, y vamos encendiendo lentamente las interacciones hasta alcanzar un estado en el cual, vuelven a estar presentes ( $\lambda = 0$ ).

En las corridas con CL, tanto para agua como para metanol, el factor  $\lambda$  se hace variar de 0 a 1 en 500 ps de simulación. Esto significa que en el cálculo de  $\Delta G_t$ , en cada uno de los solventes, se simulan 1000 ps (500 ps para el apagado, y 500 ps más para el posterior encendido de las interacciones). Por tanto, el valor final para cada aminoácido, se obtiene luego de 2000 ps (2 ns) de simulación.

## B.1 REFERENCIAS

- 1 W.F. van Gunsteren, S.R. Billeter, A.A. Eising, W.R.P. Scott, and I.G. Tironi, *Biomolecular Simulation: The Gromos96 Manual and User Guide* (Zürich: vdf Hochschulverlag AG an der ETH Zürich, 1996).
- 2 H.J.C. Berendsen, J.R. Grigera and T.P. Straatsma, *J. Phys. Chem.*, **91**, 6269 (1987).
- 3 X. Daura, P.H. Hünenberger, A. E. Mark, E. Querol, F.X. Avilés and W.F. van Gunsteren, *J. Am. Chem. Soc.*, **118**, 6285 (1996).
- 4 W.F. van Gunsteren and P.K. Weiner, *Computer Simulation of Biomolecular System*, (Leiden: ESCOM Science Publishers, 1989).
- 5 M. P. Allen and D.J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, 1987).

6 M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions*, (New York: Dover, 1965).

7 H.J.C. Berendsen, J.P.M. Postma and W.F. van Gunsteren, *Molecular Dynamics and Protein Structure*, edited by J. Hermans (Western Springs, IL: Polycrystal Book Service, 1985).

## C. VALORES DE LOS PARÁMETROS UTILIZADOS EN EL CÁLCULO DE LAS FUNCIONES DE DISTRIBUCIÓN, MEDIANTE LA TEORÍA MECÁNICO-ESTADÍSTICA.

Todos los cálculos realizados con la teoría mecánico-estadística desarrollada en los *capítulos* 3, 4 y 5 fueron efectuados considerando al sistema formado por 1 o 2 moléculas de residuo disueltas en agua. En la Tabla C.1 detallamos a los parámetros que caracterizan a las moléculas mencionadas.

Tabla C.1: parámetros que caracterizan a los residuos y al agua.

molécula	$\sigma[\text{Å}]$	$\varepsilon/k_B [^{\circ}\text{K}]$	$\mu[\text{Dy}]$	$q/e$
alanina (A)	3.794	148		
prolina (P)	4.932	120		
valina (V)	5.324	120		
metionina (M)	5.818	110		
isoleucina (I)	5.887	90		
leucina (L)	5.887	80		
fenilalanina (F)	6.277	70		
ácido aspártico (D)	3.595	160		-0.110
ácido glutámico (E)	4.200	120		-0.120
lisina (K)	5.500	100		0.185
arginina (R)	5.436	100		0.165
serina (S)	3.785	148	1.350	
cisteína (C)	4.574	120	1.850	
treonina (T)	4.741	120	1.900	
asparagina (N)	4.797	120	2.150	
glutamina (Q)	5.438	100	2.450	
histidina (H)	5.607	90	2.550	
tirosina (Y)	6.328	60	2.800	
triptófano (W)	6.835	50	3.100	
agua ( $\alpha$ )	3.166	78.2	2.351	

Al final de la Tabla C.1 agregamos los parámetros que caracterizan al agua. Estos parámetros son los correspondientes al modelo SPC/E, utilizado en todas las simulaciones de Dinámica Molecular con Gromos96.

En la teoría mecánico–estadística presentada, debemos considerar un valor para la densidad del agua. En este caso, el criterio que utilizamos está relacionado nuevamente con las simulaciones. Consideramos la densidad que se obtiene al dividir el número total de moléculas por el volumen promedio de la caja que ocupan, durante 100 ps de simulación, a presión normal (1 atm) y 300 °K de temperatura. Encontramos un valor  $\rho_\alpha = 0.0326 \text{ \AA}^{-3}$ .

Para los parámetros que caracterizan a las esferas que representan a los residuos de aminoácidos, procedemos del siguiente modo. La segunda columna de la Tabla C.1 correspondiente al diámetro de van der Waals ( $\sigma_\rho$ ), deriva de medidas experimentales realizadas por otros autores.

En los aminoácidos apolares, donde necesitamos determinar  $\varepsilon_\rho$ , hacemos simulaciones mediante Dinámica Molecular (DM) de cada uno de los residuos en agua. Luego, obtenemos la función  $g_{\rho\alpha}^{DM}(r)$ , y con ella, el número de coordinación ( $N_C$ ) a través de la expresión usual (3.19). A continuación, asignamos un valor a  $\varepsilon_\rho$ , y calculamos la  $g_{\rho\alpha}(r)$  resolviendo las ecuaciones integrales correspondientes. Finalmente, ajustamos el parámetro  $\varepsilon_\rho$  de manera que el  $N_C$  surgido de la teoría difiera en menos del 1% del  $N_C$  proveniente de la simulación.

Para los aminoácidos cargados necesitamos un parámetro más, la carga efectiva  $q_\rho$ . En estos residuos, asignamos a  $\varepsilon_\rho$  un valor cercano al de un aminoácido apolar de  $\sigma_\rho$  semejante. Luego, ajustamos el valor para la carga efectiva con el mismo procedimiento del número de coordinación descrito en el párrafo anterior. Finalmente, procuramos optimizar ambos parámetros cotejando no sólo el valor numérico de  $N_C$ , sino también, los rasgos característicos de las funciones de distribución radial.

En cuanto a los aminoácidos polares, el criterio utilizado coincide con el expresado recién para los cargados. Elegimos el  $\varepsilon_\rho$  que más se asemeje al de un

residuo apolar de tamaño próximo, y ajustamos el parámetro restante  $\mu_\rho$  buscando conseguir un buen acuerdo entre las funciones de distribución radial residuo–agua y los  $N_C$  que surgen de la teoría con los procedentes de la simulación.

Por último, mencionaremos algunos parámetros extra que utilizamos en el algoritmo con el cual resolvemos las ecuaciones integrales de la teoría mecánico–estadística.

En el espacio de las configuraciones, el paso de la función  $g_{\alpha\alpha}^{\text{exp}}(r)$  es  $\Delta r = 0.05 \text{ \AA}$ . El número total de puntos sobre los cuales evaluamos a las funciones de distribución radial y a los potenciales de campo medio es  $N = 1024$ , lo que da un rango de distancias de 0 a  $51.150 \text{ \AA}$ .