



UNIVERSIDAD NACIONAL AUTÓNOMA DE  
MÉXICO

**PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS  
QUÍMICAS**

“ESTUDIO POR DINÁMICA MOLECULAR DE LA ESFERA DE SOLVATACIÓN Y DE LAS  
INTERACCIONES POR PUENTES DE AGUA EN DIVERSAS PROTEÍNAS”

TESIS  
PARA OPTAR POR EL GRADO DE

**MAESTRO EN CIENCIAS**

PRESENTA

EDDY FRANCIS PLAZA MEDINA



TUTOR: JORGE PEÓN PERALTA

AÑO: 2007



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



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

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

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

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

Jurado Asignado:

Presidente: Dr. Armando Gómez-Puyou  
Vocal: Dra. María Eugenia Costas Basín  
Secretario: Dra. Adela Rodríguez Romero  
Primer suplente: Dr. Luis Miguel Trejo Candelas  
Segundo suplente: Dr. Ángel Piñeiro Guillén

Asesor

---

Dr. Jorge Peón Peralta

## **AGRADECIMIENTOS**

*No se puede ignorar el hecho, de que el mundo entero conspira para que cada uno de nosotros alcance alguna meta, esta es la razón por la cual hoy doy gracias a cada uno de los congéneres que de una u otra forma han contribuido a que cumpla con este logro. En primera instancia agradezco al Dr. Jorge Peón Peralta, por su paciencia, dedicación, aportes y apoyo, desde el momento en que me vinculé a su grupo de trabajo, a mis compañeros de grupo por su solidaridad, a mi familia por sus enseñanzas y apoyo incondicional, a los miembros del jurado por sus observaciones y comentarios, a la comunidad colombiana por todos los momentos compartidos, a los mexicanos por su hospitalidad y a la Universidad Autónoma de México por la oportunidad brindada. También doy mi más sincero agradecimiento al Dr. Guillermo Aguilar, porque sin su apoyo inicial no hubiera alcanzado esta meta.*

*Finalmente agradezco al CONACYT, proyecto 42663-Q, por la beca otorgada.*

*Dedico esta tesis  
A Dios y a mi Familia*

*Esta tesis se desarrolló en las instalaciones del Instituto de Química, en las salas de sistema del área de  
Fisicoquímica, bajo la dirección del Dr. Jorge Peón Peralta*

## Contenido

	página
<b>1. Introducción</b> .....	1
1.1. Las interacciones agua-proteína: Su estudio e importancia.....	1
1.2. Objetivos generales y sistemas de estudio.....	5
1.3. Objetivos específicos.....	8
<b>2. Antecedentes</b> .....	10
2.1. Simulaciones de dinámica molecular.....	10
2.1.1. Campos de fuerza.....	11
2.1.2. Integración de las ecuaciones de movimiento.....	16
2.1.3. Preparación y corrida de las simulaciones de dinámica molecular.....	17
2.2. Análisis de la simulación de dinámica molecular.....	20
2.2.1. Función de supervivencia del solvente en la esfera de solvatación de un átomo $\alpha$ de una proteína $P_{\alpha}(t)$ .....	20
2.2.2. Función de distribución radial de dos centros.....	23
2.3. Proteínas objeto de estudio.....	29
2.3.1. FKBP12.....	29
2.3.2. Tendamistat.....	29
2.3.3. Complejo barnasa-barstar.....	30
<b>3. Metodología</b> .....	34
3.1. Preparación de los sistemas.....	34
3.2. Minimización y equilibrio de los sistemas.....	36
3.3. Simulación de dinámica molecular.....	40
3.4. Ventajas de los programas empleados para el análisis de la simulación de dinámica molecular.....	46
<b>4. Resultados y discusión</b> .....	52
4.1. Resultados de la función de supervivencia.....	52

4.1.1. Forma general del decaimiento de la función $P_{\alpha}(t)$ .....	52
4.1.2. Ventajas de los archivos PDB.....	58
4.1.3. Gráficas de distribución de los tiempos de residencia.....	59
4.1.4. Comentarios generales de la distribución de los tiempos de residencia en la superficie.....	63
4.1.5. Distribución espacial del número de moléculas de aguas coordinadas.....	67
4.1.6. Comparación de los resultados de las simulaciones de dinámica molecular con los de rayos X y de RMN de la FKBP12.....	73
4.1.7. Comparación de los resultados de la simulación de la dinámica molecular con los de RMN de la proteína tendamistat.....	80
4.1.8. Comparación de los resultados de la simulación de la dinámica molecular del complejo barnasa-barstar con los resultados de la simulación de dinámica molecular de las proteínas libres.....	88
4.2. Resultados de la función de distribución radial de dos centros.....	97
<b>5. Conclusiones.....</b>	<b>103</b>
<b>Apéndice A.....</b>	<b>107</b>
A.1 Funciones del programa de la función de supervivencia.....	107
A.2 Funciones del programa de la función radial de dos centros.....	108
A.3 Funciones del script (TRAPro).....	109
A.4 Programa de la función de supervivencia.....	109
A.5 Programa de la función de distribución radial de dos centros.....	116
A.6 Script que se emplea para la ejecución del programa de la función de supervivencia.....	125
<b>Referencias bibliográficas.....</b>	<b>128</b>

## Índice de figuras

Figuras	página
<b>Figuras del capítulo 2</b>	
<b>Figura 2.1.</b> Gráfica de la función $P_{\alpha}(t)$ (función de supervivencia de las moléculas de agua en la esfera de solvatación) del átomo de oxígeno del carbonilo del residuo tirosina 103 de la proteína barnasa.....	22
<b>Figura 2.2.</b> Representación de la intersección de las esferas de solvatación de dos átomos de una proteína o de proteínas diferentes. La zona delimitada mediante líneas indica el volumen de intersección. R es la distancia entre los dos átomos, $r_1$ y $r_2$ el radio de cada esfera y $dr$ es la diferencia entre las dos esferas de un mismo átomo.....	26
<b>Figura 2.3.</b> Evaluación de la función radial de dos centros sobre los dos oxígenos del grupo carboxílico del ácido glutámico de la proteína barstar, donde cada oxígeno presenta una zona de alta densidad de agua (zona oscura), las cuales se ubican a 2.75 Å de uno de los átomos y a aproximadamente 4.8 Å del otro.....	27
<b>Figura 2.4.</b> Evaluación de la función radial de dos centros para el nitrógeno de la cadena principal de la serina 38 de la proteína barnasa y el oxígeno del grupo carbonilo de la valina 45 de la proteína barstar. La zona de alta densidad (zona oscura) es un indicador de que ambos átomos coordinan simultáneamente a una misma molécula de agua, pues la distancia de la molécula de agua con respecto a cada uno de los átomos, corresponde a una distancia típica de un puente de hidrógeno.....	28
<b>Figura 2.5.</b> Diagrama de la proteína FKBP12 plegada, constituida por 5 hojas $\beta$ antiparalelas y una hélice $\alpha$ de color verde, utilizando el programa VMD. Las hojas $\beta$ están distribuidas así: $\beta_1$ de color azul, $\beta_2$ de color plateado, $\beta_3$ de color rojo, $\beta_4$ de color rosado y $\beta_5$ de color morado.....	31



**Figura 2.6.** Diagrama la proteína tendamistat plegada, formando una estructura  $\beta$  de 6 hojas, utilizando el programa VMD. Las hojas  $\beta$  están distribuidas así:  $\beta_1$  de color azul,  $\beta_2$  de color rojo,  $\beta_3$  de color beigs,  $\beta_4$  de color rosado,  $\beta_5$  de color verde y  $\beta_6$  de color morado.....32

**Figura 2.7.** Diagrama del complejo barnasa-barstar plegado, utilizando el programa VMD. Barnasa está formada por 3 hélices  $\alpha$  ( $\alpha_1$  de color verde,  $\alpha_2$  de color plata y  $\alpha_3$  de color beigs) y 5 hojas  $\beta$  antiparalelas ( $\beta_1$  de color amarillo,  $\beta_2$  de color naranja,  $\beta_3$  de color gris,  $\beta_4$  de color rojo y  $\beta_5$  de color azul) y barstar la integran 4 hélices  $\alpha$  ( $\alpha_1$  de color blanco,  $\alpha_2$  de color rosado,  $\alpha_3$  de color morado y  $\alpha_4$  de color verde claro) y 3 hojas  $\beta$  paralelas ( $\beta_1$  de color azul claro,  $\beta_2$  de color café y  $\beta_3$  de color morado claro).....33

### **Figuras del capítulo 3**

**Figura 3.1.** Comportamiento de la energía total en la minimización y equilibrio de la proteína barnasa. Los primeros 3000 pasos corresponden a la minimización y los restante 20000 pasos, corresponden al equilibrio.....38

**Figura 3.2.** Comportamiento de la temperatura en la minimización y equilibrio de la proteína barnasa. Los primeros 3000 pasos corresponden a la minimización y los restante 20000 pasos, corresponden al equilibrio.....39

**Figura 3.3.** Gráficas de RMSD vs tiempo. La primera gráfica de barstar (barstar A) es una representación de los últimos 9000 ps de su trayectoria de 15000 ps. La segunda gráfica de barstar (barstar B) corresponde a la elección de los 800 ps más estables, considerando la gráfica de los 9000 ps (barstar A). En las restantes gráficas se representan los 800 ps más estables de cada una de las proteínas, los cuales se eligieron para el posterior análisis. Para construir cada una de estas gráficas, se consideró la primera configuración después del equilibrio, como la de referencia.....44

**Figura 3.4.** Radio de giro vs tiempo. En las cinco gráficas se representan los 800 ps que se eligieron de cada una de las proteínas, y los cuales se utilizan más adelante en el análisis de la simulación de dinámica molecular .....45

**Figura 3.5.** Función de supervivencia vs tiempo de residencia del agua en la esfera de solvatación del átomo de N de la cadena principal del aminoácido Gly1 de la FKBP12, a diferentes  $\Delta t$ . Línea continúa  $\Delta t = 0.2$  ps Línea punteada  $\Delta t = 0.4$  ps y línea discontinúa  $\Delta t = 1.0$  ps.....47

**Figura 3.6.** Datos de la función de supervivencia vs tiempo de residencia de la esfera de solvatación del átomo de oxígeno del grupo carbonilo del ácido aspártico N°79 de la proteína FKBP12, calculados con el programa de la función de supervivencia (TRAPro). El 50% y el 20% son un par de ejemplos indicados en la gráfica y los cuales corresponden a la cantidad de moléculas de agua presentes en la esfera de solvatación cuando la función de supervivencia ha alcanzado un tiempo de 3.27 ps y de 9.58 ps respectivamente.....49

#### **Figuras del capítulo 4**

**Figura 4.1.** Tiempo de decaimiento de la función  $P_{\alpha}(t)$  de un átomo de oxígeno, correspondiente a una molécula de agua del seno de la solución del complejo barnasa-barstar.....53

**Figura 4.2.** Tiempo de decaimiento de la función  $P_{\alpha}(t)$  para diferentes átomos de la proteína FKBP12 (Gly1:O del grupo carbonilo, Glu31:N de la cadena principal, Gln70:NE1 y Asp79:O del grupo carbonilo). Esta gráfica es una representación de las moléculas de agua que se intercambian rápidamente con el seno de la solución.....56

**Figura 4.3.** Tiempo de decaimiento de la función  $P_{\alpha}(t)$  para diferentes átomos de la proteína FKBP12 (Val55:O del grupo carbonilo, Ser67:OG, Thr96:O del grupo carbonilo y Gly10:N de la cadena principal). Esta gráfica es una representación de las

moléculas de agua que presentan altos tiempos de residencia en la esfera de solvatación.....57

**Figura 4.4.** Distribución de los tiempos residencia de las moléculas de agua dentro de la primera esfera de solvatación de la proteína tendamistat, sin considerar los átomos de hidrógeno y teniendo en cuenta  $\tau_2$  de decaimiento.....61

**Figura 4.5.** Distribución de los tiempos residencia de las moléculas de agua de la proteína tendamistat dentro de la primera esfera de solvatación, considerando los valores de  $P_\alpha(t=0) > 0.15$  y teniendo en cuenta  $\tau_2$  de decaimiento.....62

**Figura 4.6.** Distribución de las moléculas de agua en la proteína FKBP12, de acuerdo al tiempo de residencia de éstas en la proteína, mediante la evaluación de la función de supervivencia, teniendo en cuenta  $\tau_2$  de decaimiento. La escala de colores es un indicador del tiempo de residencia de las moléculas de agua en la primera esfera de solvatación. Esta escala va desde el rojo que corresponde a las moléculas de agua que se intercambian rápidamente con el seno de la solución, al azul que indica altos tiempos de residencia de las moléculas de agua.....64

**Figura 4.7.** Distribución de las moléculas de agua en la proteína tendamistat, de acuerdo al tiempo de residencia de éstas en la proteína, mediante la evaluación de la función de supervivencia, teniendo en cuenta  $\tau_2$  de decaimiento. La escala de colores es un indicador del tiempo de residencia de las moléculas de agua en la primera esfera de solvatación. Esta escala va desde el rojo que corresponde a las moléculas de agua que se intercambian rápidamente con el seno de la solución, al azul que indica altos tiempos de residencia de las moléculas de agua.....65

**Figura 4.8.** Distribución de las moléculas de agua en el complejo, de acuerdo al tiempo de residencia de éstas en la proteína, mediante la evaluación de la función de supervivencia, teniendo en cuenta  $\tau_2$  de decaimiento. La escala de colores es un

indicador del tiempo de residencia de las moléculas de agua en la primera esfera de solvatación. Esta escala va desde el rojo que corresponde a las moléculas de agua que se intercambian rápidamente con el seno de la solución, al azul que indica altos tiempos de residencia de las moléculas de agua. En **A** se observa el complejo, en **B** se muestra el lado de la proteína barnasa que interactúa con la proteína barstar, en **C** se muestra el lado de la proteína barstar que interactúa con barnasa.....66

**Figura 4.9.** Mapa de colores del número de aguas coordinadas de la proteína FKBP12, considerando el archivo  $P_{\alpha}(t=0)$  (número de aguas coordinadas a un átomo). La escala de colores va desde el rojo que corresponde a átomos que no están rodeados por moléculas de agua, al azul que indica átomos rodeados por 6 ó 7 moléculas de agua.....69

**Figura 4.10.** Distribución del número de moléculas de agua en la primera esfera de solvatación de los átomos de la proteína FKBP12, sin considerar los átomos de hidrógeno; los datos se obtuvieron del valor de  $P_{\alpha}(t=0)$  (promedio del número de aguas coordinadas a un átomo).....70

**Figura 4.11.** Distribución del número de moléculas de agua en la primera esfera de solvatación de los átomos de la proteína FKBP12, considerando los valores de  $P_{\alpha}(t=0) > 0.15$ , teniendo en cuenta los datos de  $P_{\alpha}(t=0)$  (número de aguas coordinadas a un átomo).....71

**Figura 4.12.** Gráfica del tiempo de residencia del agua. Comparación de los sitios de hidratación de la simulación de DM de la FKBP12, considerando  $\tau_2$  de decaimiento vs los átomos correspondientes a las clases de los sitios de hidratación reportados por Karplus. La línea punteada representa el tiempo promedio de permanencia de las moléculas de agua en la esfera de solvatación, cuando  $P_{\alpha}(t=0) > 0.15$ .....74

**Figura 4.13.** Identificación de los átomos donde se ubica la molécula de agua N°1 en el mapa de colores de la proteína FKBP12, considerando  $\tau_2$  de decaimiento. En amarillo se encuentran encerrados los átomos que están a menos de 5 Å de los N de la cadena principal de la Gly86, His94 y Ala95. La escala de colores es un indicador del tiempo de residencia de las moléculas de agua en la primera esfera de solvatación. Esta escala va desde el rojo que corresponde a las moléculas de agua que se intercambian rápidamente con el seno de la solución, al azul que indica altos tiempos de residencia de las moléculas de agua.....76

**Figura 4.14.** Identificación de los átomos donde se ubica la molécula de agua N°2 en el mapa de colores de la proteína FKBP12, considerando  $\tau_2$  de decaimiento. En amarillo se encuentran encerrados los átomos que están a menos de 5 Å del N de la cadena principal de la Tyr82. Los átomos de color verde sin encerrar están próximos al N de la cadena principal de la Gly83, al igual que el CA de la Ala81 y el C del grupo carbonilo de la Tyr80. La escala de colores es un indicador del tiempo de residencia de las moléculas de agua en la primera esfera de solvatación. Esta escala va desde el rojo que corresponde a las moléculas de agua que se intercambian rápidamente con el seno de la solución, al azul que indica altos tiempos de residencia de las moléculas de agua.....78

**Figura 4.15.** Identificación de los átomos donde se ubica la molécula de agua N°4 en el mapa de colores de la proteína FKBP12, considerando  $\tau_2$  de decaimiento. En fucsia aparecen los N de la cadena principal de los residuos Met49, Lys52, Glu54 y Val55. La escala de colores es un indicador del tiempo de residencia de las moléculas de agua en la primera esfera de solvatación. Esta escala va desde el rojo que corresponde a las moléculas de agua que se intercambian rápidamente con el seno de la solución, al azul que indica altos tiempos de residencia de las moléculas de agua.....79

**Figura 4.16.** Comparación de la conectividad  $\delta - \varepsilon$  de la Tyr46 reportada por scarselli con los resultados de la función de supervivencia, considerando  $\tau_2$  de decaimiento. En **A** se identifica la Tyr46 dentro de la proteína tendamistat. En **B** se muestran los posibles puentes de hidrógenos que puede formar el HE1 (protón blanco) de la Tyr46 y en **C** los posibles puentes de hidrógeno que puede formar el HD1 (protón blanco) de la Tyr46. La escala de colores es un indicador del tiempo de residencia de las moléculas de agua en la primera esfera de solvatación. Esta escala va desde el rojo que corresponde a las moléculas de agua que se intercambian rápidamente con el seno de la solución, al azul que indica altos tiempos de residencia de las moléculas de agua.....83

**Figura 4.17.** Diagrama de barras en el que se muestran los átomos que pueden formar puentes de hidrógeno con los HD1 y HE1 de la Tyr46 de la proteína tendamistat a una distancia inferior a los 5 Å, considerando  $\tau_2$  de decaimiento. La línea punteada indica el tiempo promedio de permanencia de las moléculas de agua dentro de la primera esfera de solvatación, considerando  $P_\alpha(t=0) > 0.15$ .....84

**Figura 4.18.** Diagrama de barras en el que se muestran los átomos que pueden formar puentes de hidrógeno con los HG1 y HD1 de la Pro50 de la proteína tendamistat a una distancia inferior a los 5 Å, considerando  $\tau_2$  de decaimiento. La línea punteada indica el tiempo promedio de permanencia de las moléculas de agua dentro de la primera esfera de solvatación, considerando  $P_\alpha(t=0) > 0.15$ .....85

**Figura 4.19.** Diagrama de barras en el que se muestran los átomos que pueden formar puentes de hidrógeno con los HA y HG de la Thr3 de la proteína tendamistat a una distancia inferior a los 5 Å, considerando  $\tau_2$  de decaimiento. La línea punteada indica el tiempo promedio de permanencia de las moléculas de agua dentro de la primera esfera de solvatación, considerando  $P_\alpha(t=0) > 0.15$ .....86

**Figura 4.20.** Diagrama de barras en el que se muestran los átomos que pueden formar puentes de hidrógeno con los HA, HG1 y HG22 de la Thr54 de la proteína tendamistat a una distancia inferior a los 5 Å, considerando  $\tau_2$  de decaimiento. La línea punteada indica el tiempo promedio de permanencia de las moléculas de agua dentro de la primera esfera de solvatación, considerando  $P_\alpha(t=0) > 0.15$ .....87

**Figura 4.21.** Interfase entre las proteínas barnasa-barstar, mediada por la hélice  $\alpha 2$  (color rosado) de barstar. Esta hélice divide a la interfase en dos zonas: una de alta complementariedad y otra de poca complementariedad. La primera la comprenden el Asp39 de barstar (color blanco) y los aminoácidos Arg83, Arg87 e His107 (color azul) de barnasa. La segunda zona corresponde al lado de la hélice  $\alpha 2$  que colinda con el rizo (color azul claro) de barnasa que interactúa con la guanina.....96

**Figura 4.22.** Evaluación de la función radial de dos centros para átomos de la interfase entre barnasa y barstar, con la finalidad de identificar moléculas de agua que medien dichas interacciones.....99

**Figura 4.23.** Evaluación de la función radial de dos centros para átomos de la interfase entre barnasa y barstar, con la finalidad de identificar moléculas de agua que medien dichas interacciones.....100

**Figura 4.24.** Evaluación de la función radial de dos centros para átomos de la interfase entre barnasa y barstar, con la finalidad de identificar moléculas de agua que medien dichas interacciones.....101

**Figura 4.25.** Evaluación de la función radial de dos centros para átomos de la interfase entre barnasa y barstar, con la finalidad de identificar moléculas de agua que medien dichas interacciones.....102

## Índice de tablas

	página
<b>Tabla 1:</b> Algunos átomos del complejo, de la tendamistat y de la FKBP12 que presentan tiempos de residencia altos.....	72
<b>Tabla 2:</b> Átomos de la interfase de la proteína barnasa que presentan altos tiempos de residencia, obtenidos de la simulación de DM del complejo, considerando $\tau_2$ de decaimiento. En gris aparecen los átomos de alto tiempo de residencia que coinciden con los resultados de la simulación de DM de barnasa libre.....	91
<b>Tabla 3:</b> Átomos de la interfase de la proteína barstar que presentan altos tiempos de residencia, obtenidos de la simulación de DM del complejo, considerando $\tau_2$ de decaimiento.....	92
<b>Tabla 4:</b> Moléculas de agua que median el acoplamiento entre barnasa y barstar, reportadas por Buckle. En negrilla aparecen los átomos de barstar y en claro los átomos de barnasa. En gris aparecen los átomos que coinciden con los resultados de la simulación de DM del complejo.....	93



## Lista de acrónimos y abreviaturas

DM	Dinámica molecular
RMN	Resonancia Magnética Nuclear
NOE	Efecto nuclear Overhauser
RMSD	Desviación de la raíz cuadrática media
Rg	Radio de giro
PDB	Archivos del banco de datos de las proteínas
NAMD	Paquetes de programas para la simulación de dinámica molecular
VMD	Programa de visualización molecular
PME	Partícula-superficie de Ewald
N	Número de partículas
V	Volumen
P	Presión
T	Temperatura
E	Energía
<i>t</i>	Tiempo
ps	Picosegundos
ns	Nanosegundos
fs	Femtosegundos
Å	Angstroms
Sc	Grado de complementariedad de dos proteínas
TIP3P	Modelo del agua
ESA	Área de superficie expuesta
TRAPro	Tiempo de residencia del agua en la proteína
Atm	atmósfera
K	Kelvin
BAKLIZ	Supercomputadora de la UNAM

## Resumen

Las moléculas de agua que forman parte del medio celular y tisular de las proteínas juegan un papel muy importante en su funcionamiento, estabilidad y estructura. Las simulaciones de dinámica molecular a diferencia de los métodos experimentales, tienen la ventaja de proporcionar información directamente en la escala de tiempo de las fluctuaciones de la esfera de solvatación además de proveer detalles a nivel atómico sobre el comportamiento del agua en torno a las proteínas. En nuestro caso se utilizará esta metodología para identificar las moléculas de agua que juegan un papel importante en las proteínas FKBP12, tendamistat, barnasa, barstar y en el complejo barnasa-barstar. Las simulaciones de dinámica molecular en el ensamble isobárico-isotérmico produjeron trayectorias que se analizaron mediante la función de supervivencia del disolvente en un sitio de hidratación de una proteína y con la función de distribución radial de dos centros  $g(r_1, r_2)$ . La primera función tiene como finalidad medir la dinámica de intercambio de las moléculas de agua entre la esfera de solvatación de cada átomo de la proteína y el seno del disolvente, y la segunda función se empleó, para identificar las moléculas de agua que están coordinadas simultáneamente a dos átomos de una proteína mediante enlaces por puentes de hidrógeno. El análisis mediante la función de supervivencia permitió diferenciar las zonas de la superficie de la proteína de acuerdo al tiempo de residencia de las moléculas de agua en torno a los átomos de la macromolécula. En este estudio el mapeo de dicha propiedad en la superficie de estas proteínas nos permitió concluir que la mayoría de las moléculas de agua se intercambian rápidamente con el seno de la solución y muy pocas moléculas permanecen en la esfera de solvatación durante tiempos mayores a algunas decenas de picosegundos. Al comparar nuestros resultados con los de difracción de rayos X y los de RMN de las proteínas FKBP12 y tendamistat respectivamente, se encontró que la mayoría de los resultados concuerdan con los de RMN pero solo algunos casos concuerdan con los sitios

identificados por difracción de rayos X. Estos resultados nos permiten concluir que aunque todas las moléculas de agua de intercambio lento son identificables por rayos X, solo una pequeña fracción de los sitios de hidratación identificados por rayos X es de intercambio lento. De la interacción entre el complejo barnasa-barstar se dedujo que la proteína barstar se enlaza a barnasa gracias a la presencia de moléculas de agua de alto tiempo de residencia localizadas en barnasa, además, se identificaron moléculas de agua de alto tiempo de residencia que median dicha interacción. Finalmente, con la función de distribución radial de dos centros  $g(r_1, r_2)$ , determinamos que algunas de las moléculas de agua de alto tiempo de residencia que median la interacción barnasa-barstar, están coordinadas simultáneamente a dos átomos del complejo mediante enlaces puentes de hidrógeno. También con esta función se determinó, que hay moléculas de agua de alto tiempo de residencia en barnasa libre que se conservan en el complejo y que incrementan notablemente la complementariedad de las superficies de reconocimiento. Adicionalmente, hay otras moléculas superficiales de disolvente que se reacomodan durante el acoplamiento entre barnasa y barstar y otras que aparecen únicamente en el complejo.

# 1. Introducción

## 1.1. Las interacciones agua-proteína: Su estudio e importancia

Gracias a su estructura en forma de "V" y a su capacidad de donar y aceptar enlaces de hidrógeno, el agua no es simplemente un medio viscoso donde las proteínas pueden permanecer en solución y difundir, sino que más aún, este vital líquido representa un factor muy específico que a nivel molecular determina la estructura, la dinámica y el funcionamiento de las proteínas.

Por ejemplo, las moléculas de agua participan directamente en la definición de la estructura tridimensional proteínica al contribuir a que la energía libre de plegamiento sea favorable mediante efectos hidrofóbicos e hidrofílicos<sup>1</sup>. Adicionalmente, existen moléculas de agua que se ubican dentro de las cavidades de las proteínas (moléculas de agua internas) y a las cuales se les ha conferido la propiedad de estabilizar las macromoléculas mediante redes de puentes de hidrógeno. Por ejemplo, se ha encontrado que mutantes de la lisozima humana que presentan moléculas de agua en las cavidades internas, son más estables que las mutantes de la lisozima humana con cavidades vacías<sup>2</sup>

También, se ha encontrado que las moléculas de agua de la superficie pueden participar estructuralmente en las uniones proteína-proteína o proteína-ligando. En este sentido, las moléculas de agua funcionan como puentes (de agua) al hacer un enlace de hidrógeno con dos átomos simultáneamente, uno con cada molécula del complejo. Tal es el caso del complejo entre la ciclofilina A y la ciclosporina A, donde se ha encontrado, por medio de cristalografía de rayos X, que 36 moléculas de agua se ubican en la interfase entre las dos moléculas y median la interacción<sup>3</sup> Otro ejemplo al respecto es la formación del complejo entre la proteína SGPB (proteasa de tipo quimotripsina) con la proteína OMTKY3 (proteína inhibidora de proteasa); en este sistema, a través de estudios de rayos X, se ha observado que para la formación del complejo, es necesaria una importante redistribución de las moléculas de agua de la

zona de reconocimiento para optimizar las interacciones por puentes de agua entre la enzima y el inhibidor<sup>4</sup>. Un tercer ejemplo es el siguiente: En el complejo entre CDK2 (quinasa) y la isopenteniladenina se encontró, a través de simulaciones de dinámica molecular, que dos moléculas de agua contribuyen en la interacción entre el inhibidor y la enzima vía enlaces de hidrógeno. Además, se observó que dos moléculas de agua en la trayectoria libre CDK2 ocupan las mismas posiciones de los nitrógenos N3 y N9 de la isopenteniladenina, concluyendo con esto que el tomar en consideración moléculas de agua fuertemente enlazadas a la proteína puede ayudar substancialmente en el diseño de nuevos inhibidores<sup>5</sup>.

El agua también influye en la actividad catalítica de las enzimas. En este sentido, se han hecho estudios con micelas en fase reversa en cuyas cavidades es posible limitar la cantidad de agua disponible para solvatar una enzima<sup>6</sup>. Estas investigaciones determinaron que la actividad enzimática depende de manera directa de la cantidad de agua disponible para interactuar con la proteína y que en ausencia de una cantidad suficiente de ésta, la actividad catalítica prácticamente desaparece<sup>7</sup>.

Asimismo, se ha contemplado la importancia de las moléculas de agua que se ubican en sitios específicos de la superficie proteínica y que realizan funciones altamente especializadas en los movimientos o fluctuaciones estructurales de las proteínas, las cuales a su vez están relacionadas con los mecanismos enzimáticos. Un ejemplo de ello es el siguiente: El sitio activo de la triosa fosfato isomerasa (TIMs), cuenta con una especie de cubierta (asa catalítica), la cual constantemente se está abriendo y cerrando para permitir la entrada y salida del sustrato. Cuando la TIM se encuentra en su forma abierta, se ha encontrado que la posición de la cubierta es compensada por 3 moléculas de agua las cuales se deben intercambiar con el seno de la solución para permitir un mecanismo de bisagra. De hecho, se ha establecido que la frecuencia de apertura-cierre queda al menos en parte determinada por la frecuencia promedio con la que se desplazan las tres aguas mencionadas anteriormente<sup>8</sup>.

Diversos estudios experimentales recientes indican que las moléculas de agua que se encuentran alrededor de una proteína no forman un cascarón semi-rígido entorno a ella, ni mucho menos difunden al unísono con la proteína manteniéndose unidas a ésta como se llegó a pensar<sup>9</sup>; sino que por el contrario, estas moléculas solvatan a la proteína con diferentes energías libres de unión en diferentes partes de la macromolécula, y que muy probablemente se intercambian a muy diversas velocidades con las moléculas del seno de la solución<sup>10</sup>.

Entre las observaciones experimentales que indican esta alta heterogeneidad en la dinámica de hidratación en la superficie de una proteína, se encuentran la resonancia magnética nuclear<sup>11</sup>, la cristalografía de rayos X<sup>12</sup>, y los estudios de fluorescencia de triptófano con resolución temporal de femtosegundos<sup>13</sup>.

Las técnicas de RMN son capaces de identificar mediante el efecto nuclear de Overhouser (NOE), a moléculas de agua (tanto internas como en cavidades superficiales) que permanecen en contacto con la proteína durante un tiempo suficiente como para que ocurra el fenómeno de transferencia de magnetización. Modelos sencillos sobre la dinámica de dicha transferencia estiman que las técnicas por NOE identifican aquellas moléculas de disolvente que se mantienen unidas a la proteína por al menos unos 300 ps<sup>14</sup>. El número de moléculas de agua que se identifican mediante dichas técnicas corresponde a un número relativamente pequeño en comparación con el número de aguas presentes en la primera esfera de solvatación. Por ejemplo, en un estudio de la proteína BPTI (inhibidor)<sup>iError! Marcador no definido.</sup> se localizaron por RMN-NOE, a 4 moléculas de agua, mientras que el número total de aguas a una distancia de menos de 4 Å de la proteína en solución se puede estimar como de alrededor de 400.

Los estudios de difracción de rayos X en proteínas detectan la presencia de un gran número de moléculas de disolvente en los cristales difractados. El número de

aguas localizadas por difracción depende del tamaño de la proteína, de las condiciones de la medición y de la resolución del experimento, y pueden ir de varias decenas a varios cientos de moléculas de agua. Por este método es posible localizar aguas tanto internas como externas a la proteína y se considera que dichas posiciones corresponden a sitios de una elevada densidad promedio del átomo de oxígeno del agua en un cristal particular (sitio de hidratación). Sin embargo, se ha observado que para la mayoría de las aguas exteriores existe una enorme variabilidad en la posición de las mismas cuando se estudia la misma proteína en diferentes empaquetamientos cristalinos<sup>15</sup>. Es decir, la ubicación de estas moléculas de disolvente depende en gran medida del tipo de arreglo tridimensional de la celda unitaria del cristal. Dentro del grupo de moléculas de agua localizadas por difracción en una proteína es posible distinguir una alta heterogeneidad en su grado de movilidad. Esto se manifiesta en que estos átomos reportan factores de temperatura (B, factor de Debye-Weller) con una variabilidad muy grande, pudiendo ir desde valores semejantes a los átomos de las cadenas principales de la proteína, hasta valores significativamente mayores<sup>16</sup>.

Por último, mediciones muy recientes del corrimiento de Stokes dinámico de la fluorescencia de triptófano<sup>17</sup>, han detectado que el grado de movilidad de las moléculas de agua en sitios específicos de la superficie de una proteína puede ser prácticamente igual que la movilidad de una molécula de agua en el seno del disolvente, mientras que en otros lugares, dicha movilidad puede ser diez o hasta cien veces menor. Estos experimentos requieren que la proteína bajo estudio contenga un solo residuo de triptófano en la superficie de la macromolécula y de hecho los resultados sólo son relevantes para las moléculas de agua muy cercanas al triptófano. Debido a esto, el método sólo ha sido aplicado a unas cuantas proteínas y aun en estos casos, no se obtiene información sobre la dinámica de hidratación de la superficie en su conjunto, sino de tan solo una muy pequeña región de esta.

## 1.2. Objetivos generales y sistemas de estudio

Todos los ejemplos arriba mencionados resaltan la idea de que el agua en algunos sitios de hidratación particulares pueden realizar un papel altamente específico en algunas proteínas, ya sea que éstas tengan una importancia estructural<sup>18,19,29</sup> de interacción con otras moléculas<sup>21</sup>, o en la dinámica de la proteína<sup>22</sup>. Debido a esto, resulta crucial la identificación y caracterización por diversas metodologías de las moléculas de agua a las que se les pueda asignar dichas propiedades. Es importante recalcar que todas las técnicas experimentales disponibles a la fecha tienen limitaciones, en uno u otro sentido, que les imposibilita proveer información detallada a nivel atómico sobre la distribución de las propiedades dinámicas del agua en contacto directo con la proteína. Esta distribución se refiere al número y a la localización de moléculas de agua de elevado tiempo de contacto con la proteína y por otro lado, a la identificación de zonas en la superficie donde el intercambio es muy rápido. En esta tesis se realizó un estudio por simulaciones de la dinámica molecular con el objeto de caracterizar las propiedades dinámicas de la primera esfera de solvatación de diversas proteínas. En particular, esta caracterización indicó la distribución espacial de las moléculas de agua con diferentes tiempos de interacción con la proteína así como su posible rol como “puentes de agua”.

Las simulaciones de la dinámica molecular proporcionan una descripción a nivel atómico y en una escala de tiempo apropiada de los mecanismos de interacción agua-proteína y en particular para estudiar la velocidad de intercambio de moléculas de disolvente entre la superficie proteínica y el seno del líquido. El análisis de la solvatación de diversas proteínas se realizó mediante la evaluación de una función de correlación temporal y una función de correlación espacial del disolvente en la primera esfera de solvatación, donde dichas funciones fueron calculadas sobre trayectorias de dinámica molecular de la proteína en equilibrio en su estado nativo, simuladas con solvente explícito.



En primer lugar, la llamada “Función de Supervivencia del Solvente”  $P_{\alpha}(t)$ , es una función de correlación temporal que permite medir el número de moléculas de disolvente (agua) que permanecen en la esfera de solvatación de un átomo del soluto ( $\alpha$ ) durante al menos un tiempo  $t$ .

Esta función, que se presenta con más detalle en la sección de métodos, permite caracterizar el tiempo de residencia promedio de las moléculas de agua que solvatan cada átomo de la proteína, mediante el cálculo de los tiempos característicos del decaimiento de la misma. La función se evaluó para cada uno de los átomos de la superficie de diversas proteínas y los tiempos característicos resultantes fueron utilizados para “mapear” (directamente en la superficie proteínica), el grado de movilidad de la primera esfera de solvatación. Dicho mapeo se realizó con la ayuda de escalas de colores para obtener información gráfica que facilite el análisis y la ubicación de zonas en la superficie macromolecular con características peculiares respecto a la movilidad del disolvente.

En segundo lugar, se evaluó la función de distribución radial de dos centros  $g(r_1, r_2)$ . Esta función permite identificar de manera sencilla las moléculas de agua que durante la simulación, realizan con alta frecuencia, enlaces de hidrógeno simultáneos con dos átomos diferentes. Estos dos átomos pueden ser de la misma proteína o bien de dos proteínas que forman un complejo. Como se mencionó en párrafos anteriores, estas moléculas que forman los llamados “puentes de agua” (transitorios o semi-permanentes), pueden ser importantes, entre otras cosas, en la mediación de interacciones en una interfase de reconocimiento.

Se estudiaron las siguientes cuatro proteínas: FKBP12 (isomerasa peptidil-prolil), tendamistat (inhibidor de la  $\alpha$  amilasa), barnasa (ribonucleasa), barstar (inhibidor de la ribonucleasa) y el complejo barnasa-barstar. Las primeras dos proteínas han sido el objeto de investigaciones experimentales detalladas respecto a

las interacciones agua-proteína y fueron estudiadas para contrastar los resultados de la dinámica molecular con los resultados experimentales disponibles. La proteína FKBP12<sup>16</sup> fue el objeto de un minucioso estudio por parte de Faerman y Karplus donde se compararon las moléculas de agua detectadas por difracción de rayos X en seis diferentes estructuras cristalográficas de la misma proteína. Al ubicar moléculas de agua que aparecen haciendo contacto con la superficie de la FKBP12 en cada uno de los diferentes empaquetamientos cristalinos, les fue posible a dichos autores diferenciar a las moléculas de agua cuya presencia depende del empaquetamiento, de aquellas que aparecen en las estructuras debido únicamente a su interacción con la proteína. Este último grupo define el conjunto de “sitios de hidratación intrínsecos de la FKBP12”. En nuestro estudio se compararon los tiempos de residencia del solvente en la esfera de solvatación de cada átomo de la FKBP12 y fue posible concluir si existen algunas tendencias particulares respecto a los sitios reportados como sitios intrínsecos.

La proteína tendamistat fue objeto de estudio experimental por técnicas de atenuación específica de señales de RMN por co-solutos paramagnéticos<sup>23</sup>. Dichos estudios pretenden ubicar experimentalmente los sitios superficiales de elevada interacción agua-proteína gracias a la identificación de residuos de la proteína, los cuales a pesar de estar expuestos al solvente, no tienen contacto con el co-soluto paramagnético; debido precisamente a que las moléculas de agua fuertemente unidas a la superficie, bloquean el acercamiento del co-soluto. En este caso fue posible comparar los sitios que la dinámica molecular predice como de alto tiempo de residencia con los ubicados por las técnicas de atenuación de RMN.

Por último, las proteínas barnasa, barstar y su complejo 1:1 son un sistema muy interesante ya que en el complejo la unión proteína-proteína es mediada por múltiples puentes de agua. El objetivo de estudiar tanto al complejo como a las proteínas individuales es el de observar si las aguas de puente se pueden ubicar como

sitios de hidratación de alto tiempo de residencia, donde algunas moléculas de agua se mantienen asociadas a una u otra proteína desde antes de que se forme el complejo o bien, si por otro lado, es imposible distinguir los sitios que eventualmente formarán puentes de agua en las proteínas individuales.

### **1.3. Objetivos específicos**

- i) Elaborar un programa de la función de supervivencia con el lenguaje de programación C en la plataforma de Linux para analizar la distribución de las moléculas de agua en torno a la proteína.
- ii) Automatizar la ejecución del programa de la función de supervivencia mediante una serie de instrucciones escritas en lenguaje bash (scripts) en la plataforma de Linux.
- iii) Delimitar las zonas de la superficie de las proteínas por medio de colores de acuerdo al tiempo de residencia de las moléculas de agua en torno a los átomos de la proteína (construcción de mapas de colores).
- iv) Demarcar las zonas de la superficie de la proteína a través de colores de acuerdo al número de moléculas de aguas coordinadas a cada átomo de la proteína.
- v) Construir un programa con la función de distribución radial de dos centros con el lenguaje de programación C, en la plataforma de Linux para identificar moléculas de agua que forman enlaces puente de hidrógenos simultáneos, con átomos de la proteína.
- vi) Confrontar los sitios de hidratación de la proteína FKBP12 identificados mediante análisis de la dinámica molecular, con los reportados por Faerman y Karplus en su estudio de consensos de sitios de hidratación por cristalografía de rayos X.
- vii) Comparar la distribución y localización de las moléculas de agua de alto tiempo de residencia, obtenidas del análisis de la simulación de DM de la proteína tendamistat, con los reportados por Scarselli y colaboradores en su estudio

experimental por técnicas de atenuación específica de señales de RMN por co-solutos paramagnéticos.

- viii) Identificar las moléculas de agua que median la interacción del complejo barnasa-barstar y compararlas con aquellas que aparecen en la zona de interacción de las mismas proteínas no acomplexadas, esto con la finalidad de examinar la participación de las moléculas de agua en el reconocimiento proteína-proteína.
- ix) Compartir las herramientas diseñadas durante el desarrollo de este trabajo con otros grupos de investigación, a través de una página web, ya que la metodología es de aplicabilidad generalizada y los programas son suficientemente robustos para ser difundidos en la comunidad científica.

## 2. Antecedentes

En este capítulo se revisan brevemente algunos temas relacionados con el proyecto, incluyendo aspectos básicos y técnicos de las simulaciones de dinámica molecular<sup>24,25</sup>. También se presentan las funciones de correlación espacial y temporal que se emplearon para el análisis, y se realiza una breve presentación de las proteínas de estudio.

### 2.1. Simulaciones de dinámica molecular

En los últimos años el incremento significativo del poder computacional junto con la disponibilidad de métodos teóricos incluidos en sofisticados programas de simulación, han permitido expandir la aplicabilidad de la aproximación computacional a los sistemas biológicos. La expansión es de tal dimensión que en el campo de la bioquímica, las herramientas computacionales son utilizadas frecuentemente para dilucidar la estructura secundaria de las proteínas, para hacer modelados de macromoléculas, para estudiar el modo y la forma como se acoplan proteína-proteína o proteína-ligando (docking) y para estudiar el comportamiento a nivel atómico de las proteínas o biomoléculas. Actualmente las simulaciones de dinámica molecular (DM) representan una herramienta valiosa para realizar estudios que van más allá de la accesibilidad experimental, pues los métodos prácticos que hasta el momento se han desarrollado para investigar las propiedades dinámicas de las proteínas, suelen proporcionar información con pocos detalles sobre el nivel atómico de las fluctuaciones (en nuestro caso, de la esfera de solvatación)<sup>26</sup>. Por lo tanto, las simulaciones de DM representan una valiosa herramienta complementaria a las técnicas experimentales.

La dinámica molecular tiene sus bases en la segunda ley de Newton o en la ecuación del movimiento<sup>27,28</sup>.

$$F_i = m_i a_i \quad (1)$$

donde  $F_i$  es la fuerza ejercida en la partícula  $i$ ,  $m_i$  es la masa de la partícula  $i$  y  $a_i$  es la aceleración de la partícula  $i$ . La fuerza también puede ser expresada en términos del gradiente de la energía potencial:

$$F_i = -\nabla_i V \quad (2)$$

### 2.1.1 Campos de fuerza

Para calcular la energía potencial del sistema es indispensable definir los campos de fuerza que rigen el movimiento atómico, los cuales están determinados por las contribuciones energéticas tanto de átomos unidos covalentemente (estiramiento del enlace, deformación de ángulos y deformación fuera del plano) como de átomos no unidos covalentemente (fuerzas de Van der Waals e interacciones electrostáticas); de acuerdo a esto la variación energética del sistema va a depender de las desviaciones de estas contribuciones con respecto a sus valores de equilibrio. Considerando las contribuciones de átomos enlazados y de átomos no enlazados, la energía potencial se puede expresar como:

$$V(R) = E_{enlace} + E_{no-enlace} \quad (3)$$

$$E_{enlace} = E_{estiramiento} + E_{deformación} + E_{torsión} \quad (4)$$

$$E_{no-enlace} = E_{van-der-Waals} + E_{electrostática} \quad (5)$$

Cada uno de los términos de la ecuación (3) definen el campo de fuerza de la dinámica molecular, existen varios campos de fuerza como son: AMBER, CHARMM, GROMOS, OPLS/AMBER, entre otros. Éstos tienen en común cada uno de los términos de la ecuación 3; pero difieren en los términos que se le añadan o en las constantes que se empleen (los parámetros). Las dinámicas moleculares realizadas para la elaboración de esta tesis se efectuaron con el programa de NAMD, el cual

trabaja con los campos de fuerza y los parámetros de CHARMM, del cual a continuación se hace una breve descripción.

### **Alargamiento del enlace**

El primer término de la ecuación (4) considera el alargamiento del enlace el cual implica átomos separados por un enlace covalente (relación 1,2). Éste está representado por la ecuación del oscilador armónico:

$$E_{\text{estiramiento}} = \sum_{\text{enlace}} \frac{K_b}{2} (b - b_0)^2 \quad (6)$$

Esta ecuación es una representación de la energía potencial en función del desplazamiento, donde  $K_b$  es la constante que determina la fuerza del enlace, y  $b_0$  es la distancia entre dos núcleos atómicos en la configuración de mínima energía. Tanto  $K_b$  como  $b_0$  son específicos para cada par de átomos enlazados.

### **Deformación del ángulo**

El segundo término de la ecuación (4) se refiere a la deformación del ángulo  $\theta$  con respecto al valor ideal  $\theta_0$  (relación 1,3), también está representado por la ecuación del oscilador armónico:

$$E_{\text{deformación}} = \sum_{\text{ángulo}} \frac{K_\theta}{2} (\theta - \theta_0)^2 \quad (7)$$

Los valores de  $K_\theta$  y de  $\theta$  dependen del tipo de átomo involucrado.

### **Término de Urey-Bradley**

Este término corresponde a la interacción entre dos átomos separados por dos enlaces (interacción 1,3), éste es una variante del campo de fuerza de CHARMM con

respecto al campo de fuerza general de las simulaciones de dinámica molecular y se representa con la ecuación del oscilador armónico:

$$E_{UB} = \sum_{UB} \frac{K_{UB}}{2} (S - S_0)^2 \quad (8)$$

donde  $K_{UB}$  es la constante de Urey-Bradley y  $S$  es la distancia entre los átomos 1 y 3 (distancia de Urey-Bradley).

### Torsión

En el término de  $E_{torsión}$  de la ecuación (4) se incluyen tanto el ángulo diedro como el ángulo diedro impropio.

$$E_{torsión} = \sum_{diedro} \frac{V_n}{2} [1 + \cos(nw - \gamma)] + \sum_{impropio} \frac{K_{impropio}}{2} (\varphi - \varphi_0)^2 \quad (9)$$

El primer término de la ecuación (9), representa el cambio de la energía asociada a la rotación de los ángulos diedro, los cuales se forman entre tres enlaces covalentes (relaciones 1,4). Este potencial tiene un comportamiento periódico y se puede representar mediante una serie de Fourier, donde  $V_n$  determina el tamaño de la barrera rotacional,  $n$  indica la periodicidad,  $w$  es el ángulo diedro y  $\gamma$  hace referencia al ángulo de torsión cuando se encuentra en su mínima energía. El segundo término de la ecuación (9) corresponde al ángulo diedro impropio, representado por la función armónica, donde  $K_{impropio}$  es la constante de fuerza del ángulo impropio y  $\varphi$  es el ángulo de torsión impropio. Este término gobierna la quiralidad y la planaridad de la molécula<sup>29</sup>.



## Las interacciones de van der Waals

Las interacciones de van der Waals describen la repulsión o atracción entre dos átomos que no están unidos covalentemente, con frecuencia estas fuerzas se modelan utilizando el potencial de Lennard-Jones:

$$E_{\text{van-der-Waals}} = \sum \varepsilon_{ij} \left[ \left( \frac{R_{ij}}{r_{ij}} \right)^{12} - 2 \left( \frac{R_{ij}}{r_{ij}} \right)^6 \right] \quad (10)$$

donde  $R_{ij}$  es la distancia de Lennard-Jones de mínima energía, el parámetro  $\varepsilon_{ij}$  es el valor mínimo en la curva de interacción y  $r_{ij}$  es la distancia entre los dos átomos ( $i, j$ ). Las fuerzas atractivas son de largo alcance (mucho menos que las fuerzas electrostáticas) y se deben a las fuerzas de dispersión de London (dipolos instantáneos) y las fuerzas repulsivas son de corto alcance y se deben a la repulsión de las nubes electrónicas.

## Interacción electrostática

La interacción electrostática, interacción de no enlace, es una consecuencia de la distribución interna de los electrones; donde los átomos de las moléculas tienden a adquirir carga. La interacción electrostática entre un par de átomos es representada por el potencial de Coulomb:

$$E_{\text{electrostática}} = \frac{q_i q_j}{\varepsilon_D r_{ij}} \quad (11)$$

donde  $\varepsilon_D$ , es la constante dieléctrica del medio y  $r_{ij}$  es la distancia entre dos átomos de cargas  $q_i$  y  $q_j$ .

## Condiciones periódicas de frontera

Las condiciones periódicas de frontera permiten que la simulación de DM se ejecute utilizando un número relativamente pequeño de partículas y de tal forma que éstas experimenten fuerzas como si estuvieran en el seno de la solución. Un sistema con condiciones periódicas es semejante a una caja central con algunos átomos, la cual es rodeada por otras cajas semejantes a la central. Estas condiciones son importantes en el momento en que una partícula salga de la caja durante la simulación, en cuyo caso la partícula es reemplazada por otra partícula que debe entrar por el lado opuesto de la caja. Adicionalmente, estos métodos permiten el cálculo de interacciones (de no enlace) con las múltiples réplicas de las partículas de la caja principal. En un sistema con condiciones periódicas es posible que las interacciones de largo alcance sean tratadas mediante métodos basados en las sumas de Ewald<sup>30</sup>. El programa NAMD utiliza el método de PME (particle-mesh Ewald) para agilizar el cálculo de las sumas de Ewald. Las sumas de Ewald involucran las siguientes expresiones:

$$E_{Ewald} = E_{dir} + E_{rec} + E_{self} + E_{surface} \quad (12)$$

$$E_{dir} = \frac{1}{2} \sum_{i,j=1}^N q_i q_j \sum_{\vec{n}_r} \frac{\text{erfc}(\beta |\vec{r}_i - \vec{r}_j + \vec{n}_r|)}{|\vec{r}_i - \vec{r}_j + \vec{n}_r|} - \sum_{(i,j) \in \text{Excluded}} \frac{q_i q_j}{|\vec{r}_i - \vec{r}_j + \vec{v}_{ij}|} \quad (13)$$

$$E_{rec} = \frac{1}{2\pi V} \sum_{\vec{m} \neq 0} \frac{\exp(-\pi^2 |\vec{m}|^2 / \beta^2)}{\vec{m}^2} \left| \sum_{i=1}^N q_i \exp(2\pi i \vec{m} \times r_i) \right|^2 \quad (14)$$

$$E_{self} = -\frac{\beta}{\sqrt{\pi}} \sum_{i=1}^N q_i^2 \quad (15)$$

$$E_{surface} = \frac{2\pi}{(2\epsilon_s + 1)V} \left| \sum_{i=1}^N q_i \vec{r}_i \right|^2 \quad (16)$$

donde  $E_{dir}$  sumas directas, considera un sistema periódico embebido en un material con una constante dieléctrica infinita,  $E_{rec}$  sumas recíprocas, representa el potencial electrostático en un punto, debido a la distribución de carga,  $E_{self}$  es una constante de corrección y  $E_{surface}$ , término de superficie, es normalmente ignorado al considerar las condiciones periódicas, debido en parte, a que la constante dieléctrica de el agua ( $\epsilon_s \approx 80$ ) es mucho mayor que uno. En las ecuaciones de (12-16)  $q_i$  representa la carga del átomo  $i$ ,  $\vec{r}_i$  es la posición del átomo  $i$ ,  $\vec{n}_r$  es el vector relacionado con la caja de simulación,  $\vec{v}_{ij}$  denota el vector para el par  $(i, j)$  que minimiza  $|\vec{r}_i - \vec{r}_j + \vec{v}_{ij}|$ ,  $\beta$  es un parámetro que distribuye la carga computacional entre los términos de las sumas directas y sumas recíprocas.  $\epsilon_s$  es la constante dieléctrica de los alrededores,  $V$  es el volumen de la caja de simulación y  $\vec{m}$  es el vector recíproco. En el método de PME (particle-mesh Ewald)  $\beta$  se elige, de tal forma que la mayor carga computacional se realice sobre las sumas recíprocas  $E_{rec}$ , las cuales se resuelven mediante la transformada de Fourier. El método PME, distribuye la carga en una rejilla de puntos de acuerdo a la distancia de estos puntos a la carga, empleando para ello funciones de peso.

### 2.1.2 Integración de las ecuaciones de movimiento

Como la energía potencial es una función de las posiciones atómicas de todos los átomos del sistema, es imposible encontrar una solución analítica para la ecuación de movimiento. La función de movimiento se puede resolver numéricamente por medio de algoritmos de integración; los cuales asumen que las posiciones, velocidades y aceleraciones pueden ser aproximadas por una serie de expansión de Taylor. Hay varios algoritmos para resolver dicha aproximación, pero uno de los más empleados es el algoritmo de Verlet, el cual usa las posiciones y aceleraciones en el tiempo  $t$  y las posiciones para el tiempo  $t - \Delta t$  para calcular las nuevas posiciones en el tiempo

$t + \Delta t$ ; en este algoritmo no están explícitas las velocidades. El esquema de integración de Verlet se basa en una expansión de Taylor de tercer orden:

$$\bar{r}_i(t + \Delta t) = 2\bar{r}_i(t) - \bar{r}_i(t - \Delta t) + \Delta t^2 \frac{d^2 \bar{r}_i(t)}{dt^2} \quad (17)$$

La integración de las ecuaciones de movimiento produce una trayectoria que describe las posiciones, velocidades y aceleraciones de las partículas conforme ellas cambian con el tiempo. Si estos datos se conocen para cada uno de los átomos en un determinado instante, entonces el estado del sistema se puede predecir en algún tiempo, futuro o pasado. Debido a que las simulaciones de este proyecto se realizaron considerando la dinámica de Langevin en lugar de la Newtoniana, el método de integración no es el de Verlet simple, si no una sencilla extensión del mismo realizada por Brügger Brooks y Karplus donde la fórmula recursiva de la posición de los átomos toma en cuenta los términos adicionales de la ecuación de Langevin<sup>31</sup> (ver a continuación).

### 2.1.3 Preparación y corrida de las simulaciones de dinámica molecular

La mayoría de las simulaciones de dinámica molecular son ejecutadas en un ensamble microcanónico, es decir bajo condiciones de número de partículas (N), volumen (V) y energía (E) constante. Sin embargo, es frecuente usar otros ensambles tales como: el canónico (NVT) o el isotérmico-isobárico (NPT). En este trabajo se realizaron las simulaciones de dinámica molecular manteniendo el número de moléculas, la presión y la temperatura constante, es decir un ensamble (NPT). El programa NAMD para un ensamble NPT, considera la ecuación de Langevin:

$$m_i \frac{d^2 x_i(t)}{dt^2} = F_i \{x_i(t)\} - \gamma_i \frac{dx_i(t)}{dt} m_i + R_i(t) \quad (18)$$

En la ecuación de Langevin, la fuerza sobre la partícula se divide en tres términos: El primer término  $F_i \{x_i(t)\}$  representa la fuerza, debida a las interacciones entre los

átomos del sistema; este término es idéntico al de la dinámica molecular Newtoniana.

El segundo término  $\gamma_i \frac{dx_i(t)}{dt} m$  es la fuerza de fricción, la cual se debe a la resistencia

de la partícula por efecto del medio, en este término  $\gamma_i$  es el coeficiente de fricción. El

tercer término  $R_i(t)$  corresponde a las colisiones al azar entre las moléculas del

solvente y el soluto (fuerza al azar o fuerza estocástica). Los dos últimos términos son

propios de la dinámica molecular de Langevin, básicamente representan los efectos

del medio y se pueden emplear para mantener constante la temperatura del sistema.

Al integrar la ecuación de Langevin por el método de Brünger-Brooks-Karplus (BBK) se obtiene la siguiente expresión:

$$r_{n+1} = r_n + \frac{1 - \gamma\Delta t/2}{1 + \gamma\Delta t/2} (r_n - r_{n-1}) + \frac{1}{1 + \gamma\Delta t/2} \Delta t^2 \left[ M^{-1} F(r_n) + \sqrt{\frac{2\gamma k_B T}{\Delta M}} Z_n \right] \quad (19)$$

donde  $M$  es la masa,  $F$  es la fuerza,  $r$  la posición,  $\gamma$  es el coeficiente de fricción,  $k_B$

es la constante de Boltzman,  $T$  es la temperatura y  $Z_n$  es un juego de variables al

azar de una gaussiana de punto cero y una varianza de uno.

Después de establecer las condiciones en las cuales se ejecuta el sistema, de

definir la ecuación de movimiento para la ejecución de la dinámica molecular y de fijar

el método de integración para resolver la ecuación de movimiento, damos paso a una

breve descripción de la ejecución de la dinámica molecular, la cual comprende las

siguientes etapas: inicialización, minimización, calentamiento, equilibrio y desarrollo de

la dinámica molecular, a continuación se describen cada una de estas etapas.

### **Inicialización**

La ecuación de movimiento de Langevin requiere los valores iniciales, tanto de las

coordenadas como de las velocidades para dar inicio a la integración. Las

coordenadas iniciales son usualmente obtenidas de estructuras moleculares

determinadas principalmente por experimentos de rayos X o de RMN, y las velocidades iniciales se asignan al azar de acuerdo a la distribución de Maxwell-Boltzman a una temperatura T; normalmente estas asignaciones se hacen a baja temperatura para posteriormente incrementar la temperatura del sistema.

### **Minimización, calentamiento y equilibrio del sistema**

La minimización se encarga de reducir la energía potencial de una conformación, mitigando sus tensiones locales. Durante la minimización las velocidades atómicas se ajustan a cero y como consecuencia de esto los movimientos de todos los átomos se limitan, de esta forma se evitan los grandes desplazamientos de la estructura, con respecto a sus coordenadas iniciales. Una vez se conocen las coordenadas y las velocidades iniciales del sistema, empieza el calentamiento, para lo cual, primero se asignan las velocidades a bajas temperaturas, posteriormente se hacen ligeros incrementos de la temperatura y para cada una de ellas se van asignando las nuevas velocidades, este proceso se repite hasta alcanzar la temperatura deseada. Después del calentamiento continúa la etapa del equilibrio, que es una simulación de dinámica molecular, cuyo objetivo es asegurar que el sistema se encuentre en equilibrio a la temperatura y presión de trabajo, y libre de fluctuaciones erráticas. Durante el equilibrio varias propiedades son monitoreadas; en particular la estructura, presión, temperatura y energía del sistema.

### **Desarrollo de la simulación de dinámica molecular**

Durante esta etapa se ejecuta la simulación del sistema en un periodo de tiempo dado, en el transcurso de este tiempo se van registrando las diferentes configuraciones en un número significativo de pasos  $\Delta t$ , es decir cada vez que transcurre una fracción de tiempo  $\Delta t$ , se integra la ecuación de movimiento y los valores de las posiciones y de las velocidades se van guardando. En esta fase ya no es necesario hacer el escalonamiento de las velocidades, debido a que el método de integración no las

requiere. La simulación de dinámica molecular arroja información acerca de las propiedades conformacionales de los sistemas y la manera como la estructura cambia con el tiempo.

## 2.2 Análisis de la simulación de dinámica molecular

### 2.2.1 Función de supervivencia del solvente en la esfera de solvatación de un átomo $\alpha$ de una proteína $P_\alpha(t)$

Para el análisis de la simulación de DM se empleó la función de supervivencia definida originalmente por Impey y colaboradores<sup>32</sup>, quienes inicialmente la emplearon para describir la movilidad de la esfera de solvatación de iones en solución. Hay varias investigaciones en las que se ha reportado el uso de esta función para el estudio de los sistemas agua-soluto<sup>33,34,35</sup>. En nuestro caso, la utilizamos para caracterizar la movilidad local de la esfera de solvatación de cada átomo  $\alpha$  de una proteína, el rótulo átomo  $\alpha$  se refiere al átomo que se desea estudiar y es independiente de la nomenclatura que se emplea para nombrar a los átomos de la cadena lateral de un aminoácido. Esta función se define con la siguiente ecuación:

$$P_\alpha(m) = \sum_{j=1}^{j=NW} \frac{1}{NC-m} \sum_{n=1}^{n=NC-m} P(n, n+m) \quad (20)$$

donde  $\alpha$  es un átomo en particular de la proteína,  $NW$  es el número de aguas en la caja de simulación,  $m$  es un índice de configuración tal que el tiempo al que está evaluada la función  $P_\alpha$  es  $t = m\Delta t$  (donde  $\Delta t$  es el intervalo entre dos configuraciones consecutivas); y donde  $NC$  es el total de configuraciones. La función  $P$  toma los valores de 1 cuando la molécula de agua  $j$  (su átomo de oxígeno) reside en la esfera del átomo  $\alpha$  de manera continua en el intervalo de configuraciones  $n$  a  $n+m$  y de cero cuando esto no se obedece.  $P_\alpha$  es una función de correlación de tiempo de residencia que indica el promedio de moléculas de agua que permanecen en la esfera

de coordinación por al menos un tiempo  $t$ . Cuando  $t = 0$  ( $m = 0$ ),  $P_\alpha$  representa el número promedio de moléculas de agua que residen en la esfera de solvatación del átomo, es decir, es un promedio del número de aguas coordinadas al átomo  $\alpha$  en cualquier instante dado. Como cualquier función de correlación temporal,  $P_\alpha(t)$  considera cada una de las configuraciones de la simulación y evalúa la probabilidad de permanencia de las moléculas de agua en la esfera de solvatación del átomo  $\alpha$  a partir de este momento.

La primera esfera de solvatación alrededor de un átomo de una proteína se definió como una esfera de radio de 4 Å, de acuerdo a los estudios realizados por Gu y Schoenborn con la función de distribución radial unidimensional en la proteína mioglobina<sup>36</sup>.

A modo de ejemplo, en la figura 2.1 se muestra el resultado del cálculo de la función  $P_\alpha(t)$ , para el átomo de oxígeno del carbonilo del residuo 103 de la tirosina de la proteína barnasa. Como se puede observar, el valor al tiempo cero de esta función es de 1.6 (número de coordinación promedio de este átomo), y debido al intercambio con las moléculas del seno de la solución, la función decae de forma aproximadamente exponencial como función del tiempo. En este caso, la función vale cero a partir de los 250 ps. Es decir, que en ningún momento de la simulación existió alguna molécula de agua que permaneciera solvatando a este átomo por más de 250 ps. También se puede apreciar que el tiempo de vida media de permanencia del grupo de aguas de solvatación es aproximadamente de tan solo unos 20 ps. Los tiempos característicos del decaimiento de la función  $P_\alpha(t)$  se conocen como tiempos de residencia de agua en la esfera de solvatación del átomo  $\alpha$ .



## 2.2.2 Función de distribución radial de dos centros

Con el motivo de complementar la información proporcionada por la función de supervivencia, se utilizó la función de distribución radial de dos centros, la cual suministra información sobre moléculas de agua que forman enlaces de tipo puentes de hidrógeno simultáneos a dos átomos de la proteína, con la particular ventaja de que no considera ningún centro de referencia, ya que la proteína está en continuo movimiento, sino que identifica las moléculas de agua con respecto a dos átomos de la proteína previamente seleccionados, lo cual permite la identificación de los llamados puentes de agua. Es importante notar que debido a las fluctuaciones de las interacciones agua-proteína, la inspección visual directa de la trayectoria de la proteína solvatada no hace posible la identificación de los puentes de agua transitorios, si no que se requiere un método que cuantifique estadísticamente la formación de puentes de hidrógeno simultáneos de una misma agua hacia dos átomos diferentes de una proteína o de un complejo.

Para explicar la función de distribución radial de dos centros, primero se debe traer a colación la función de distribución radial unidimensional definida por la ecuación:

$$g_i(r, dr) \equiv \frac{\rho_i(r, dr)}{\rho_{gas\ ideal}} \quad (21)$$

donde  $g_i(r, dr)$  es la función de distribución radial de las moléculas de agua alrededor de un átomo de la proteína en la configuración  $i$  entre el radio  $r$  y  $r + dr$ ;  $\rho_i(r, dr)$  y  $\rho_{gas\ ideal}$  se definen en términos del átomo de oxígeno del agua por medio de

$$\rho_i(r, dr) \equiv \frac{\text{número de oxígenos entre } r \text{ y } r + dr}{\text{Volumen}_{r, dr}} \quad (22)$$

$$\rho_{gas\ ideal} = \frac{\text{número de oxígenos totales}}{\text{Volumen de simulación total}} \quad (23)$$

La función de distribución radial de dos centros considera el número de moléculas de agua que existen simultáneamente entre el radio  $r_1$  y  $r_1 + dr_1$  del átomo A y el número de moléculas de agua entre el radio  $r_2$  y  $r_2 + dr_2$  del átomo B de la configuración  $i$  y se expresa por medio de la ecuación:

$$\rho(r_1, dr_1, r_2, dr_2) = \frac{\text{número de oxígenos de H}_2\text{O entre } r_1 \text{ y } r_1 + dr_1 \text{ y entre } r_2 \text{ y } r_2 + dr_2}{\text{Volumen de intersección de los elementos de volumen}} \quad (24)$$

donde  $\rho(r_1, dr_1, r_2, dr_2)$  representa la densidad en el volumen de intersección entre los átomos A y B. El volumen de intersección se refiere al volumen ocupado simultáneamente por los segmentos de esfera alrededor de los dos átomos. En la figura 2.2 se presenta un esquema de cómo los dos segmentos de esfera delimitan este volumen de intersección (nótese que dicho volumen depende tanto de  $r_1$ , como de  $r_2$  y de la distancia entre los dos centros: R, dependiendo de estos valores, la fórmula para el cálculo del volumen de intersección cambia, existiendo 10 diferentes casos para el cálculo del volumen de intersección considerados en el programa respectivo).

A manera de introducción, en las figuras 2.3 y 2.4 se muestran dos ejemplos del resultado de la evaluación de la función  $\rho(r_1, dr_1, r_2, dr_2)$ , considerando el promedio de todas las configuraciones de la simulación. En la figura 2.3 se calcula la densidad del agua como función de la posición relativa (simultánea) de los dos oxígenos del grupo carboxílico del ácido glutámico (residuo 8) de la proteína barstar. Este residuo se encuentra en la superficie de la proteína con la cadena lateral totalmente proyectada hacia el seno del líquido. La función  $\rho(r_1, dr_1, r_2, dr_2)$  se muestra como una

gráfica de curvas de nivel donde la densidad se grafica en una escala de tonos de gris como función de la distancia al átomo 1 (Oxígeno OE1) y de la distancia al átomo 2 (Oxígeno OE2). Como se puede apreciar, en el caso de este grupo carboxílico existen dos zonas con elevada densidad de agua (del átomo de oxígeno de la misma). Ambas se ubican a aproximadamente 2.75 Å de uno de los átomos y a aproximadamente 4.8 Å del otro (nótese que fuera de estos dos puntos del diagrama, la densidad del agua se aproxima a su valor promedio en el seno del líquido, esto es: 0.0338 Aguas/Å<sup>3</sup>) para el modelo TIP3P a 1 atm y a 300 K ).

En el segundo ejemplo (figura 2.4), se muestra el diagrama de  $\rho(r_1, dr_1, r_2, dr_2)$  para dos átomos del complejo barnasa-barstar donde el átomo asociado al eje vertical es el oxígeno carbonílico de la Val45 de barstar, y el eje horizontal se refiere al nitrógeno de la cadena principal de la Ser38 de la barnasa. Estos dos átomos se encuentran precisamente en la interfase del complejo y en la gráfica se puede apreciar claramente que ambos átomos coordinan simultáneamente a una misma molécula de agua, ya que para ambos átomos, la distancia al máximo de densidad corresponde a una distancia típica de un puente de hidrógeno. En comparación, para el caso anterior del grupo carboxílico, es claro que los dos átomos de oxígeno no coordinan simultáneamente a una misma molécula de agua sino que se crea una red local de puentes de hidrógeno donde cada oxígeno se coordina con un agua diferente.

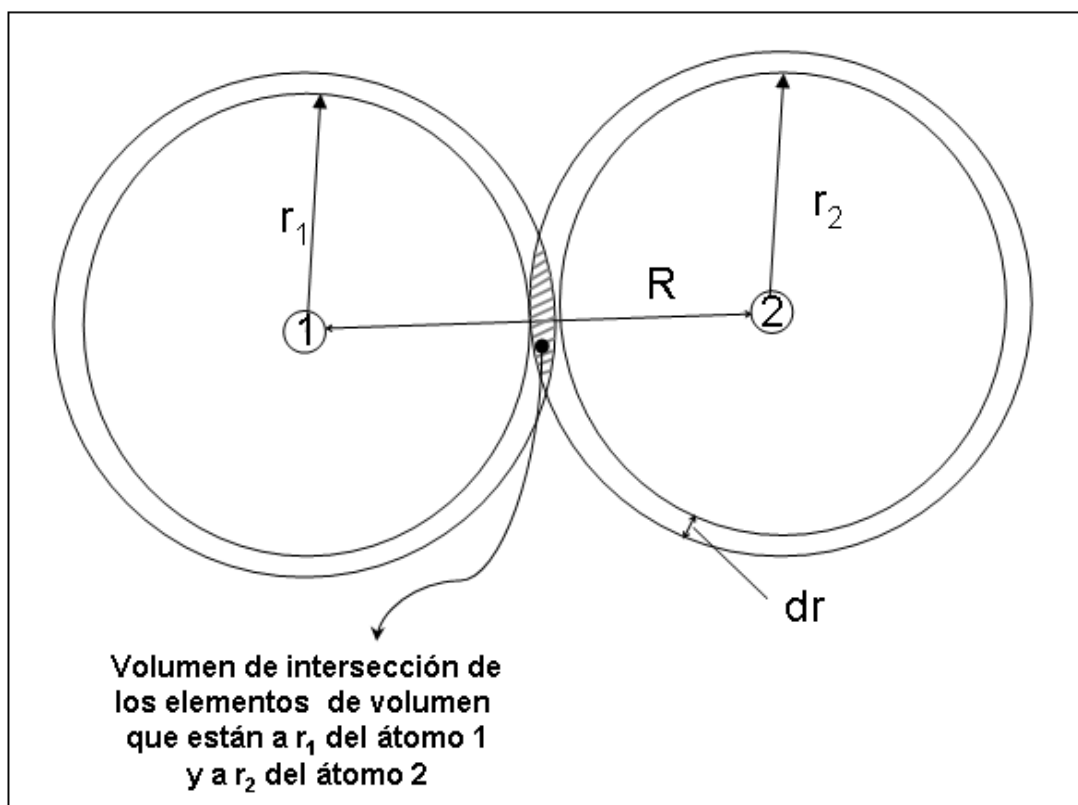


Figura 2.2. Representación de la intersección de las esferas de solvatación de dos átomos de una proteína o de proteínas diferentes. La zona delimitada mediante líneas indica el volumen de intersección.  $R$  es la distancia entre los dos átomos,  $r_1$  y  $r_2$  el radio de cada esfera y  $dr$  es la diferencia entre las dos esferas de un mismo átomo.

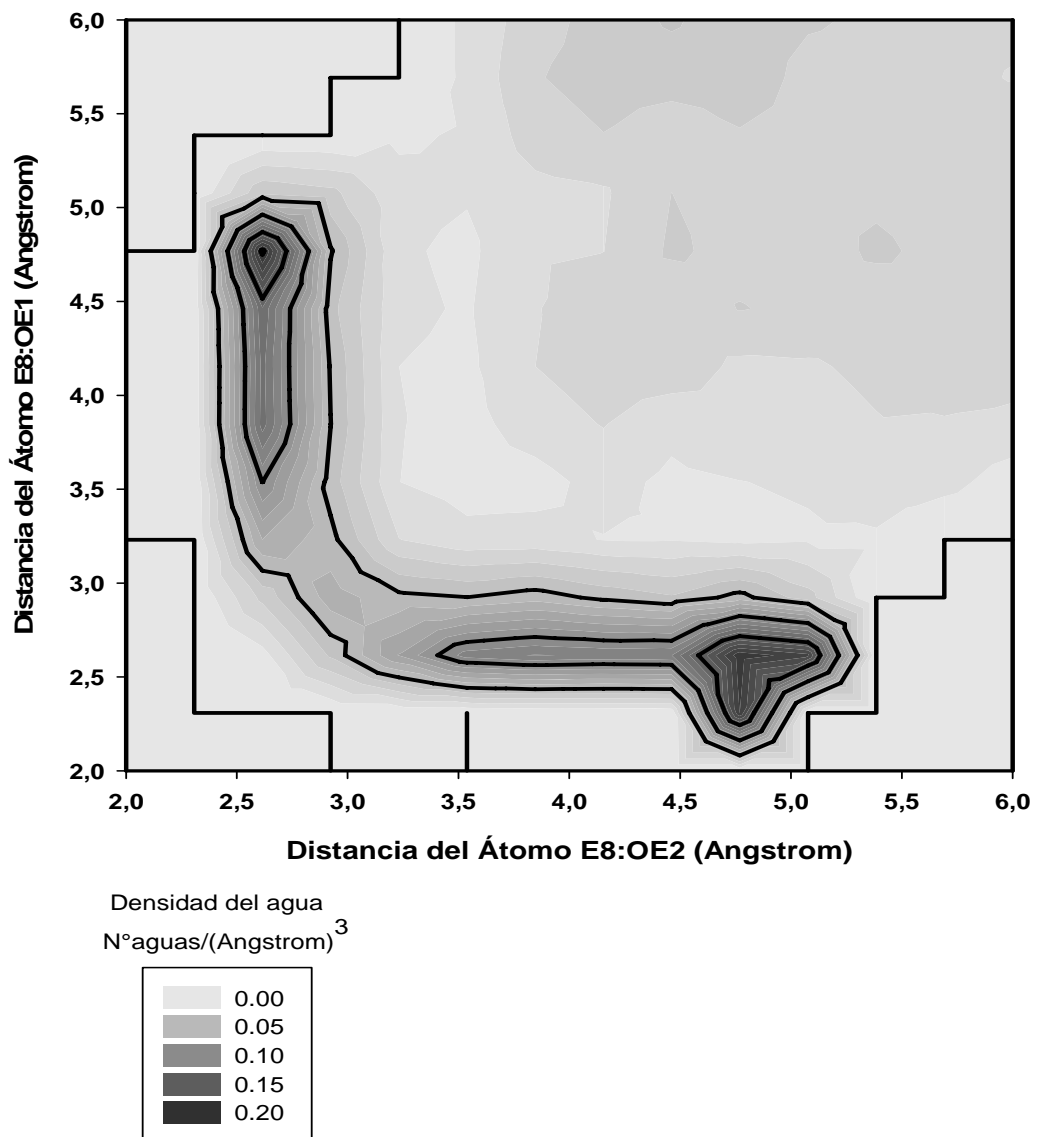


Figura 2.3. Evaluación de la función radial de dos centros sobre los dos oxígenos del grupo carboxílico del ácido glutámico (residuo 8) de la proteína barstar, donde cada oxígeno presenta una zona de alta densidad de agua (zona oscura), las cuales se ubican a  $2.75 \text{ \AA}$  de uno de los átomos y a aproximadamente  $4.8 \text{ \AA}$  del otro.

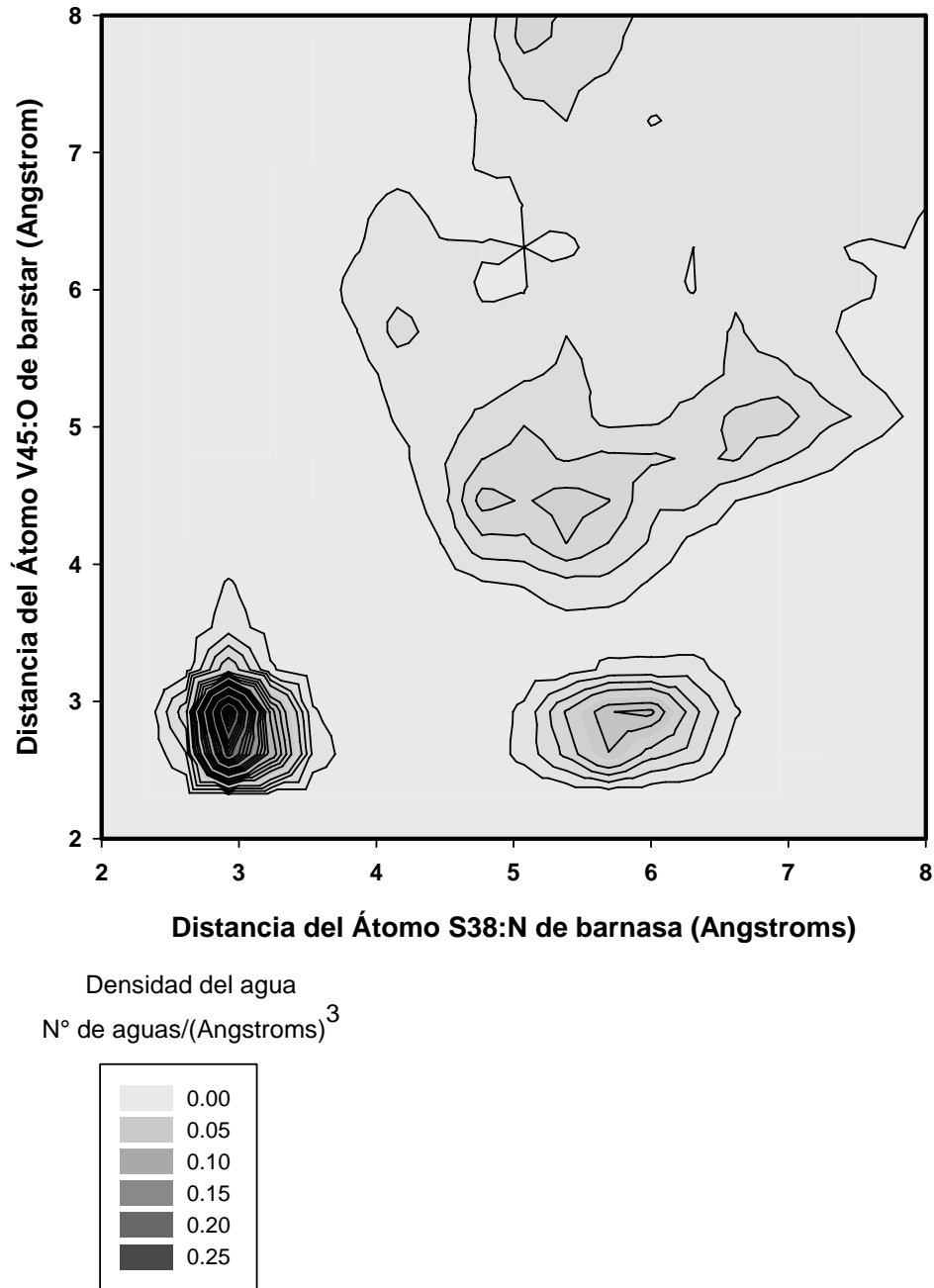


Figura 2.4. Evaluación de la función radial de dos centros para el nitrógeno de la serina 38 de la cadena principal de la proteína barnasa y el oxígeno del grupo carbonilo de la valina 45 de la proteína barstar. La zona de alta densidad (zona oscura) es un indicador de que ambos átomos coordinan simultáneamente a una misma molécula de agua, pues la distancia de la molécula de agua con respecto a cada uno de los átomos, corresponde a una distancia típica de un puente de hidrógeno.

### 2.3 Proteínas objeto del estudio

Para el desarrollo de este trabajo se emplearon las proteínas barnasa, barstar, la tendamistat y la FKBP12. El complejo barnasa-barstar es un buen sistema de estudio debido a la influencia del agua en la interacción proteína-proteína. De hecho aparece en la literatura que el grado de complementariedad  $Sc^{37}$  entre barnasa-barstar sin agua es de 0.7 y con agua es de 0.82, mientras que en los casos más comunes del Sc sin agua aparecen en el rango de 0.70-0.76 para proteasa-inhibidor y de 0.64-0.68 para el antígeno-anticuerpo, lo que resalta el importante papel del agua en la interacción barnasa-barstar. También, se estudiaron las proteínas tendamistat y FKBP12 con el objetivo de comparar los resultados, porque de la primera existe evidencia experimental por RMN<sup>23</sup> y de la segunda existe evidencia experimental tanto por difracción de rayos X como por RMN<sup>16,38</sup>.

#### 2.3.1 FKBP12

La FKBP12 es una isomerasa peptidil-prolil de 107 residuos que se enlaza fuertemente a los inmunosupresores FK506 y rapamicina, originándose la inhibición de la actividad rotamasa y por ende a un bloqueo en la activación y proliferación de las células T. Esta enzima cataliza la isomerización cis-trans (rotamerización) de los enlaces amido del peptidil-prolil<sup>39,40</sup>. La estructura secundaria de la FKBP12 está caracterizada por 5 hojas  $\beta$  antiparalelas y una hélice  $\alpha^{41}$  (ver figura 2.5). Esta proteína es secretada por el ser humano y su sitio activo lo constituyen 13 aminoácidos<sup>42</sup> (Tyr26, Phe36, Asp37, Arg42, Phe46, Phe48, Asn53, Glu54, Ile56, Trp59, Tyr82, His87, Phe99).

#### 2.3.2 Tendamistat

La tendamistat es una proteína de 74 aminoácidos secretada por *Streptomyces tendae*. Esta proteína puede inhibir a la  $\alpha$ -amilasa de mamífero, lo cual ha sido de

gran interés en el tratamiento terapéutico de la diabetes millitus<sup>43,44</sup>. Los residuos 11-73 de la tendamistat están plegados en una estructura  $\beta$  de seis hojas (ver figura 2.6). Su sitio activo lo conforman 15 aminoácidos distribuidos en 4 fragmentos en las cadenas polipeptídicas (segmento 1: Tyr15, Trp18, Arg19, Tyr20; segmento 2: Leu44, Tyr46; segmento 3: Gln52, Ile53, Thr54 y segmento 4: Asp58, Gly59, Tyr60, Ile61, Gly62).

### **2.3.3 Complejo barnasa-barstar**

Barnasa es una ribonucleasa extracelular conformada por 110 residuos y secretada por el *Bacillus amyloliquefaciens*. El mismo organismo produce un polipéptido de 89 residuos, barstar, el cual se enlaza fuertemente a barnasa e inhibe su actividad ribonucleasa en el interior de la célula<sup>45</sup>. La estructura secundaria de barnasa está organizada en 3 hélices  $\alpha$  y 5 hojas  $\beta$  antiparalelas y la de barstar en 4 hélices  $\alpha$  y 3 hojas  $\beta$  paralelas, ver figura 2.7.

El sitio activo de la barnasa lo constituyen los residuos Glu73, His102, Arg87, Lys27 y el asa entre los residuos 56-62, que es indispensable en el reconocimiento de sustratos<sup>46</sup>. Barstar se enlaza a barnasa a través del asa (Pro27-Glu32) y la hélice  $\alpha 2$ <sup>47,48</sup>.



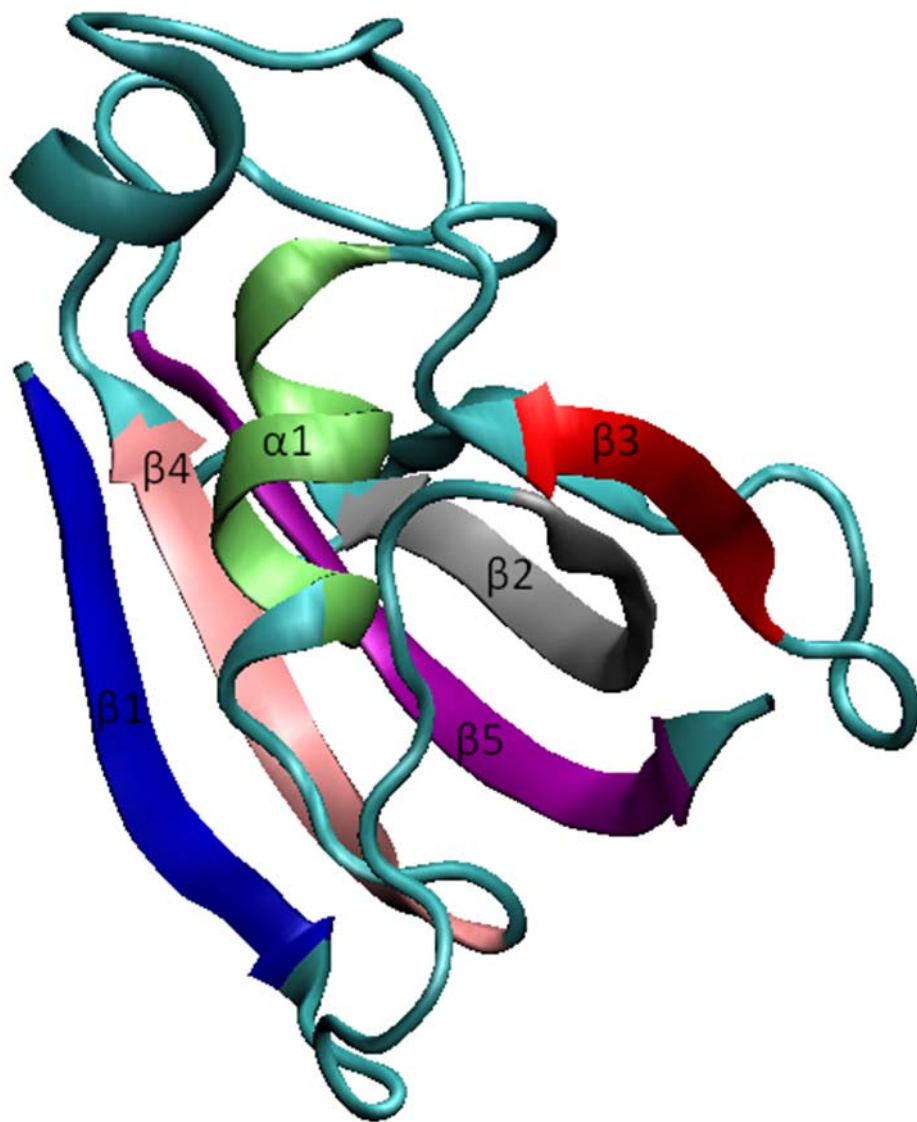


Figura 2.5. Diagrama de la proteína FKBP12 plegada, constituida por 5 hojas  $\beta$  antiparelas y una hélice  $\alpha$  de color verde, utilizando el programa VMD. Las hojas  $\beta$  están distribuidas así:  $\beta 1$  de color azul,  $\beta 2$  de color plateado,  $\beta 3$  de color rojo,  $\beta 4$  de color rosado y  $\beta 5$  de color morado.

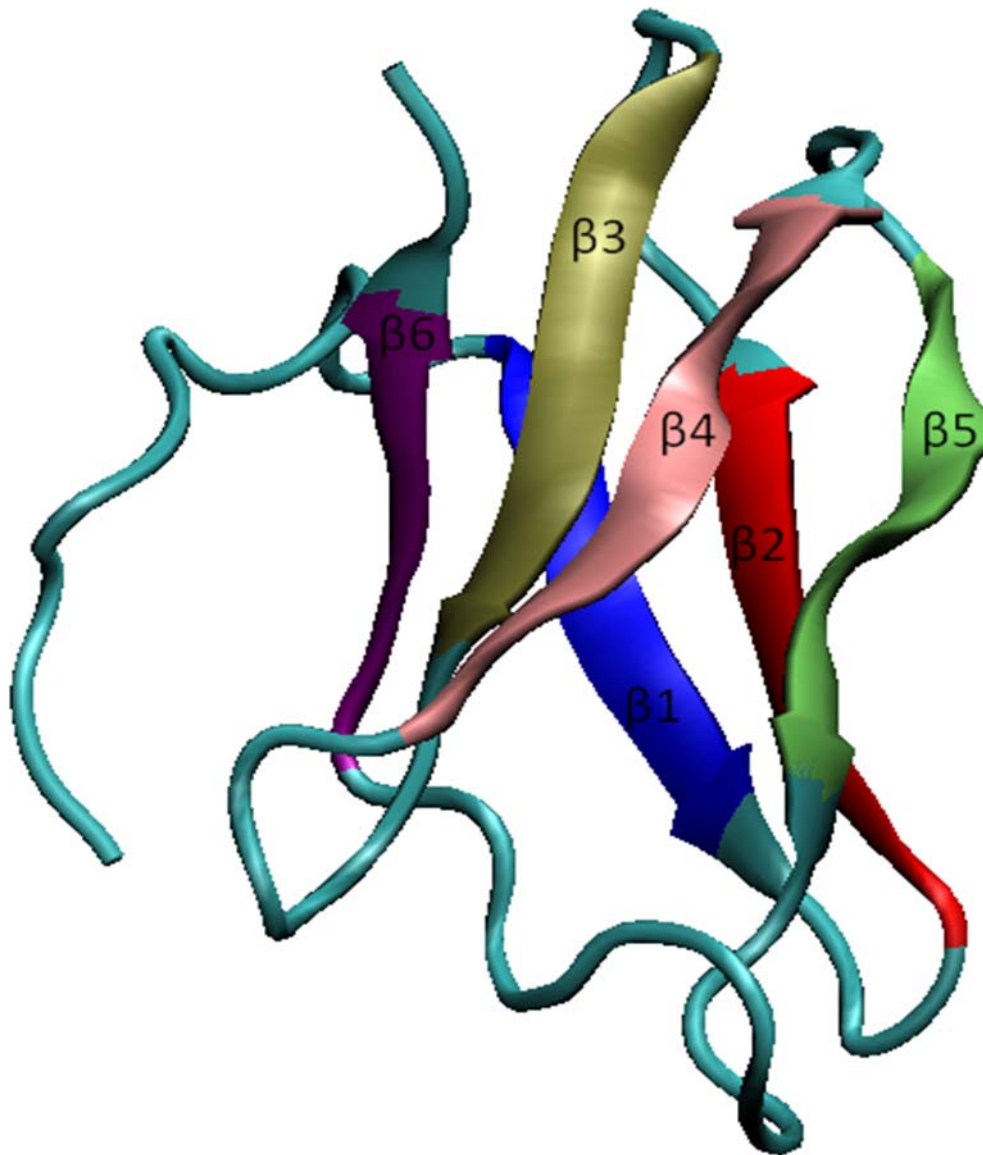


Figura 2.6. Diagrama de la proteína tendamistat plegada, formando una estructura  $\beta$  de 6 hojas, utilizando el programa VMD. Las hojas  $\beta$  están distribuidas así:  $\beta 1$  de color azul,  $\beta 2$  de color rojo,  $\beta 3$  de color beige,  $\beta 4$  de color rosado,  $\beta 5$  de color verde y  $\beta 6$  de color morado.

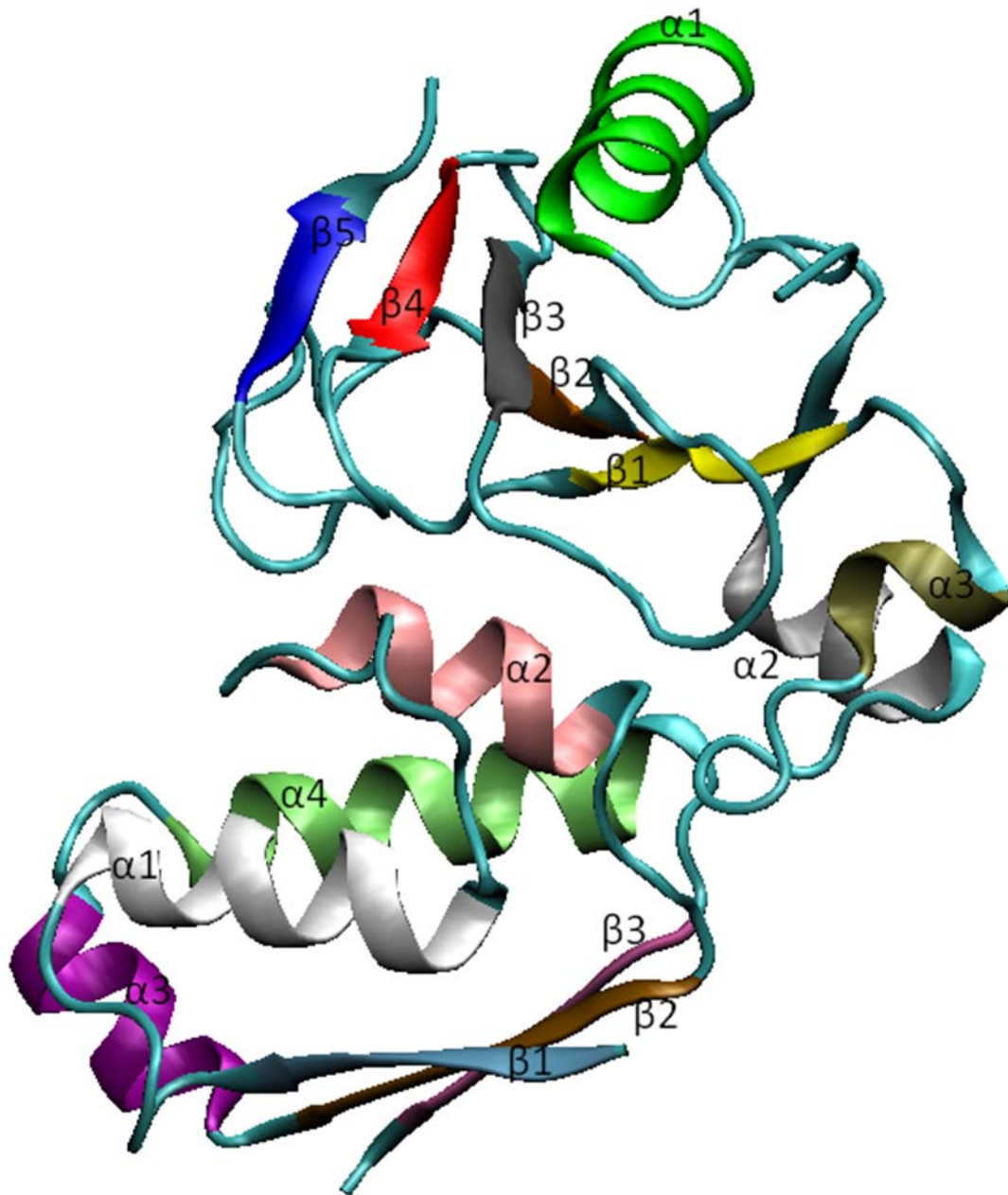


Figura 2.7. Diagrama del complejo barnasa-barstar plegado, utilizando el programa VMD. Barnasa está formada por 3 hélices  $\alpha$  ( $\alpha 1$  de color verde,  $\alpha 2$  de color plata y  $\alpha 3$  de color beigs) y 5 hojas  $\beta$  antiparalelas ( $\beta 1$  de color amarillo,  $\beta 2$  de color naranja,  $\beta 3$  de color gris,  $\beta 4$  de color rojo y  $\beta 5$  de color azul) y barstar la integran 4 hélices  $\alpha$  ( $\alpha 1$  de color blanco,  $\alpha 2$  de color rosado,  $\alpha 3$  de color morado y  $\alpha 4$  de color verde claro) y 3 hojas  $\beta$  paralelas ( $\beta 1$  de color azul claro,  $\beta 2$  de color café y  $\beta 3$  de color morado claro).

### 3. Metodología

En este capítulo se hace una descripción detallada de la preparación de los diferentes sistemas de estudio, de las condiciones en las que se llevó a cabo la minimización, equilibrio y la simulación de dinámica molecular<sup>49,50,52</sup>. También se muestran e interpretan algunas gráficas del monitoreo que se realizaron durante la minimización, equilibrio y la simulación de dinámica molecular. Finalmente se mencionan las ventajas de las herramientas diseñadas en el desarrollo de esta tesis, para el análisis de la simulación de dinámica molecular.

#### 3.1. Preparación de los sistemas

Las coordenadas iniciales de la proteína tendamistat se obtuvieron de la estructura cristalográfica de rayos X, a una resolución de 0.93 Å con código de acceso al PDB (protein data-bank) 1OK0<sup>44</sup>. A partir de estas coordenadas y eliminando las aguas cristalográficas (la estructura no contiene aguas internas, solo superficiales o cavidades de fácil acceso para el agua), se realizó la protonación de la proteína con el paquete psfgen de VMD en las posiciones estándares a pH= 7; las histidinas se protonaron en el nitrógeno  $\delta$ . La proteína posteriormente fue incluida en una caja de agua del modelo TIP3P, previamente equilibrada a 300 K y 1 atm, y las moléculas de agua a menos de 2.4 Å de distancia de cualquier átomo de la proteína fueron removidas. La proteína se solató de tal forma que existieran al menos 10 Å de agua en cada dirección en una caja rectangular, cuyas dimensiones resultantes fueron 42 x 49 x 51 Å. De esta forma se obtuvo un sistema con la proteína y 3436 moléculas de agua.

Las coordenadas iniciales de la FKBP12 se obtuvieron de una estructura dilucidada por RMN con código de acceso al PDB 1FKS<sup>52</sup>. Como estas coordenadas incluyen a los protones pero no incluyen a las moléculas de agua, entonces la proteína

se solvató (10 Å de agua en cada dirección) directamente, formando una caja con dimensiones 71 x 51 x 55 Å, un sistema con 5429 moléculas de agua.

Del complejo barnasa-barstar se empleó el archivo PDB con código de acceso 1BRS<sup>53</sup> resuelto por cristalografía de rayos X a una resolución de 2.0 Å. En esta estructura, la proteína barstar presenta la característica de ser una doble mutante Cys→Ala(40,82). La estructura de este PDB incluye a tres complejos barnasa-barstar; referidos como los complejos A-D, el B-E y el C-F que presentan las siguientes variantes: Las cadenas A y C de barnasa carecen de los aminoácidos 1 y 2, desde el N terminal. Este desorden del N terminal, es una característica relativamente frecuente, sin embargo, estos residuos sí están presentes en la cadena B de barnasa del complejo barnasa-barstar. Las cadenas D y E de barstar carecen de los residuos 64 y 65, debido también al gran desorden de la región. Para preparar los sistemas se empleó la cadena C de barnasa, que carece de los aminoácidos Ala1 y Gln2, y la cadena F de barstar del PDB 1BRS. Las cadenas C-F se eligieron primero porque la cadena F no carece de ningún residuo y segundo porque la inserción de dos aminoácidos N terminales en la cadena C no implican un rompimiento a la cadena principal como hubiera sido el caso donde se elijan las cadenas D o E de barstar, lo que originaría un mayor desajuste de los ángulos de torsión. También se incluyeron las moléculas de agua cristalográficas que aparecen a 5 Å de distancia de cada una de las proteínas y a 5 Å de distancia de la interfase barnasa-barstar, para un total de 97 moléculas de agua, de las cuales 23 corresponden a la región cercana a la interfase. Posteriormente, a la cadena C se le adicionaron los dos aminoácidos faltantes con el programa Swiss PDB viewer, conservando los ángulos de torsión permitidos según el diagrama de Ramachandran<sup>54</sup>. Como se realizaron simulaciones tanto del complejo como de las proteínas en forma independiente, se obtuvieron las coordenadas para el complejo, para la barnasa (cadena C) y para la barstar (cadena F). A continuación, estas estructuras se protonaron con el programa CNSsolve

(programa de cristalografía de rayos X y RMN), en esta ocasión, este programa se utilizó, porque el paquete psfgen de VMD protonaba las moléculas de agua orientándolas todas en el mismo sentido, mientras CNSolve (generate) las protonó considerando los átomos que estaban alrededor, y de esta forma estas moléculas alcanzaron orientaciones energéticamente favorables. Con el anterior procedimiento se obtuvieron los sistemas del complejo, de barnasa y de barstar con los protones, que posteriormente se solvataron, como en los casos anteriores con el paquete solvate de VMD, formando cajas de agua con dimensiones de 57.8 x 52 x 76 Å para el complejo, de 55.7 x 51.8 x 51.8 Å para barnasa y de 46.5 x 52.8 x 49.4 Å para barstar; en cada caso se obtuvieron los archivos con 3159 moléculas de agua para el complejo, 1727 para barnasa y 1432, para barstar. A raíz de que el complejo tiene una carga de -4, barnasa una carga de +2 y barstar una carga de -6, se neutralizó cada sistema, adicionando iones Cl<sup>-</sup> o Na<sup>+</sup> según el caso, a una distancia de 5 Å de la proteína. Para ello se utilizó el paquete ionize del programa VMD.

### **3.2. Minimización y equilibrio de los sistemas**

Se aplicó un algoritmo de minimización de energía a todas las proteínas, posteriormente los sistemas se equilibraron empleando el paquete de NAMD con el campo de fuerza de CHARMM y considerando el modelo TIP3P del agua. Las minimizaciones de los diferentes sistemas se realizaron en 3000 pasos empleando el algoritmo estándar del programa NAMD para minimizaciones.

Una vez que se alcanzó la configuración donde se han eliminado los contactos de muy alta energía, se comenzaron las simulaciones realizando un calentamiento gradual de cada sistema durante la primera etapa de la DM. Se inició el calentamiento asignando una temperatura de 100 K mediante la primera asignación de las velocidades (obedeciendo la distribución Maxwell-Boltzman correspondiente), posteriormente se fueron reasignando las velocidades con incrementos de 25 K cada 1000 pasos de dinámica molecular hasta completar 20000 pasos. Los incrementos se

continuaron hasta alcanzar una temperatura de 300 K. Durante esta etapa, para tratar las interacciones electrostáticas y de van der Waals se consideró una distancia de corte de 12 Å, usando una función de ajuste a cero a partir de los 10 Å, de tal forma que sus valores se aproximen a cero en la distancia de corte. Para agilizar el tratamiento de las interacciones de no enlace se construyeron listas de pares de átomos que están a 13.5 Å y solo sobre estos átomos, se evaluó la distancia interatómica; estas listas se renovaron cada 10 fs. Los cálculos de las interacciones de van der Waals se efectuaron cada 1 fs, mientras las interacciones electrostáticas se calcularon cada 2 fs. El algoritmo de integración se calculó para tiempos de 1 fs puesto que todos los enlaces se consideraron como no vibrantes mediante el algoritmo RATTLE<sup>55</sup>. Las condiciones periódicas se tuvieron en cuenta, al igual que el traslado de las coordenadas de las moléculas, en el caso de que éstas salieran de la caja. El archivo de salida de configuraciones (archivo de coordenadas) se grabó cada 0.5 ps, el archivo correspondiente a las energías se grabó cada 0.1 ps y el archivo del registro de las temperaturas se grabó cada 0.1 ps. En la figura 3.1 se muestra a manera de ejemplo, una gráfica de energía total contra pasos de la simulación (los primeros 3000 son de minimización y los siguientes 20000 son de calentamiento) para la proteína barnasa (monómero). En la figura 3.2 se muestra una gráfica de la temperatura del sistema contra los pasos de la simulación. Como se puede apreciar, cuando se realizan incrementos (reasignaciones) de velocidades, existen incrementos (picos) de temperatura los cuales se relajan parcialmente. Estos picos tienen una menor amplitud en las últimas etapas del calentamiento, cuando las reasignaciones corresponden a la misma temperatura (300 K).

Las diferentes minimizaciones y equilibrios de la energía se ejecutaron en un computador con 2 procesadores Xeon de 2.4 GHz en un tiempo de aproximadamente 10 horas para la tendamistat, 11 horas para la FKBP12, 24 horas para el complejo y 12 horas para cada una de las proteínas barnasa y barstar.

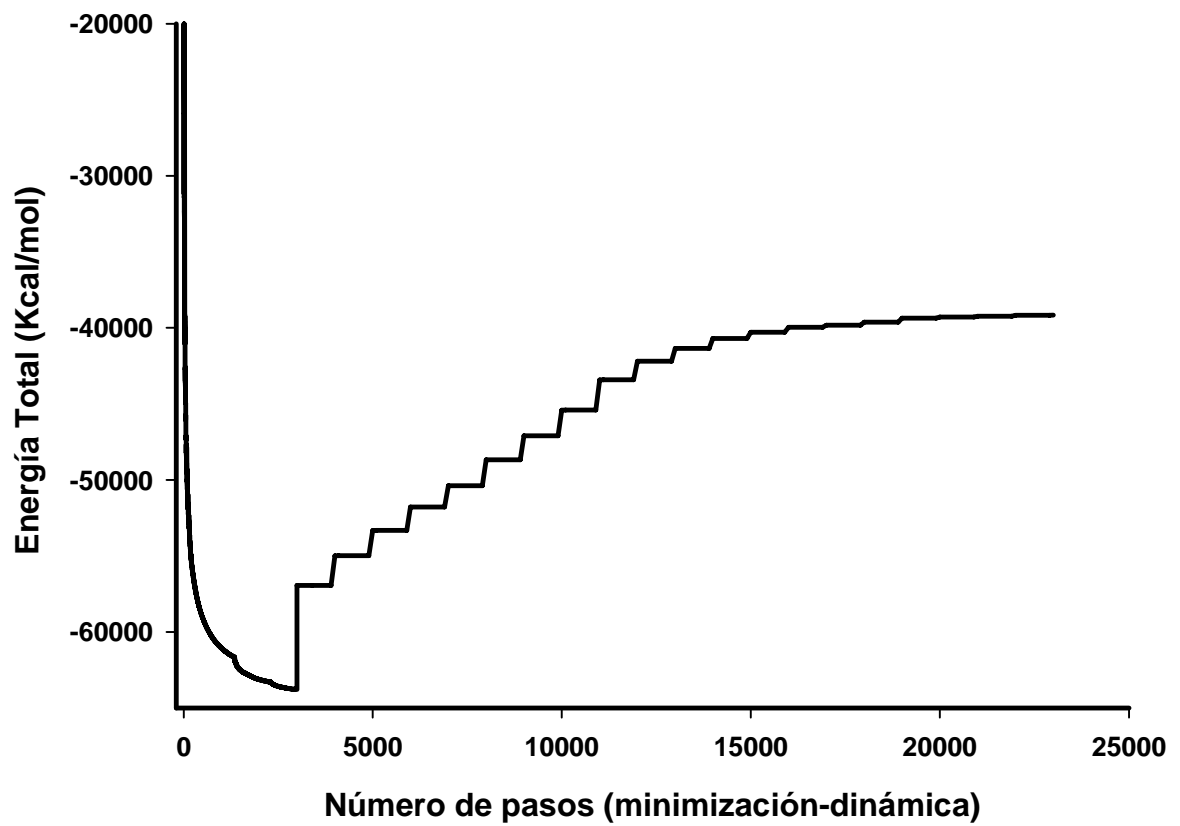


Figura 3.1. Comportamiento de la energía total en la minimización y equilibrio de la proteína barnasa. Los primeros 3000 pasos corresponden a la minimización y los restante 20000 pasos, corresponden al equilibrio.



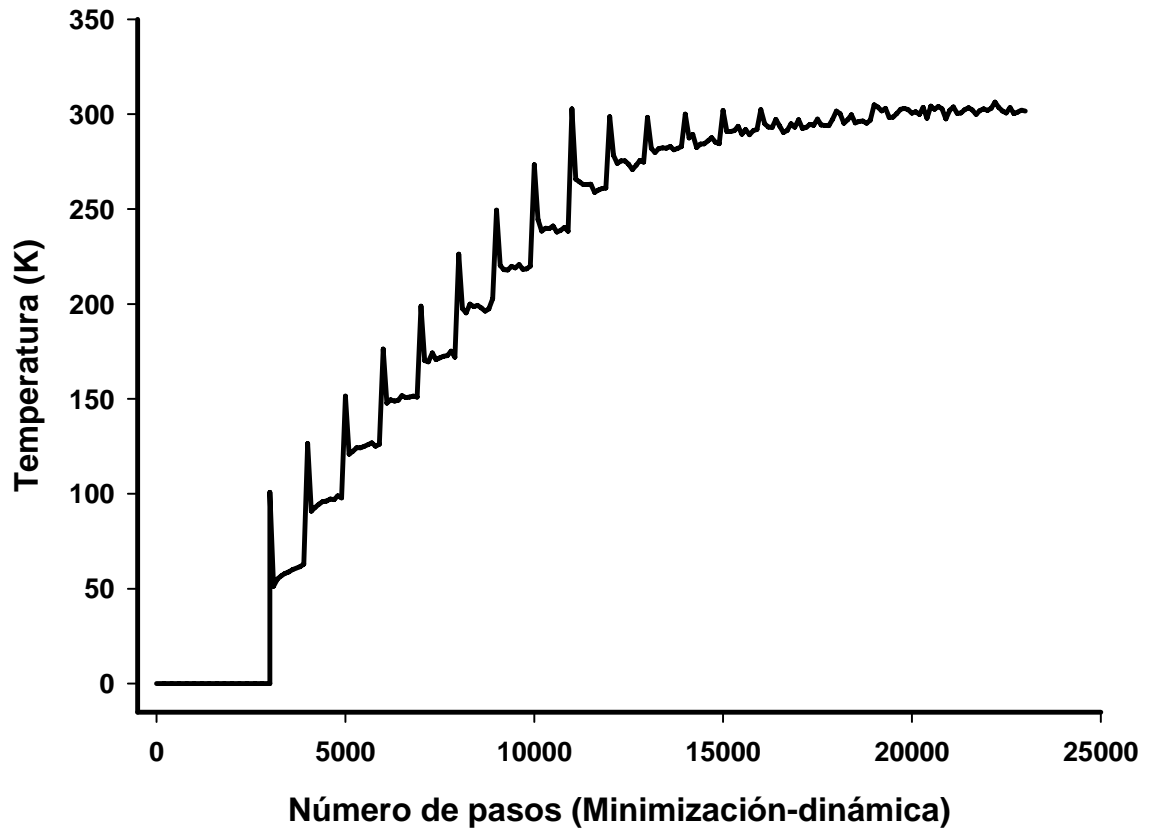


Figura 3.2. Comportamiento de la temperatura en la minimización y equilibrio de la proteína barnasa. Los primeros 3000 pasos corresponden a la minimización y los restante 20000 pasos, corresponden al equilibrio.

### 3.3. Simulación de dinámica molecular

Después de que cada sistema llegó al equilibrio de temperatura y densidad, se corrieron 3 ns de DM para la tendamistat, 3 ns para la FKBP12, 7 ns para el complejo, 6 ns para barnasa y 15 ns para barstar, bajo las siguientes condiciones: En el cálculo de las interacciones electrostáticas y de van der Waals se estimó una distancia de corte de 13 Å, empleando una función de ajuste a cero a partir de los 10 Å, de tal forma que los valores de esta función tiendan a cero en la proximidad de la distancia de corte. La distancia interatómica correspondiente a las interacciones de no enlace se evaluó sobre una lista de pares de átomos, que se encuentran a menos 15 Å de distancia entre sí. Estas listas se renovaron cada 20 fs. La integración de las ecuaciones de movimiento se efectuó cada 2 fs, pero haciendo la consideración de que los enlaces son rígidos (no vibrantes), mediante el algoritmo de RATTLE. Las fuerzas de van der Waals y las fuerzas electrostáticas de corto alcance se calcularon cada 2 fs, y las fuerzas electrostáticas de largo alcance se determinaron cada 4 fs. Se tuvieron en cuenta las condiciones periódicas de frontera, al igual que el traslado de las coordenadas de las moléculas, en el caso de que éstas salieran de la caja. Las fuerzas electrostáticas se calcularon mediante sumas de Ewald por el método PME, empleando una rejilla de dimensiones 32 x 32 x 32 Å para distribuir la carga. En todas las simulaciones se utilizó la dinámica de Langevin, pues éstas se efectuaron a presión constante de 1 atm y a temperatura constante de 300 K (ensamble isobárico-isotérmico), empleando el barostato y el termostato de Langevin. La fuerza de fricción de Langevin (resistencia de la partícula por efecto del solvente) se cuantificó con un coeficiente de fricción de 5 ps<sup>-1</sup>. La dinámica de Langevin no se realizó sobre los átomos de hidrógeno. El periodo de oscilación del pistón de Langevin fue de 100 fs y el tiempo de fricción del pistón fue de 50 fs. Los archivos de salida como son: el de configuraciones, el de energía y el de presión, se grabaron cada 0.2 ps.

Las simulaciones se ejecutaron en paralelo utilizando la supercomputadora de la UNAM, BAKLIZ, en la plataforma de LINUX, en un tiempo de aproximadamente 24 horas para el complejo y 12 horas para las otras macromoléculas, por cada ns, de cada proteína.

Las diferentes simulaciones efectuadas durante el desarrollo de este trabajo, se monitorearon a través del RMSD (Desviación de la raíz cuadrática media de las posiciones atómicas) y del radio de giro, para verificar la estabilidad de las macromoléculas. El RMSD es una medida numérica de la diferencia entre dos estructuras. Ésta es definida como:

$$RMSD = \sqrt{\frac{\sum_{j=1}^{N_t} \sum_{\alpha=1}^{N_\alpha} \left( \vec{r}_\alpha(t_j) - \langle \vec{r}_\alpha \rangle \right)^2}{N_\alpha}} \quad (25)$$

donde  $N_\alpha$  y  $N_t$  es el número de átomos y el número de etapas sobre las cuales las posiciones atómicas están siendo comparadas, respectivamente,  $\vec{r}_\alpha(t_j)$  es la posición del átomo  $\alpha$  en el tiempo  $t_j$  y  $\langle \vec{r}_\alpha \rangle$  es el valor de la posición del átomo  $\alpha$  sobre el cual la posición  $\vec{r}_\alpha(t_j)$  está siendo comparada.

El radio de giro (Rg) es una medida de las dimensiones de la proteína, definido como:

$$Rg = \sqrt{\frac{\sum_i m_i \left( \vec{r}_i - \vec{r}_c \right)^2}{\sum_i m_i}} \quad (26)$$

donde  $m_i$  es la masa de cada partícula  $r_i$  y  $r_c$  son las posiciones de los átomos  $i$  y del centro de masa  $c$ , respectivamente.

En la figura 3.3 se observan las gráficas de RMSD vs tiempo para cada uno de los sistemas de estudio. La primera gráfica de barstar (barstar A) es una representación de los últimos 9000 ps de su trayectoria de 15000 ps. En esta gráfica se observa que durante este tiempo las diferentes configuraciones de la simulación fluctúan alrededor de los 1.2 Å sin superar este valor (la estructura base es la del tiempo cero de la simulación después del calentamiento inicial). Sin embargo, es bastante notorio que las configuraciones de los primeros 4000 ps oscilan mucho más con respecto a la primera configuración, que las configuraciones de periodo de 4000-9000 ps, tornándose más estable este tramo; esta es la razón por la cual, para la proteína barstar se tomaron de los últimos 3000 ps de la gráfica, los primeros 800 ps para el análisis de la dinámica molecular, como se puede ver en la segunda gráfica de barstar (barstar B). El mismo argumento se empleó para definir los 800 ps para el análisis de la dinámica molecular de las restantes proteínas, de tal forma que para barnasa se tomaron los primeros 800 ps de los últimos 3 ns de una trayectoria de 6 ns, para el complejo los primeros 800 ps de los últimos 3 ns de una trayectoria de 7 ns y para la FKBP12 y la tendamistat los últimos 800 ps de sus trayectorias de 3 ns. Estos fragmentos de 800 ps para las proteínas barnasa, el complejo, la FKBP12 y la tendamistat se observan en la figura 3.3. Es importante mencionar que se observó mediante diversos ensayos, que las funciones de correlación que se calcularán quedarán bien definidas empleando únicamente este número de configuraciones (lo cual depende de la frecuencia natural de intercambio de las moléculas de agua).

El comportamiento del radio de giro vs tiempo, de cada una de las proteínas se observa en la figura 3.4. En estas gráficas se muestra que el radio de giro, de los 800 ps definidos anteriormente para cada una de las proteínas, oscila en un intervalo de 0.5 Å en sus respectivas escalas, lo que nos permite asegurar que las diferentes proteínas

no dieron indicio de un proceso de desnaturalización durante este tiempo, sino que por el contrario conservaron su estructura tridimensional.

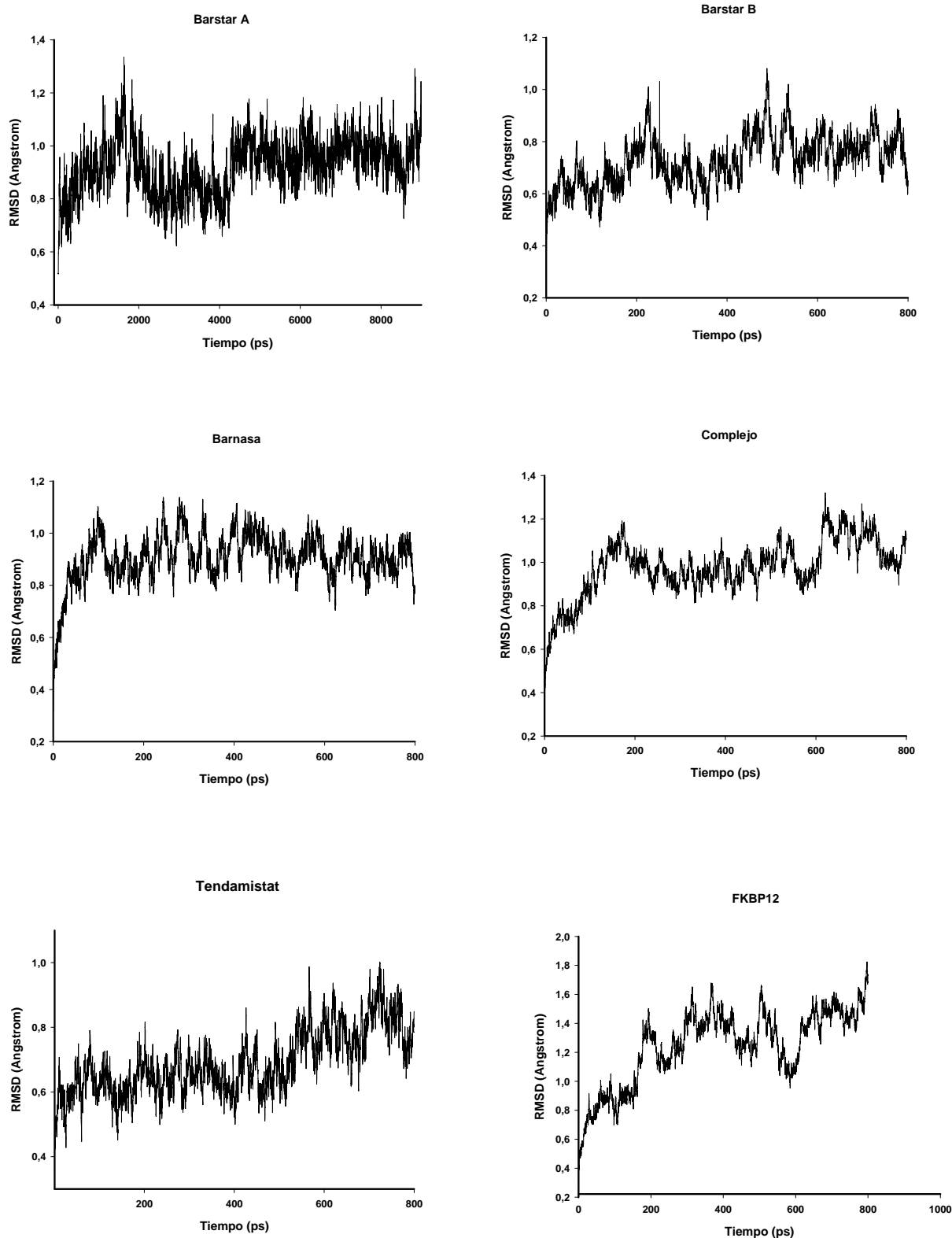


Figura 3.3. Gráficas de RMSD vs tiempo. La primera gráfica de barstar (barstar A) es una representación de los últimos 9000 ps de su trayectoria de 15000 ps. La segunda gráfica de barstar (barstar B) corresponde a la elección de los 800 ps más estables, considerando la gráfica de los 9000 ps (barstar A). En las restantes gráficas se representan los 800 ps más estables de cada una de las proteínas, los cuales se eligieron para el posterior análisis. Para construir cada una de estas gráficas, se consideró la primera configuración después del equilibrio de la dinámica molecular, como la de referencia.

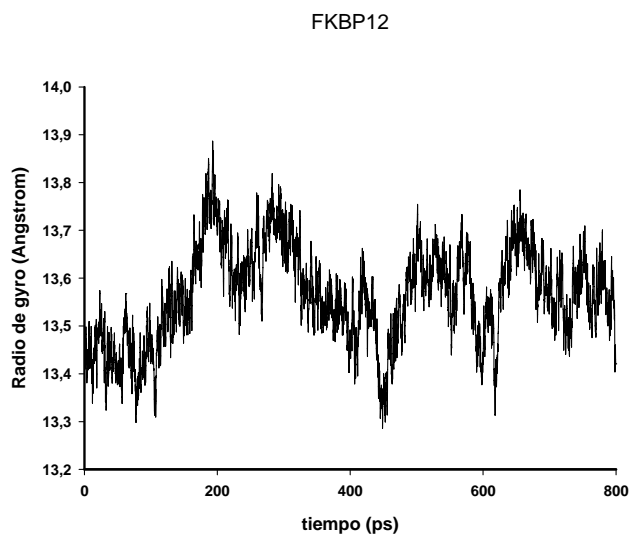
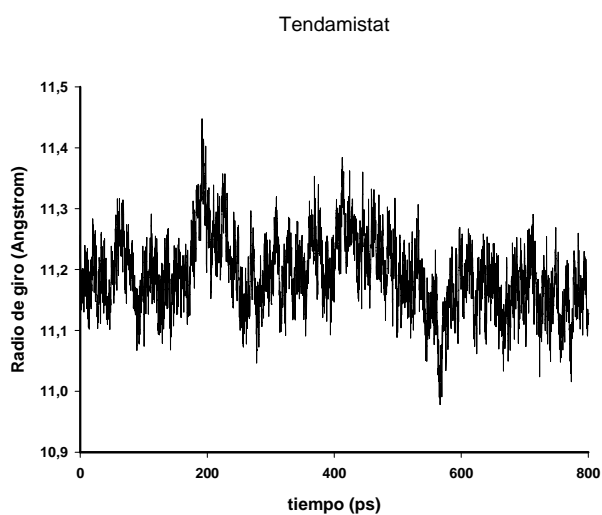
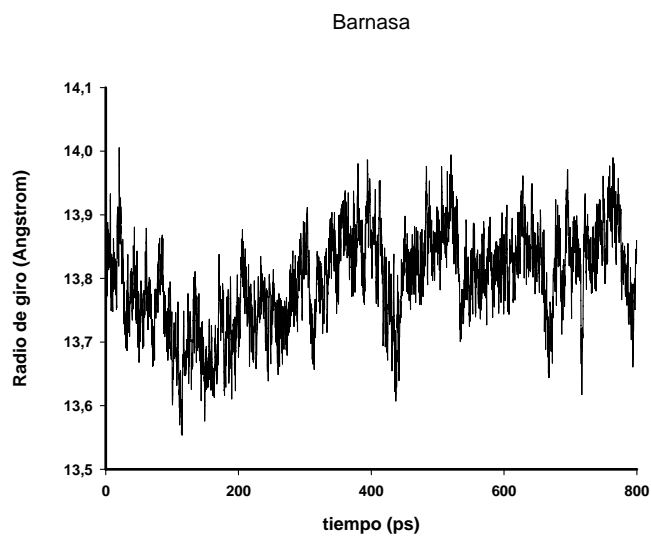
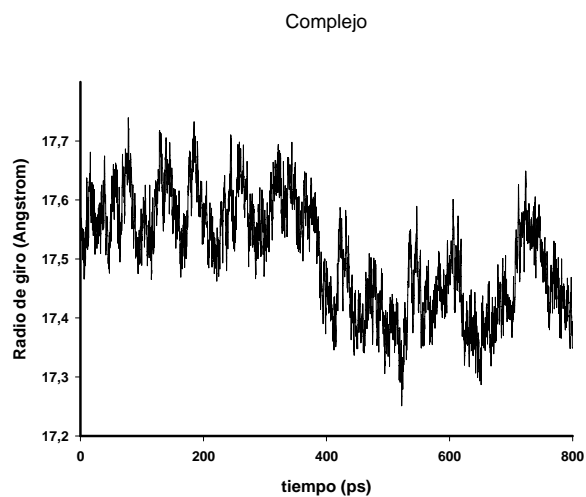
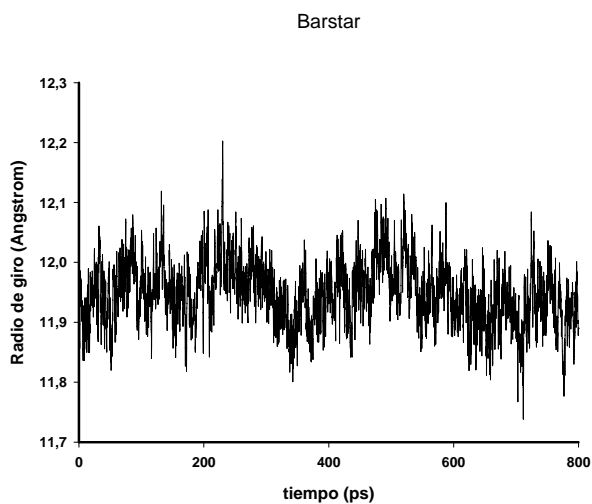


Figura 3.4. Radio de giro vs tiempo. En las cinco gráficas se representan los 800 ps que se eligieron de cada una de las proteínas, y los cuales se utilizan más adelante en el análisis de la simulación de dinámica molecular.

### 3.4. Ventajas de los programas empleados para el análisis de la simulación de dinámica molecular

Las trayectorias arrojadas por las simulaciones de dinámica molecular, se analizaron con el programa de la función de supervivencia (programa denominado TRAPro) escrito en C<sup>56</sup> bajo la plataforma de Linux, y su ejecución se automatizó mediante una serie de instrucciones escritas en lenguaje bash<sup>57,58</sup> (script TRAPro) en la plataforma de Linux. Para el análisis de las configuraciones de la simulación de DM también se elaboró el programa de la función radial de dos centros escrito igualmente en C. Las funciones tanto de los programas como del script, se especifican en el apéndice A.

Como se mencionó en la sección anterior, el análisis de la dinámica molecular sobre las trayectorias de las proteínas, se efectuó solo sobre los 800 ps definidos para cada una de las macromoléculas en la sección pasada. Estos archivos con un  $\Delta t = 0.2$  ps y un total de 4000 configuraciones; fueron modificados con el programa CATDcd, de tal forma que el  $\Delta t$  entre configuración y configuración aumentó a 1.0 ps, completando un total de 800 configuraciones. Estas modificaciones se realizaron con la finalidad de acelerar el cálculo computacional de la función de supervivencia. La figura 3.5 muestra un ejemplo del cálculo de la función de supervivencia del átomo de nitrógeno de la cadena principal de la Gly1 de la FKBP12 a diferentes  $\Delta t$  ( $\Delta t = 0.2$  ps,  $\Delta t = 0.4$  ps y  $\Delta t = 1.0$  ps), en esta figura se observa que el comportamiento de la función para los tres  $\Delta t$  presentan la misma tendencia, lo que nos permite considerar cualquiera de las tres situaciones para cuantificar los tiempos de decaimiento de la función. La razón de las ligeras diferencias de las gráficas de la figura 3.5, es debido a la exclusión de algunas configuraciones, es decir que si el programa de la función de supervivencia se evalúa sobre las 4000 configuraciones iniciales ( $\Delta t = 0.2$  ps), se hace un registro completo de la salida y entrada de las moléculas de agua, pero si se consideran solo 2000 configuraciones ( $\Delta t = 0.4$  ps),



entonces se van omitir las entradas y salidas de las moléculas de agua correspondientes a las configuraciones eliminadas, lo mismo sucede si tenemos en cuenta las 800 configuraciones ( $\Delta t = 1.0$  ps). Desde luego se van obviar muchos más registros de entradas y salidas de las moléculas de agua, pero como se indicó anteriormente, a pesar de estas omisiones, se observa claramente que el comportamiento de las gráficas se conserva y por lo tanto cualquiera de ellas es representativa de la dinámica de la esfera de solvatación de un átomo  $\alpha$  de la proteína.

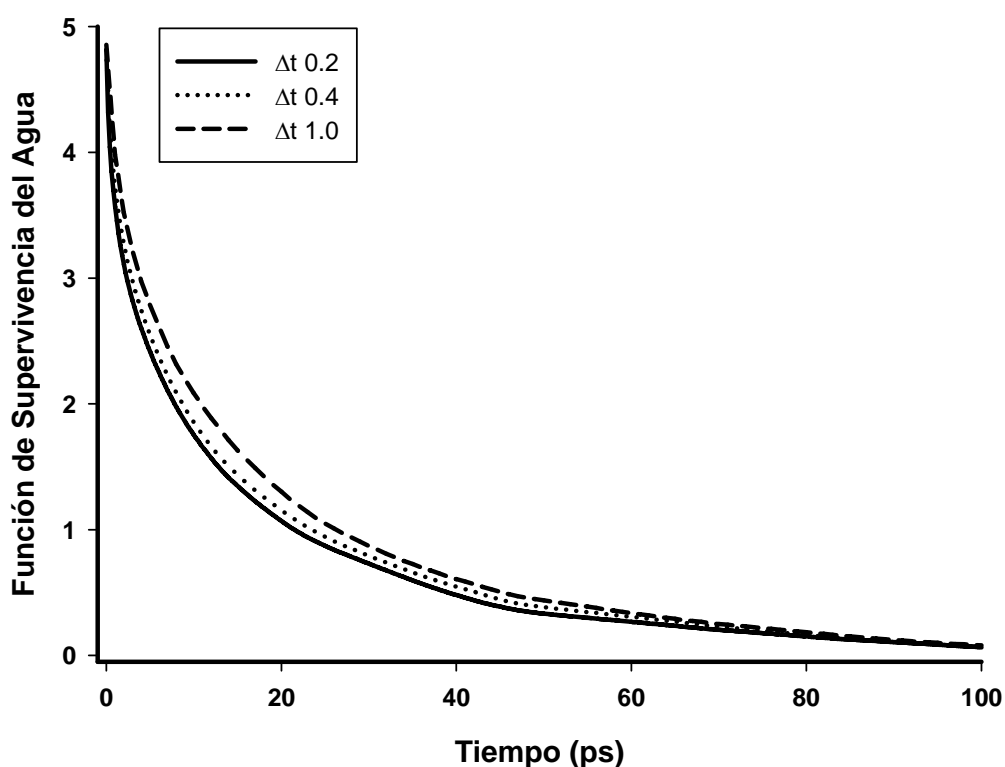


Figura 3.5. Función de supervivencia vs tiempo de residencia del agua en la esfera de solvatación del átomo de N de la cadena principal del aminoácido Gly1 de la FKBP12, a diferentes  $\Delta t$ . Línea continua  $\Delta t = 0.2$  ps, línea punteada  $\Delta t = 0.4$  ps y línea discontinua  $\Delta t = 1.0$  ps.

La principal ventaja de la forma como nosotros hemos estudiado las velocidades de intercambio de las moléculas de agua de la primera esfera de solvatación con las del seno de la solución, radica en el programa que se construyó con la función de supervivencia, al ser este un programa independiente, posibilita su optimización, de tal forma que se encuentren las mejores condiciones para su ejecución.

Para que este programa funcione se requiere de un archivo de salida por cada átomo de la proteína, asimismo después de su evaluación el programa arroja para cada átomo de la macromolécula dos archivos. El primero de estos archivos contiene los valores de la función de supervivencia  $P_{\alpha}(t)$  vs el tiempo de residencia que da origen a las gráficas de decaimiento, ver figura 2.1. El segundo archivo tiene nueve valores, correspondientes a los porcentajes de decaimiento desde el 90% hasta el 10%. Para ilustrar la forma como el programa de la función de supervivencia determina estos valores recurrimos a la figura 3.6, esta gráfica es una representación del decaimiento de la función de supervivencia, donde el número de moléculas de agua a  $t = 0$  corresponde a 3, es decir el 100% de las moléculas de agua presentes en la esfera de solvatación del átomo de oxígeno del grupo carbonilo del ácido aspártico N° 79 de la proteína FKBP12 en el momento inicial. De acuerdo con esto, al calcular el 50% de decaimiento encontramos que hay 1.5 moléculas de agua en la esfera de solvatación, si interpolamos este valor en la gráfica determinamos el tiempo en el cual esta cantidad de moléculas de agua está presente en la esfera, en este caso de 3.27 ps. Si hacemos lo mismo con el 20% de decaimiento, encontramos que hay 0.6 moléculas de agua en un tiempo de 9.58 ps. De esta forma se calcularon los porcentajes de decaimiento desde el 90% hasta el 10%, es decir se obtuvo la versión resumida de los valores de decaimiento.

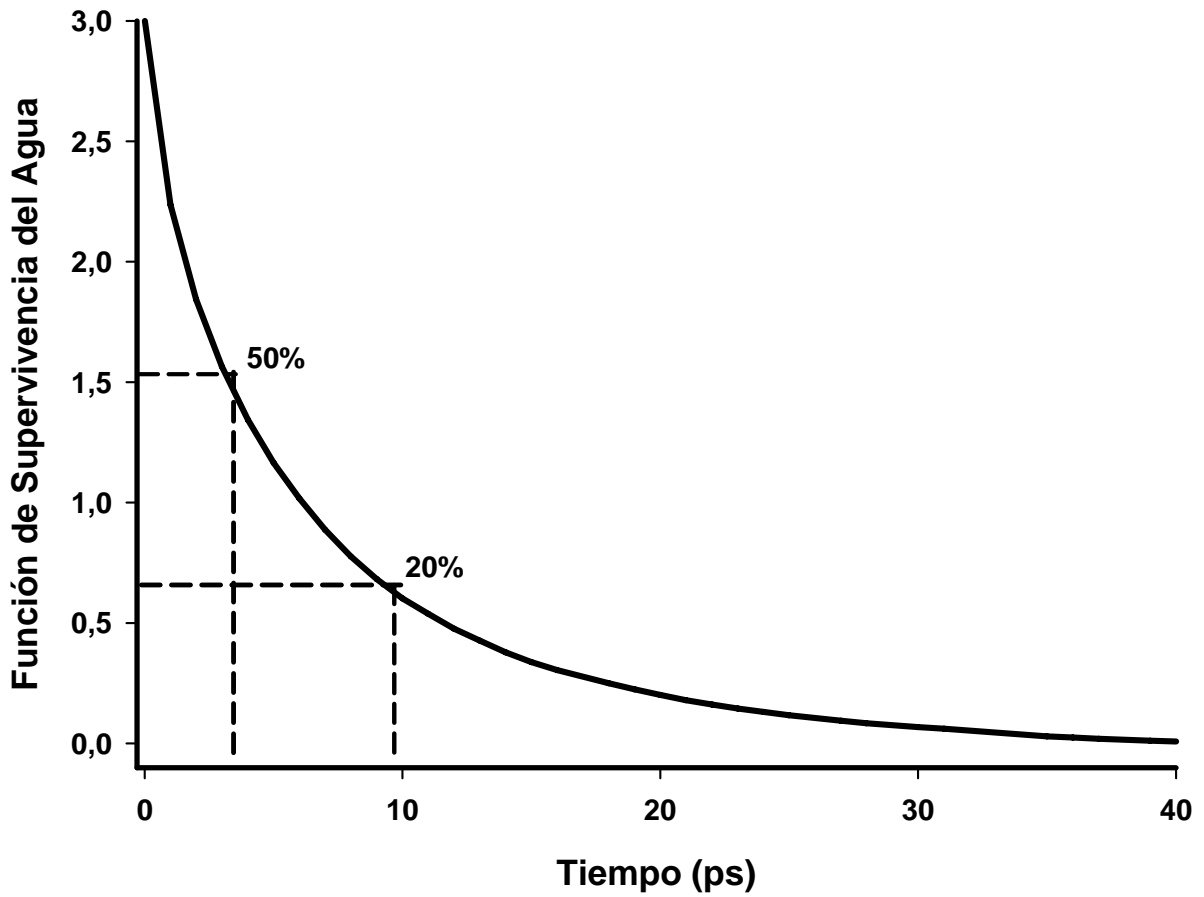


Figura 3.6. Datos de la función de supervivencia vs tiempo de residencia de la esfera de solvatación del átomo de oxígeno del grupo carbonilo del ácido aspártico N° 79 de la proteína FKBP12, calculados con el programa de la función de supervivencia (TRAPro). El 50% y el 20% son un par de ejemplos indicados en la gráfica y los cuales corresponden a la cantidad de moléculas de agua presentes en la esfera de solvatación cuando la función de supervivencia ha alcanzado un tiempo de 3.27 ps y de 9.58 ps respectivamente.

El script diseñado para el desarrollo esta tesis es una herramienta muy valiosa, debido a que contribuye a la automatización de la ejecución del programa de la función de supervivencia. Hay varias razones por las cuales se hace esta aseveración, primero el script se encarga de crear un archivo de salida que se utiliza en la ejecución del programa, este archivo se requiere para cada átomo de la proteína, como ya se había indicado. Segundo crea una carpeta para cada átomo de la proteína donde ubica el archivo de salida y los dos archivos que el programa proporciona después de su ejecución, los cuales se explicaron anteriormente. Tercero genera 9 archivos de coordenadas PDB de la proteína utilizando los porcentajes de decaimiento desde el 90% hasta el 10%, uno por cada porcentaje de decaimiento y con cuyos archivos es posible construir los mapas de colores (ver más adelante). Un archivo PDB consta de 11 columnas, las cuales tienen información sobre los átomos, sus coordenadas, la ocupancia y el factor B de temperatura. A el factor B le corresponde la última columna (indica el nivel de movilidad de los átomos en la estructura cristalina) la cual se puede sustituir. En nuestro caso el script lo que hace es reemplazar esta columna con los porcentajes de decaimiento, por ejemplo coloca los porcentajes de decaimiento al 90% de cada átomo de la proteína en la columna del factor B de temperatura y de esta forma construye el archivo PDB con los porcentajes de decaimiento al 90% y así sucesivamente hace lo mismo con los demás porcentajes de decaimiento, de tal forma que se completen 9 archivos PDB para cada proteína.

También el hecho de usar el programa NAMD tiene sus beneficios, debido a que este es un código diseñado para la ejecución de las simulaciones de dinámica molecular en paralelo, igualmente trabajar en el ambiente Linux tiene grandes ventajas. En conjunto, el programa de la función de supervivencia, el script, el programa NAMD y la plataforma de Linux facilitan la manera de representar el tiempo

de residencia de las moléculas de agua por medio de un código de colores asignado a cada átomo de la proteína.

El programa de la función de distribución radial de dos centros  $g(r_1, r_2)$  también es un programa independiente que permite su optimización, de tal forma que se encuentren las mejores condiciones para su ejecución. Otra virtud de este programa radica en la posibilidad que se tiene de escoger los 2 átomos, de los cuales se sospecha pueden coordinarse simultáneamente a una molécula de agua, mediante enlaces puente de hidrógeno, gracias a la información que nos proporciona el programa de la función de supervivencia. Vale la pena señalar que el programa de distribución radial de dos centros  $g(r_1, r_2)$  es una herramienta complementaria a la información proporcionada por el programa de la función de supervivencia, porque de no tener la información que nos proporciona este último, entonces estaríamos a la deriva en cuanto a la búsqueda de moléculas de agua coordinadas simultáneamente a dos átomos de la proteína y desde luego la identificación de estos casos demandaría mucho tiempo, pues se pueden hacer muchas combinaciones de pares de átomos.

En vista de las ventajas que representan las herramientas aquí diseñadas como son el programa de la función de supervivencia (TRAPro), el programa de la función de distribución radial de dos centros y el script, y las limitaciones que se tienen actualmente para estudiar la distribución de las moléculas de agua, nosotros pretendemos difundir a la comunidad científica, nuestras herramientas, a través de una página web, de tal forma que aquellos investigadores que se dedican a estudiar la forma y manera como las proteínas se asocian entre sí (docking) o al diseño de fármacos puedan en muy poco tiempo obtener información sobre la distribución de las moléculas de agua e identificar las que son de su interés.

## 4. Resultados y discusión

En este capítulo se interpretan los resultados obtenidos del análisis de la simulación de dinámica molecular a través de la función de supervivencia y la función de distribución radial de dos centros. Los resultados de la FKBP12 y de la tendamistat, aquí obtenidos, se compararon con la evidencia experimental existente, mientras que los resultados del complejo y las proteínas libres barnasa, barstar se confrontaron entre sí.

### 4.1. Resultados de la función de supervivencia

#### 4.1.1. Forma general del decaimiento de la función $P_{\alpha}(t)$

Con la finalidad de tener un punto de referencia con el cual comparar los gráficos de decaimiento de los diferentes átomos de las proteínas, elegimos una molécula de agua del seno de la solución y estudiamos su esfera de solvatación. La gráfica de la figura 4.1 corresponde al átomo de oxígeno de la molécula de agua N° 229, ubicada en el seno de la solución del sistema del complejo durante toda la simulación, aquí se puede apreciar que las 8.5 moléculas de agua que forman parte de la esfera de solvatación de este átomo, salen rápidamente de la esfera la cual se consideró como de 4.0 Å, de tal forma que antes de que se completen los 150 ps, no hay presencia de ninguna molécula de agua; es decir las moléculas de agua del seno de la solución se intercambian rápidamente entre ellas.

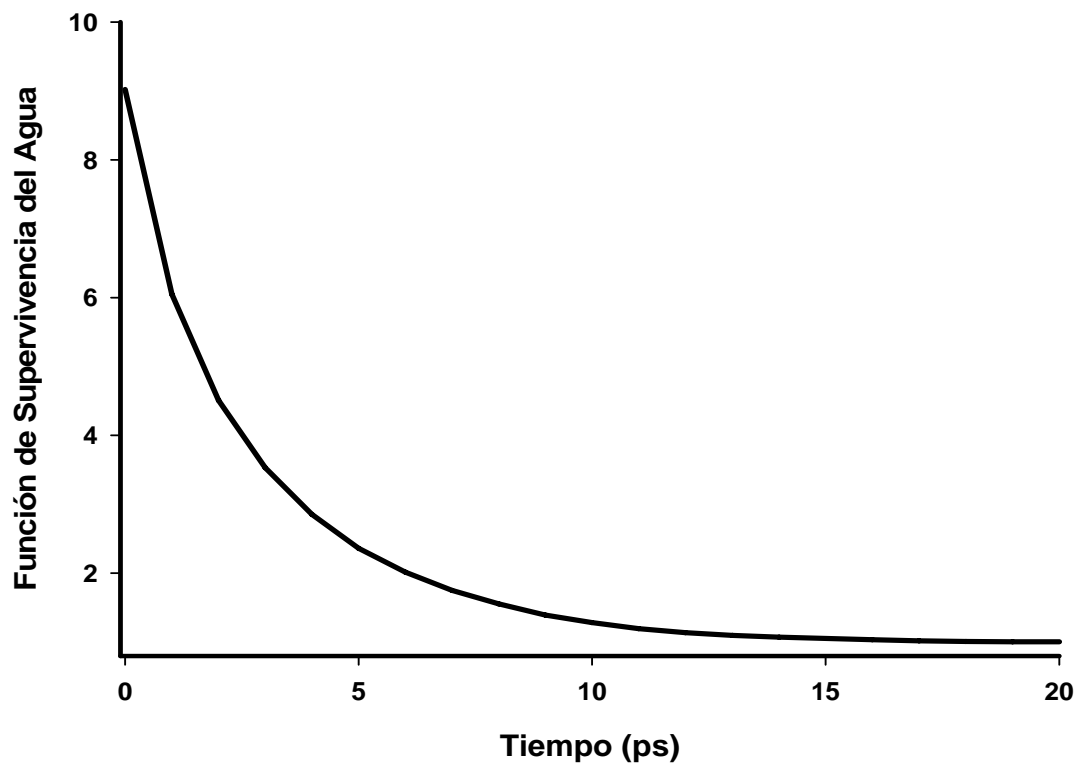


Figura 4.1. Tiempo de decaimiento de la función  $P_{\alpha}(t)$  de un átomo de oxígeno, correspondiente a una molécula de agua del seno de la solución del complejo barnasa-barstar.

Las moléculas de agua de las diversas esferas de solvatación de los átomos de las proteínas se intercambian a diferentes velocidades con el seno de la solución, es decir que hay moléculas de agua que permanecen largo tiempo en la primera capa de solvatación, hay otras que permanecen un tiempo promedio y otras que se intercambian rápidamente con el seno de la solución, prácticamente como si se tratase de moléculas de agua localizadas en el bulbo (con decaimientos semejantes a los de la figura 4.1).

Las gráficas de decaimiento pueden comportarse en forma exponencial o en forma biexponencial, estas diferencias en las gráficas nos permite clasificar las moléculas de agua como de intercambio rápido o de alto tiempo de permanencia, respectivamente. Esta variedad de comportamientos de las moléculas de agua se pueden observar en las gráficas de la figura 4.2 y figura 4.3. En la figura 4.2, se observa a modo de ejemplo, un grupo de átomos donde la función decae de forma aproximadamente exponencial, lo que es un referente de que la mayoría de las moléculas de agua de las esferas de solvatación del oxígeno del grupo carbonilo del Asp79, del nitrógeno E2 de la Gln70, del nitrógeno de la cadena principal del Glu31 y del oxígeno del grupo carbonilo de la Gly1 de la proteína FKBP12, se intercambian rápidamente con el seno de la solución. De hecho se observa que la función  $P_{\alpha}(t)$  que tarda más tiempo en decaer no sobre pasa los 40 ps (Asp79:O) y la más rápida no supera los 10 ps (Gly1:O). Sin embargo, de todas las esferas de solvatación de los átomos anteriormente citados, sólo la esfera de solvatación de la Gly1:O se comporta como si se tratase de las moléculas de agua del seno de la solución. Estas diferencias de la dinámica de solvatación de los diferentes átomos de la proteína, posiblemente se deban a qué tan expuestos estén y si éstos están o no cerca de átomos electronegativos que contribuyan en la formación de los enlaces puentes de hidrógeno.



En la figura 4.3 se muestra que algunas moléculas de agua de las esferas de solvatación del oxígeno del grupo carbonilo de la Val55, del oxígeno  $\gamma$  de la Ser67, del oxígeno del grupo carbonilo de la Thr96 y del nitrógeno de la cadena principal de la Gly10 de la proteína FKBP12, tardan un tiempo superior al promedio en intercambiarse con el seno de la solución, pues el tiempo de decaimiento de la función  $P_{\alpha}(t)$  de este grupo de átomos va desde 30 ps a tiempos superiores a los 100 ps. Además, las gráficas de estos átomos presentan un comportamiento biexponencial, lo que indica que en la esfera de solvatación hay moléculas de agua de alto tiempo de residencia y tardan un mayor tiempo en intercambiarse con el seno de la solución. El comportamiento bimodal de las gráficas se debe a que inicialmente las moléculas de agua salen de la esfera de solvatación rápidamente, pero en ella permanecen moléculas de agua que pueden estar ubicadas en cavidades de la proteína, estar rodeadas de átomos electronegativos que contribuyan en la formación de enlaces puentes de hidrógeno o estar coordinadas simultáneamente por enlaces puentes de hidrógeno a dos átomos de la proteína y por lo tanto tienen tiempos de intercambio mucho mayores.

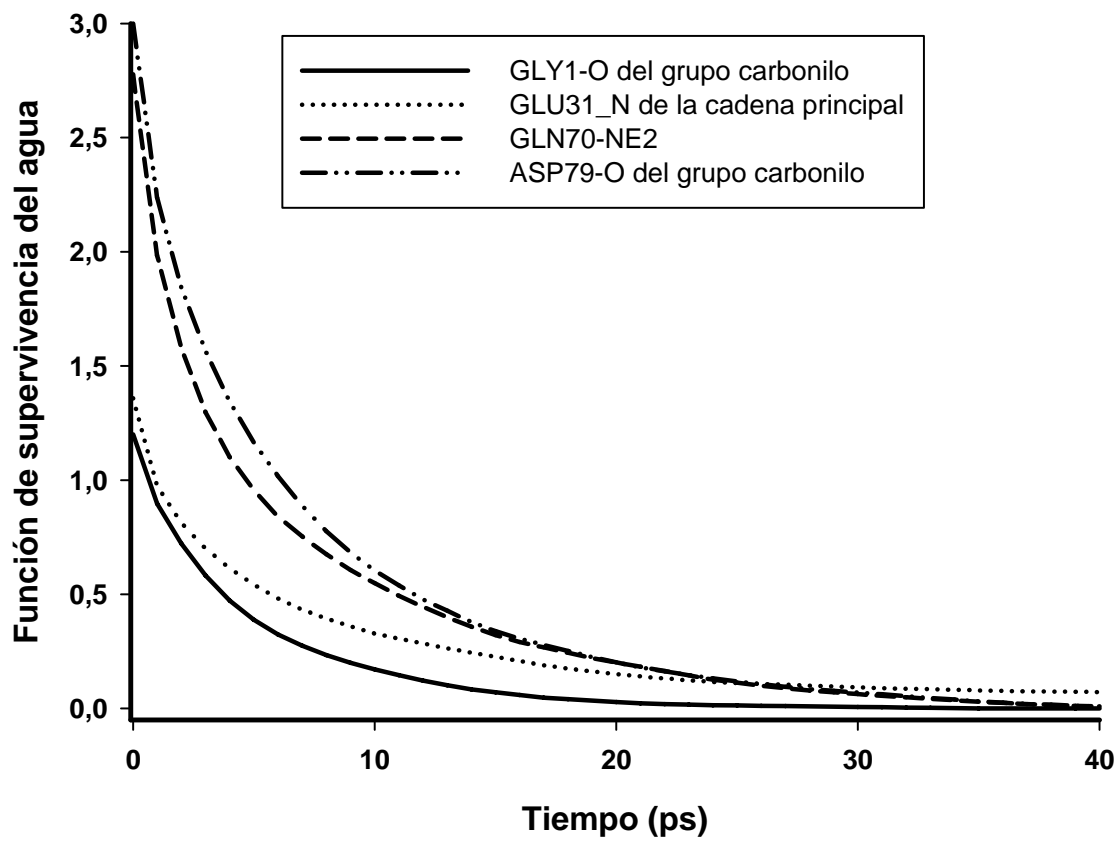


Figura 4.2. Tiempo de decaimiento de la función  $P_{\alpha}(t)$  para diferentes átomos de la proteína FKBP12 (Gly1:O del grupo carbonilo, Glu31:N de la cadena principal, Gln70:NE1 y Asp79:O del grupo carbonilo). Esta gráfica es una representación de las moléculas de agua que se intercambian rápidamente con el seno de la solución.

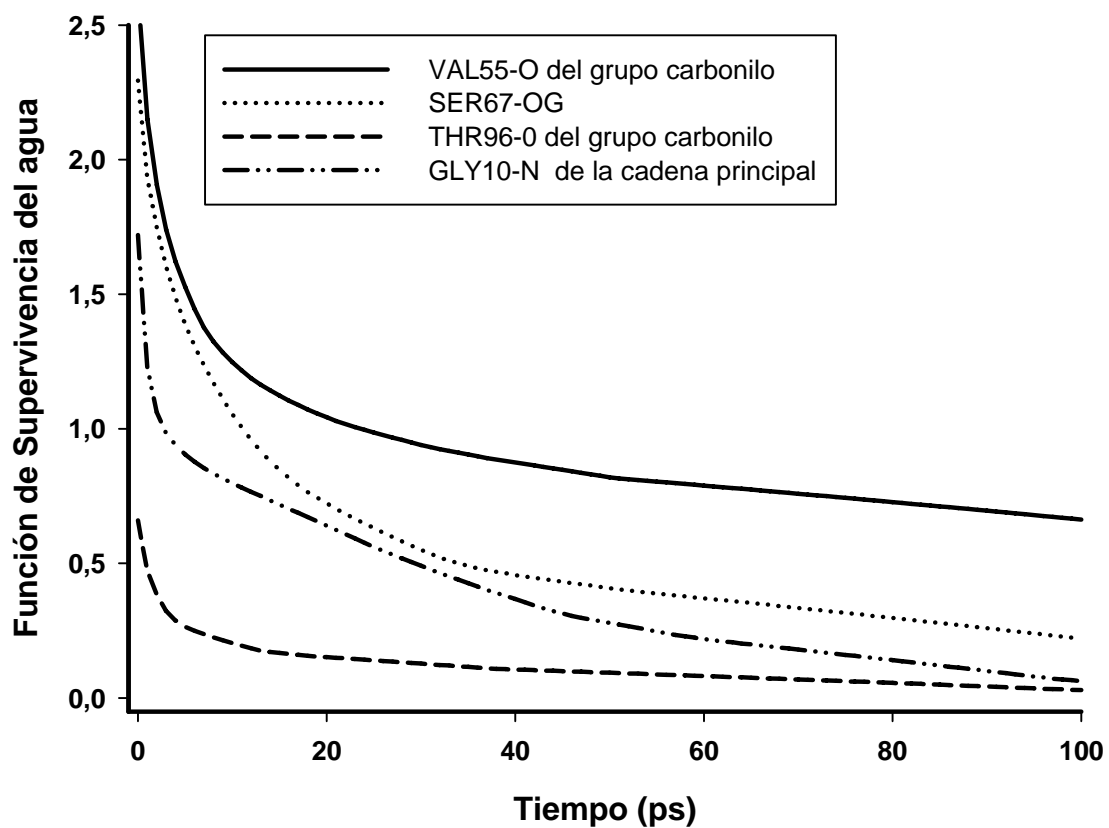


Figura 4.3. Tiempo de decaimiento de la función  $P_\alpha(t)$  para diferentes átomos de la proteína FKBP12 (Val55:O del grupo carbonilo, Ser67:OG, Thr96:O del grupo carbonilo y Gly10:N de la cadena principal). Esta gráfica es una representación de las moléculas de agua que presentan altos tiempos de residencia en la esfera de solvatación.

#### 4.1.2. Ventajas de los archivos PDB

Al analizar la trayectoria de la simulación de dinámica molecular con la función de supervivencia obtenemos por cada átomo de la proteína un gráfico de la función  $P_\alpha(t)$  vs tiempo (ver figura 2.1). Si deseamos analizar esta información entonces tendríamos que revisar por cada átomo de la proteína un gráfico de esta índole, proceso que demandaría mucho tiempo. Esta es la ventaja de los 9 archivos PDB que se construyeron, éstos contienen información resumida del decaimiento de la función.

El tiempo  $\tau_x$  característico de decaimiento lo podemos definir así:

$$\tau_x = \text{tiempo tal que } P_\alpha(\tau_x) = \frac{x}{10}(P_\alpha(0)) \quad (27)$$

donde  $\tau_x$  corresponde al tiempo de decaimiento de la función, y  $P_\alpha(t=0)$  el número de moléculas de agua que están inicialmente presentes en la primera capa de solvatación.

Con los archivos PDB a diferentes  $\tau_x$  se construyen los mapas de colores que proporcionan información inmediata del comportamiento de las moléculas de agua en la esfera de solvatación (ver más adelante). Considerando los archivos PDB a diferentes tiempos de decaimiento, podemos estudiar el comportamiento de la esfera de solvatación de manera gráfica teniendo en cuenta el rango de  $\tau_2$  a  $\tau_9$  o simplemente se puede elegir un  $\tau_x$  en particular para observar el comportamiento de la esfera de solvatación con base en un solo dato.

Los análisis que posteriormente presentamos, se hicieron sobre los archivos PDB que tienen tiempos de decaimiento al 20%, es decir  $\tau_2$ . Los archivos PDB con valor de  $\tau_2$  se eligieron puesto que ellos contienen información acumulada que abarca desde el tiempo de decaimiento del 90% al 20%, además de que nos permiten hacer una mejor discriminación de las moléculas de agua de alto tiempo de residencia entre sí.

Vale la pena aclarar que las gráficas de decaimiento son muy útiles sobre todo al evaluar la esfera de solvatación de algún átomo en particular, es decir con los mapas de colores podemos, a simple vista, identificar los átomos de la proteína de nuestro interés y posteriormente, para más detalles podemos revisar únicamente las gráficas de decaimiento de  $P_{\alpha}(t)$  de estos átomos que nos pueden proporcionar información relevante.

#### **4.1.3. Gráficas de distribución de los tiempos de residencia**

Si las moléculas de agua no forman un cascarón en torno a la proteína y tampoco actúan al unísono con ella, entonces ¿cómo se distribuyen las moléculas de agua en torno a la proteína? ¿Qué fracción de ellas se intercambia rápidamente y que fracción permanece unida a la proteína durante tiempos mayores? Para responder a estas preguntas debemos fijarnos en los tiempos de residencia de las moléculas de agua.

La distribución de los tiempos de residencia de las moléculas de agua dentro de la primera esfera de solvatación de cada una de las proteínas se muestra en las figuras 4.4 y 4.5. La distribución de la figura 4.4 se construyó sin considerar los hidrógenos de la proteína y la distribución de la figura 4.5 se elaboró únicamente considerando los átomos con valores de  $P_{\alpha}(t) > 0.15$  a  $t = 0$  (número de coordinación promedio), esto con la finalidad de que los átomos que se tengan en cuenta, por los menos presenten algún contacto con las moléculas de agua del medio durante la simulación (el valor de 0.15 es arbitrario). Independientemente de las discriminaciones que se hicieron en la elaboración de las gráficas de cada una de las macromoléculas de estudio; se observa que la mayoría de las moléculas de agua están caracterizadas por tener tiempos de residencia mucho menores a 40 ps, lo que indica que estas moléculas de agua se intercambian rápidamente con el seno de la solución. Las moléculas de agua de rápido intercambio en la esfera de solvatación, son de vital importancia para el funcionamiento de las proteínas. Por ejemplo, en el momento de la

formación del complejo entre la FKBP12 y el ligando ascomicina, es necesario que las moléculas de agua ubicadas en la zona de interacción, se desplacen, de tal forma que las macromoléculas puedan interactuar entre sí. Esto es posible gracias a las moléculas de agua que se intercambian en el seno de la solución en el rango de 10-40 ps, debido a que la entropía de estas moléculas aumenta en el proceso de desolvatación. Estas moléculas de rápido intercambio también pueden contribuir a que las enzimas adquieran varias conformaciones, lo que favorece su actividad biológica, igualmente pueden ayudar en la apertura y cierre de canales de proteínas de membrana.

Hay muy pocas moléculas de agua que presentan un tiempo de residencia mayor a los 40 ps, de hecho, se puede apreciar que hay átomos para los cuales la función de supervivencia no ha decaído a los 100 ps, indicando que durante este periodo, por lo menos una molécula de agua de la esfera de solvatación del átomo no se ha intercambiado con el seno del solvente. Muchas de estas moléculas de disolvente están atrapadas dentro de huecos o cavidades de la superficie de la proteína, lo cual disminuye la capacidad de intercambio de las moléculas de agua con el seno del solvente. Como se verá más tarde, la posibilidad de que una molécula de agua forme puentes de hidrógeno simultáneos a dos átomos de la proteína, también es un factor determinante en los altos tiempos de residencia que pueden presentar las moléculas de agua.

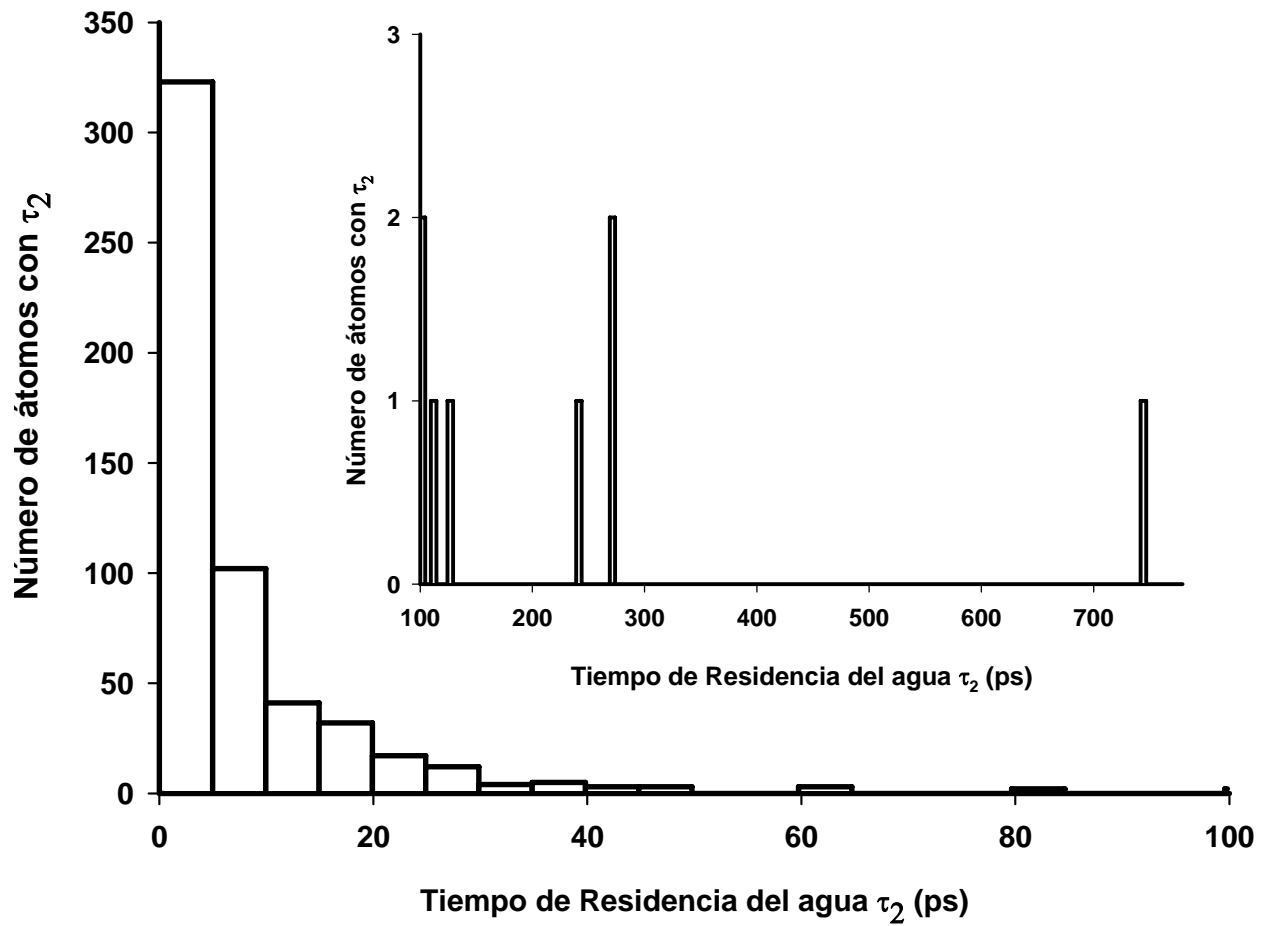


Figura 4.4. Distribución de los tiempos residencia de las moléculas de agua dentro de la primera esfera de solvatación de la proteína tendamistat, sin considerar los átomos de hidrógeno y teniendo en cuenta  $\tau_2$  de decaimiento.

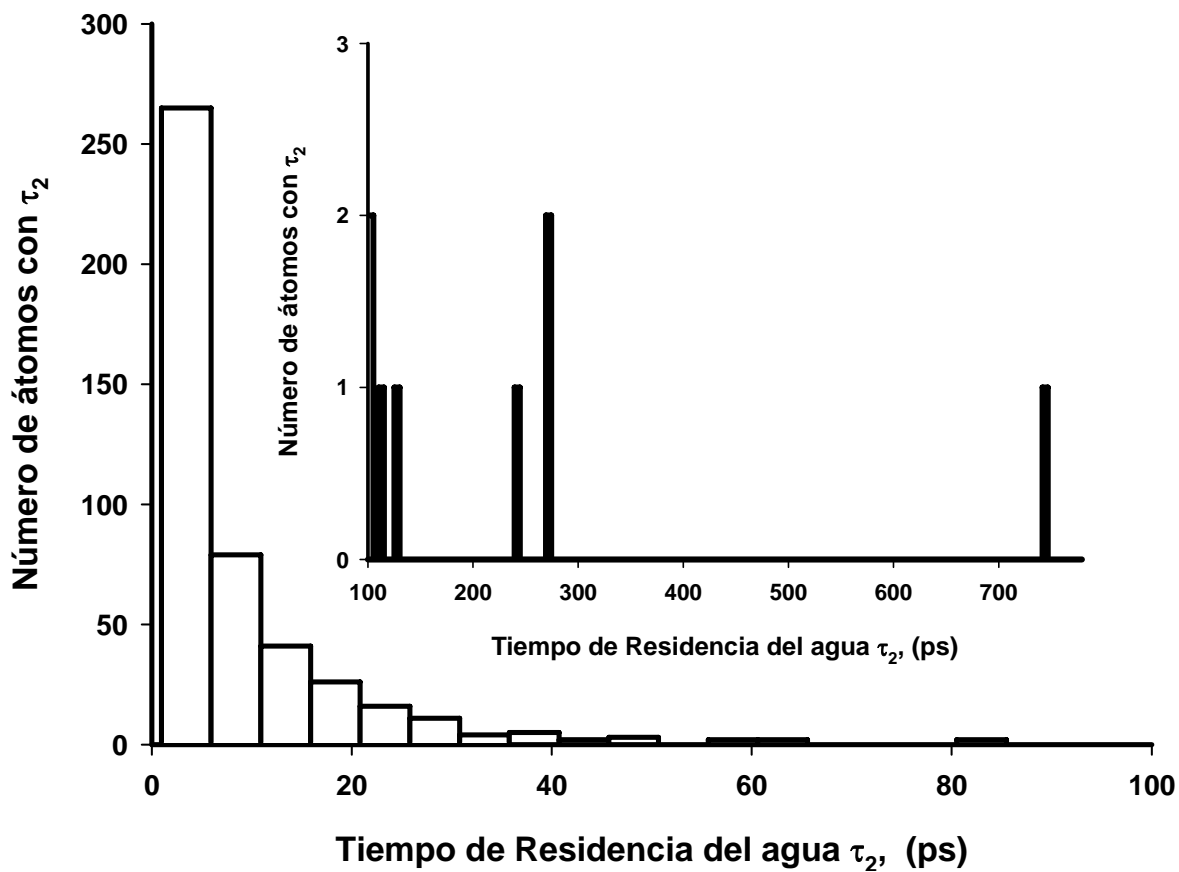


Figura 4.5. Distribución de los tiempos residencia de las moléculas de agua de la proteína tendamistat dentro de la primera esfera de solvatación, considerando los valores de  $P_\alpha(t=0) > 0.15$  y teniendo en cuenta  $\tau_2$  de decaimiento.



#### **4.1.4. Comentarios generales de la distribución de los tiempos de residencia en la superficie**

Una forma práctica y sencilla de observar el comportamiento de las moléculas de agua en la primera esfera de solvatación de una proteína, es a través de los mapas de colores, los cuales se pueden visualizar con el programa VMD gracias a los archivos PDB que se formaron durante la ejecución del programa de la función de supervivencia (programa TRAPro). El visualizador VMD nos permite explorar tanto la superficie como la parte interna de la proteína.

En la figuras 4.6-4.8 se observan los mapas de colores de las proteínas FKBP12, tendamistat y el complejo, en ellos se aprecian una serie de tonalidades de colores que van desde el rojo, el cual es un indicador de que en esta zonas de la proteína se ubican moléculas de agua que se intercambian rápidamente con el seno de la solución (tiempos de residencia  $\tau_2$  menores a 10 ps), hasta el azul el cual es un referente de que en estas zonas se localizan moléculas de agua que presentan altos tiempos de residencia (tiempos  $\tau_2$  mayores a los 30 ps). En estas figuras también se puede observar que son muy pocas las zonas donde se ubican moléculas de agua de alto tiempo de residencia, posiblemente este comportamiento se deba a que son muy escasas las partes de la proteína que tienen las condiciones propicias para que las moléculas de agua residan por largo tiempo, como son por ejemplo, la conjunción de varios átomos cargados positiva o negativamente a distancias particularmente adecuadas entre ellos para la formación de enlaces tipo puente de hidrógeno con una misma molécula de disolvente. Es importante destacar que la mera existencia de una cavidad superficial en la proteína no implica que en esta existan moléculas de agua de intercambio lento.

Es bastante notorio que en los mapas de colores, las moléculas de agua de alto tiempo de residencia se ubican sobre todo en grietas donde hay varias cadenas laterales cercanas con capacidad de formar puentes de hidrógeno. Tanto en la proteína FKBP12 como en la tendamistat se observa que las zonas de alto tiempo de residencia de las moléculas de agua están muy dispersas, mientras que en la zona de interacción del complejo, se aprecia que los sitios de alto tiempo de residencia están más aglomerados, esto se debe a que la zona de interacción del complejo es principalmente de naturaleza altamente polar<sup>48</sup>.

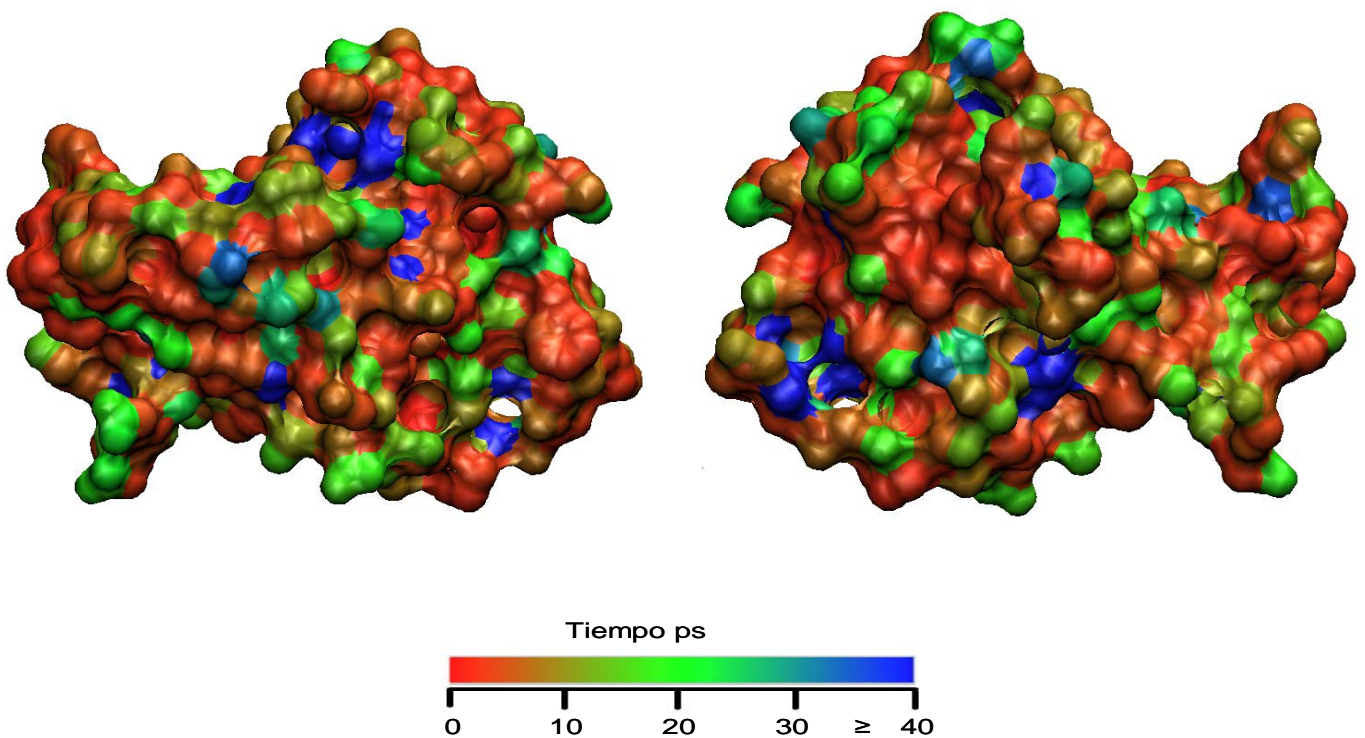


Figura 4.6. Distribución de las moléculas de agua en la proteína FKBP12, de acuerdo al tiempo de residencia de éstas en la proteína, mediante la evaluación de la función de supervivencia, teniendo en cuenta  $\tau_2$  de decaimiento. La escala de colores es un indicador del tiempo de residencia de las moléculas de agua en la primera esfera de solvatación. Esta escala va desde el rojo que corresponde a las moléculas de agua que se intercambian rápidamente con el seno de la solución, al azul que indica altos tiempos de residencia de las moléculas de agua.

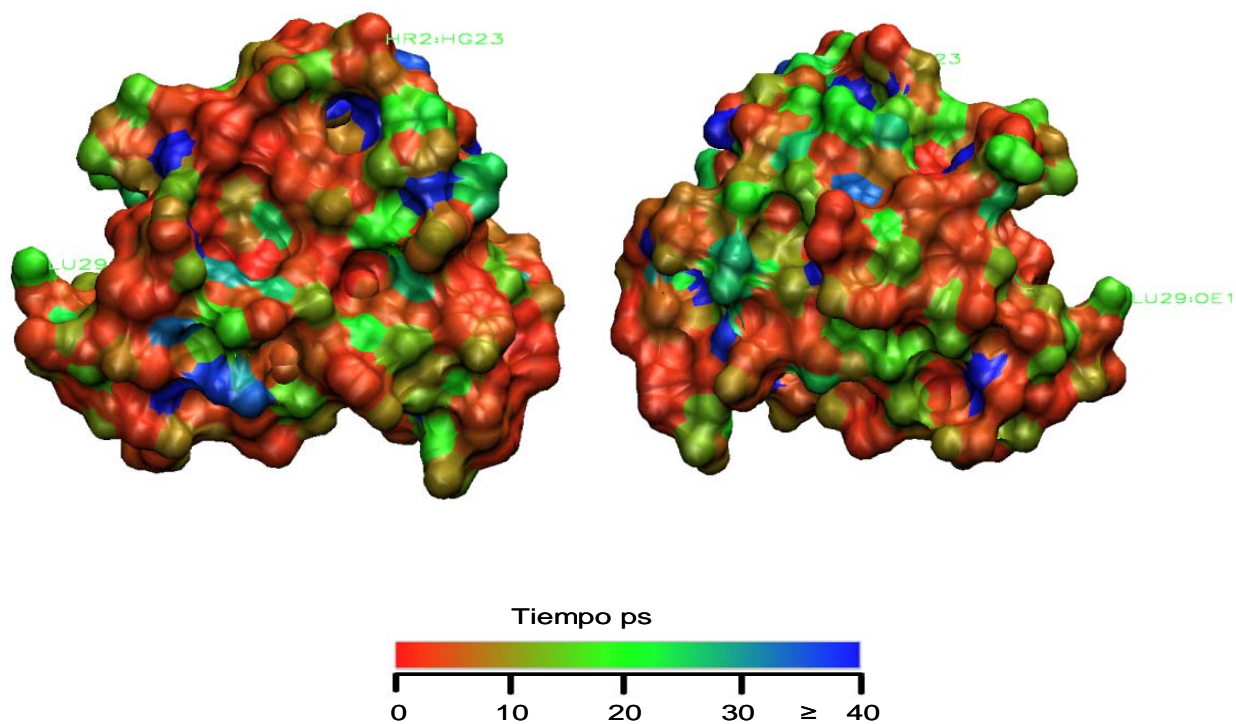


Figura 4.7. Distribución de las moléculas de agua en la proteína tendamistat, de acuerdo al tiempo de residencia de éstas en la proteína, mediante la evaluación de la función de supervivencia, teniendo en cuenta  $\tau_2$  de decaimiento. La escala de colores es un indicador del tiempo de residencia de las moléculas de agua en la primera esfera de solvatación. Esta escala va desde el rojo que corresponde a las moléculas de agua que se intercambian rápidamente con el seno de la solución, al azul que indica altos tiempos de residencia de las moléculas de agua.

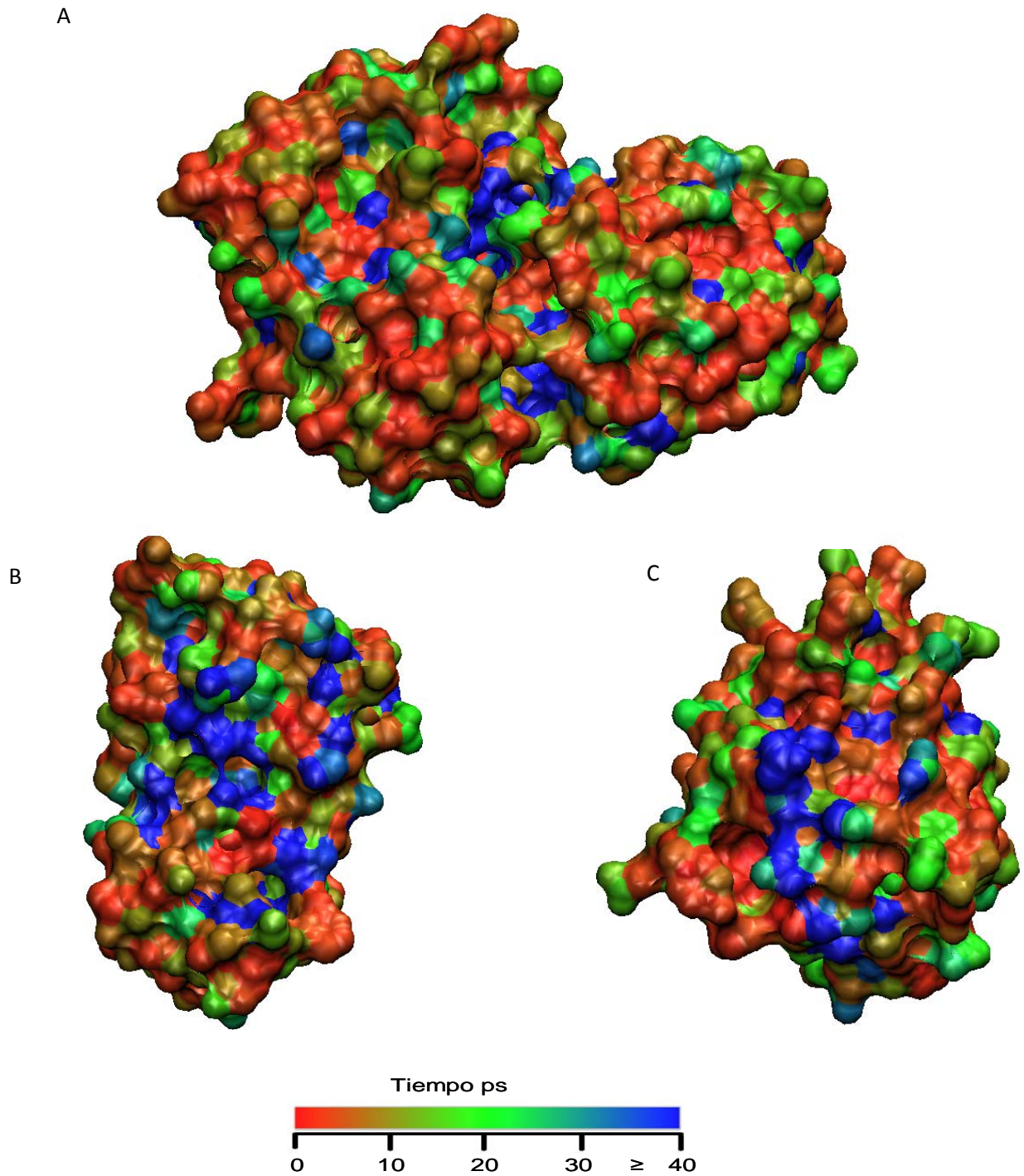


Figura 4.8. Distribución de las moléculas de agua en el complejo, de acuerdo al tiempo de residencia de éstas en la proteína, mediante la evaluación de la función de supervivencia, teniendo en cuenta  $\tau_2$  de decaimiento. La escala de colores es un indicador del tiempo de residencia de las moléculas de agua en la primera esfera de solvatación. Esta escala va desde el rojo que corresponde a las moléculas de agua que se intercambian rápidamente con el seno de la solución, al azul que indica altos tiempos de residencia de las moléculas de agua. En **A** se observa el complejo, en **B** se muestra el lado de la proteína barnasa que interactúa con la proteína barstar, en **C** se muestra el lado de la proteína barstar que interactúa con barnasa.

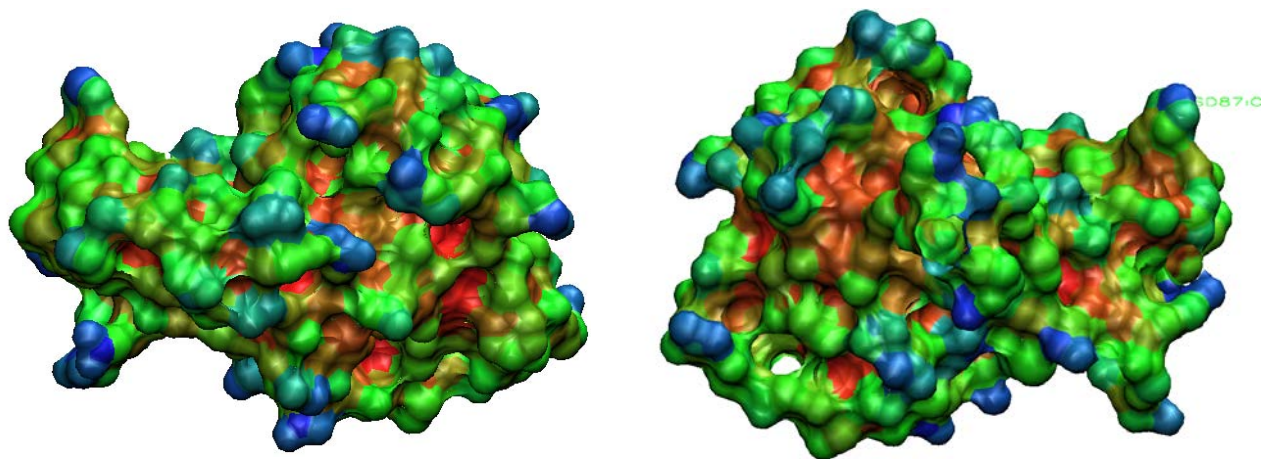
#### 4.1.5. Distribución espacial del número de moléculas de aguas coordinadas

Con el PDB del  $P_\alpha(t=0)$ , arrojado en la ejecución del programa de la función de supervivencia, y el visualizador de VMD, se construyeron los mapas de colores que indican el número promedio de aguas coordinadas a cada átomo de la proteína, ver figura 4.9. La coloración azul, en este caso, advierte que el átomo está rodeado de 6-7 moléculas de agua y la coloración roja que el átomo no se coordina con ninguna molécula de agua. La información que proporcionan estos mapas de colores es homóloga al área de accesibilidad del solvente, que se encarga de describir el área sobre la cual la proteína puede entrar en contacto con el solvente. Sin embargo, en este caso se observa el número de moléculas de agua que pueden estar coordinadas a un átomo como una cantidad promedio a lo largo de una simulación de dinámica molecular. Es importante mencionar que el cálculo del número de coordinación promedio es mucho menos costoso que el cálculo del área expuesta, ya que únicamente requiere del conocimiento de la distancia de las moléculas de agua a cada átomo de la proteína. En estas figuras se aprecia que la mayoría de los átomos expuestos de las proteínas están coordinados por lo menos a 3 ó 4 moléculas de agua.

Los histogramas de los valores de  $P_\alpha(t=0)$  correspondientes a las figuras 4.10 y 4.11, en el primer caso se construyeron sin hidrógenos y en el segundo considerando los valores de  $P_\alpha(t=0) > 0.15$ . Estos histogramas sugieren que la mayoría de los átomos de cada una de las proteínas, en promedio, por lo menos se coordinan con media molécula de agua y muy pocos átomos se encuentran coordinados con 6 ó 7 moléculas de agua, a pesar de las discriminaciones que se hicieron.

Al confrontar los mapas de colores del número de aguas coordinadas, con los mapas de colores de los tiempos de residencia, se observa que la mayoría de las moléculas de agua de alto tiempo de residencia no se ubican en las zonas con el mayor número de aguas coordinadas, es decir en las zonas más expuestas, sino que por el contrario tienden a alojarse en cavidades o grietas, como es el caso de los átomos que presentan los tiempos de residencia más altos para el complejo, la FKBP12 y la tendamistat ver tabla 1, donde se aprecia claramente que ninguno de estos átomos supera el valor de 2.5 moléculas de aguas coordinadas a su esfera de solvatación en  $t = 0$ .

Al revisar los valores de  $P_{\alpha}(t = 0)$  mayores a 4 en las diferentes proteínas, se observa que para la tendamistat, la FKBP12, barnasa libre y barstar libre, los tiempos de residencia no superan los 100 ps, para ninguno de los casos, pero en el caso del complejo hay varios átomos con número de coordinación superior a las 4 moléculas de aguas coordinadas, que tienen tiempos de residencia en el intervalo de 100-500 ps.



Número de aguas coordinadas al átomo  $\alpha$  de una proteína



Figura 4.9. Mapa de colores del número de aguas coordinadas de la proteína FKBP12, considerando el archivo  $P_{\alpha}(t=0)$  (número de aguas coordinadas a un átomo). La escala de colores va desde el rojo que corresponde a átomos que no están rodeados por moléculas de agua, al azul que indica átomos rodeados por 6 ó 7 moléculas de agua.

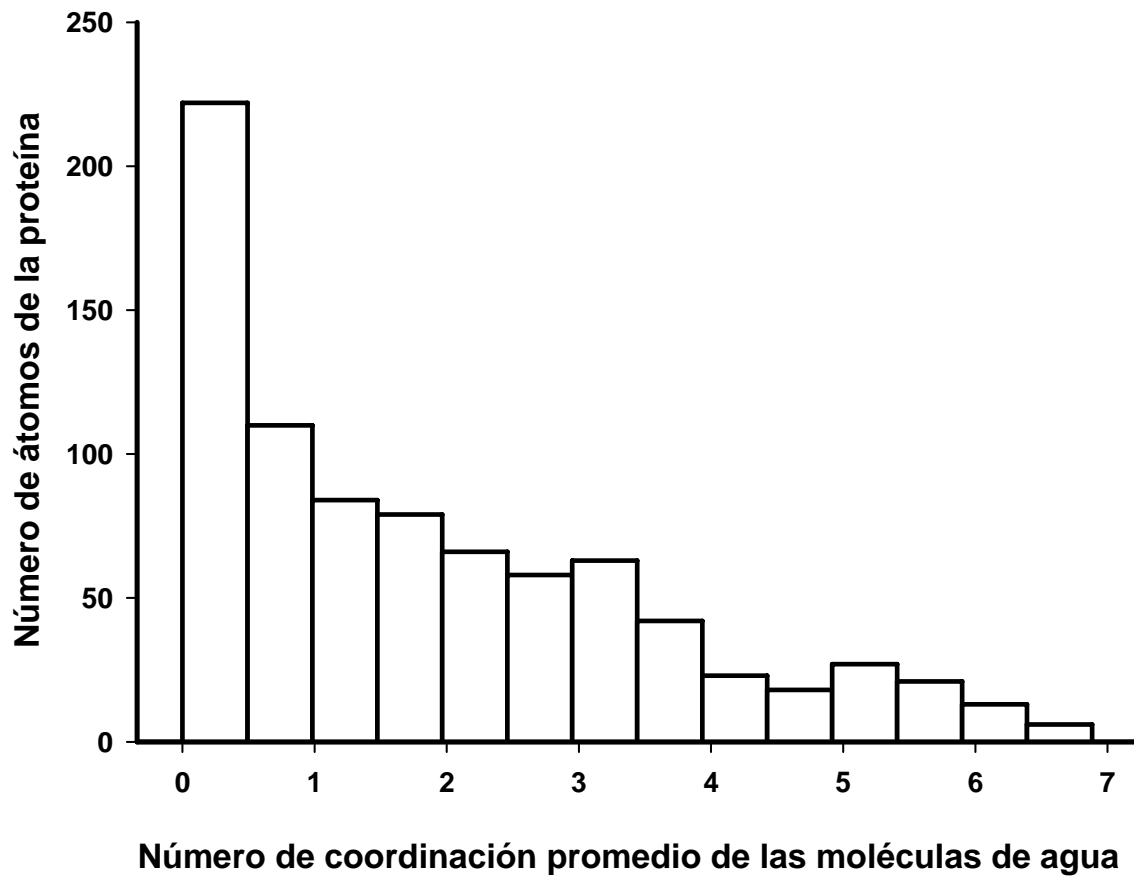


Figura 4.10. Distribución del número de moléculas de agua en la primera esfera de solvatación de los átomos de la proteína FKBP12, sin considerar los átomos de hidrógeno; los datos se obtuvieron del valor de  $P_{\alpha}(t=0)$  (promedio del número de aguas coordinadas a un átomo).



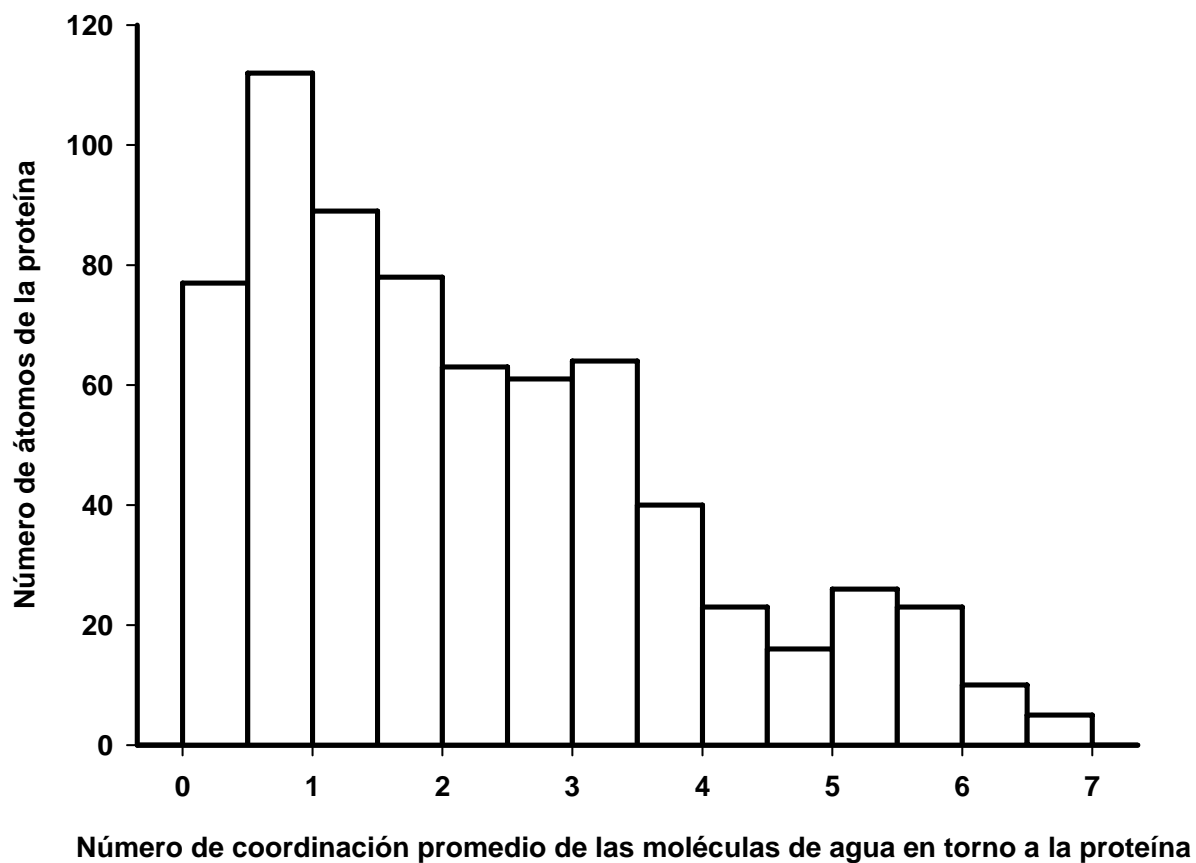


Figura 4.11. Distribución del número de moléculas de agua en la primera esfera de solvatación de los átomos de la proteína FKBP12, considerando los valores de  $P_{\alpha}(t=0) > 0.15$ , ( $P_{\alpha}(t=0)$  número de aguas coordinadas a un átomo).

**Tabla 1:** Algunos átomos del complejo, de la tendamistat y de la FKBP12 que presentan tiempos de residencia altos.

	Índice	átomo	aminoácido	N° del aminoácido	$P_{\alpha}(t=0)$	Tiempo de residencia (ps)
<b>Complejo Barnasa</b>	414	N	LYS	27	1.82	730.73
<b>Complejo Barnasa</b>	535	NE1	TRP	35	1.29	773.02
<b>Complejo Barnasa</b>	775	O	Gly	52	1.255	≥800
<b>Complejo Barstar</b>	2361	OD2	ASP	39	1.98	728.92
<b>Complejo Barstar</b>	2638	N	PHE	56	2.42	763.21
<b>Tendamistat</b>	839	O	GLY	59	1.25	746.68
<b>FKBP12</b>	1415	O	PRO	92	1.14	474.00

#### **4.1.6. Comparación de los resultados de las simulaciones de dinámica molecular con los de rayos X y de RMN de la FKBP12**

Como se mencionó anteriormente, la proteína FKBP12 se eligió para el desarrollo de esta tesis con la finalidad de comparar nuestros resultados teóricos obtenidos de la DM, con los resultados experimentales tanto de rayos X como de RMN. A continuación se describen los experimentos, y se comparan estos resultados con los teóricos.

A través de una serie de estudios<sup>16</sup> de difracción de rayos X se identificaron los sitios de hidratación de 6 complejos de la FKBP12 con diferentes ligandos (FK506, rapamicina, Sb8, Sb13, Sb9, cpFK506h). Este estudio consistió en superponer 8 estructuras, las cuales se obtuvieron de complementar las 6 estructuras iniciales con moléculas de agua relacionadas simétricamente y con los átomos polares de la proteína que se ubican dentro de una capa de 4 Å. A partir de la superposición, se identificaron 251 sitios de hidratación, los cuales se agruparon en clases (2/8, 3/8, ..., 8/8) que reflejan el número de estructuras en las cuales cada sitio de hidratación se observó, es decir: 11 sitios de hidratación son comunes a las 8 estructuras (clase 8/8), 10 a 7 estructuras (clase 7/8), 11 a 6 estructuras (clase 6/8), 5 a 12 estructuras (clase 5/8), 16 a 4 estructuras, (clase 4/8) 55 a 3 estructuras (clase 3/8) y 136 a 2 estructuras (clase 2/8).

Los sitios de hidratación de las clases 8/8 a la 4/8 los seleccionamos para confrontar nuestros resultados porque éstos son independientes de la cristalización. Es importante tener en cuenta que durante la cristalización existe la posibilidad de que átomos polares de la proteína se ubiquen en las zonas de hidratación de las macromoléculas en solución. Los investigadores en este estudio minimizaron ésta posibilidad y ésta es la razón por la cual las clases 8/8 a la 4/8 son una buena representación de las moléculas de agua estructurales, las cuales los autores han denominado como "sitios de hidratación de consenso". La figura 4.12 muestra los

átomos de las clases 8/8 a 4/8 con respecto a los tiempos de residencia de la primera capa de solvatación de cada uno de estos átomos y se encontró que no todas las moléculas de agua identificadas en rayos X corresponden a aguas de intercambio lento. De los 84 átomos correspondientes a los sitios de hidratación identificados en rayos X (átomos de la proteína formando puentes de hidrógeno con alguna de las aguas cristalográficas de consenso, es decir, de las clases 8/8 a 4/8), únicamente 23 corresponden a coloraciones azules en los mapas de colores (tiempos de residencia mayores a 30 ps) y 31 átomos a coloraciones verdes (14 ps-30 ps).

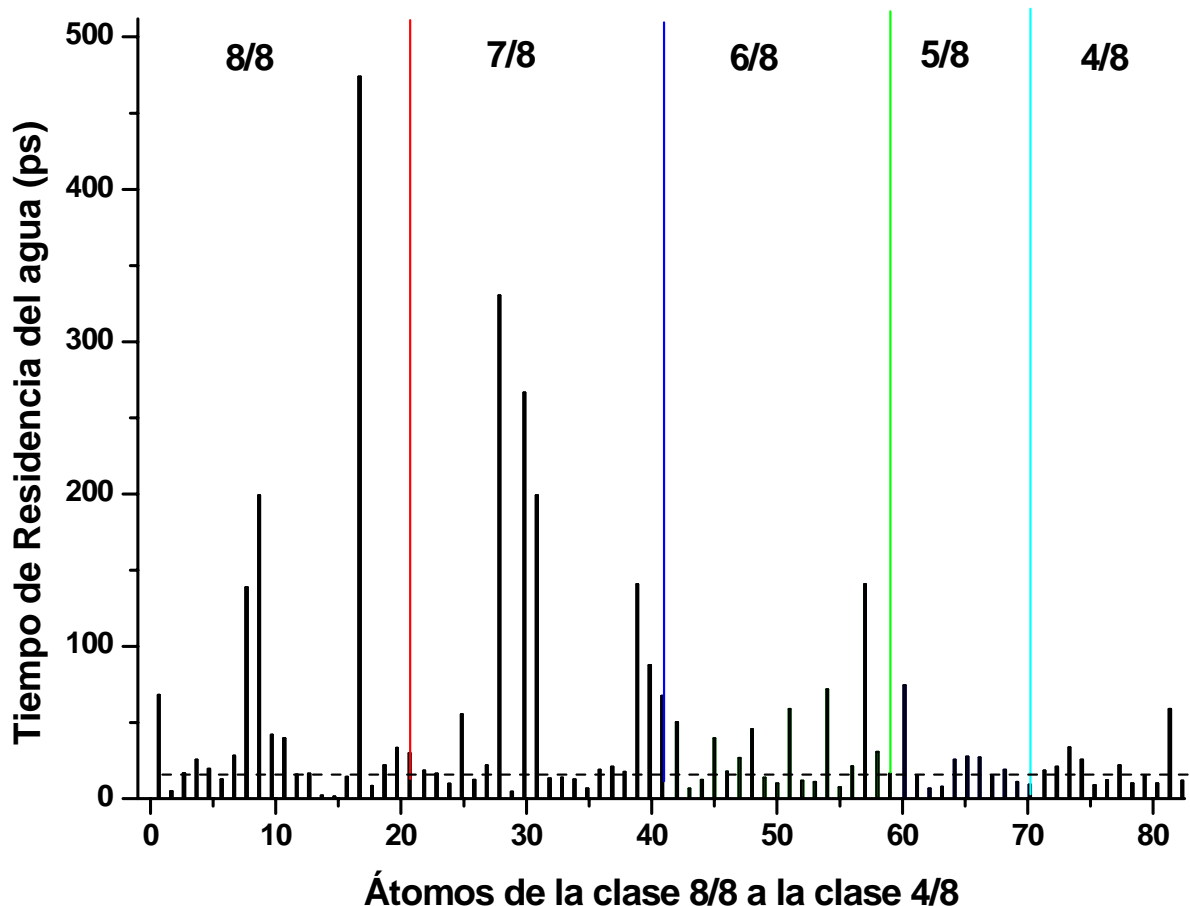


Figura 4.12. Gráfica del tiempo de residencia del agua. Comparación de los sitios de hidratación de la simulación de DM de la FKBP12, considerando  $\tau_2$  de decaimiento vs los átomos correspondientes a las clases de los sitios de hidratación reportados por Karplus<sup>16</sup>. La línea punteada representa el tiempo promedio de permanencia de las moléculas de agua en la esfera de solvatación, cuando  $P_\alpha(t=0) > 0.15$ .

En un estudio del comportamiento de las moléculas de agua de la FKBP enlazada al inmunosupresor FK506 y a la ascomicina a través de RMN-NOE 3D, se encontró que hay 3 moléculas de agua (moléculas de agua N° 1, 2 y 4) fuertemente enlazadas en el interior del complejo que forman parte de su estructura, a través de enlaces puentes de hidrógeno<sup>38</sup>. Estas moléculas de agua se encuentran distribuidas así: La molécula de agua 1 se ubica entre las amidas de la Gly86, His94, Arg95 y los protones  $\alpha$  de la Pro93, y  $\beta$  de la Tyr82. La molécula de agua 2 entre las amidas de la Tyr82, Gly83 y los protones  $\beta$  de la Tyr82,  $\alpha$  de la Pro93 y  $\gamma$  de la Ile91 y la molécula de agua 4 (de la numeración original del artículo) entre las amidas Met49, Lys52, Glu54, Val55 y los protones  $\delta$  de la Phe48 y  $\gamma$  de la Val55. Las anteriores moléculas de agua se han identificado en la estructura de rayos X FKBP/ascomicina. Los sitios de cada una de las moléculas de agua anteriormente localizados, se analizaron teniendo en cuenta el mapa de colores de la FKBP12 (ver figura 4.7) así: Para este análisis se tuvieron en cuenta los átomos de N del grupo amido de cada uno de los aminoácidos anteriormente mencionados y a partir del átomo de N previamente seleccionado, se identificaron los átomos que están a menos de 5 Å de distancia. Se consideró esta distancia para el análisis, puesto que los experimentos NOESY o ROESY implican acoplamiento espaciales en este rango. El análisis considerando los mapas de colores, arrojó los siguientes resultados:

En la figura 4.13 se observa que los átomos que conforman el sitio de ubicación de la molécula de agua N°1, se caracteriza por tener altos tiempos de residencia. De los N de la cadena principal de los aminoácidos Gly86, His94 y Arg95 tan sólo el primero presenta un tiempo de residencia superior a los 30 ps, que se ve favorecido por 3 átomos muy cercanos que presentan tiempos de residencia superiores al promedio de los átomos (tiempos de residencia mayores a los 14 ps).

Los N de los residuos His94 y Arg95 no tienen tiempos de residencia relevantes, pero en su entorno hay átomos con altos tiempos de residencia, de hecho comparten dos átomos; uno con un tiempo de residencia mayor a los 30 ps (Pro92: C) y el otro con un tiempo de residencia de aproximadamente de 18 ps (Pro93:O).

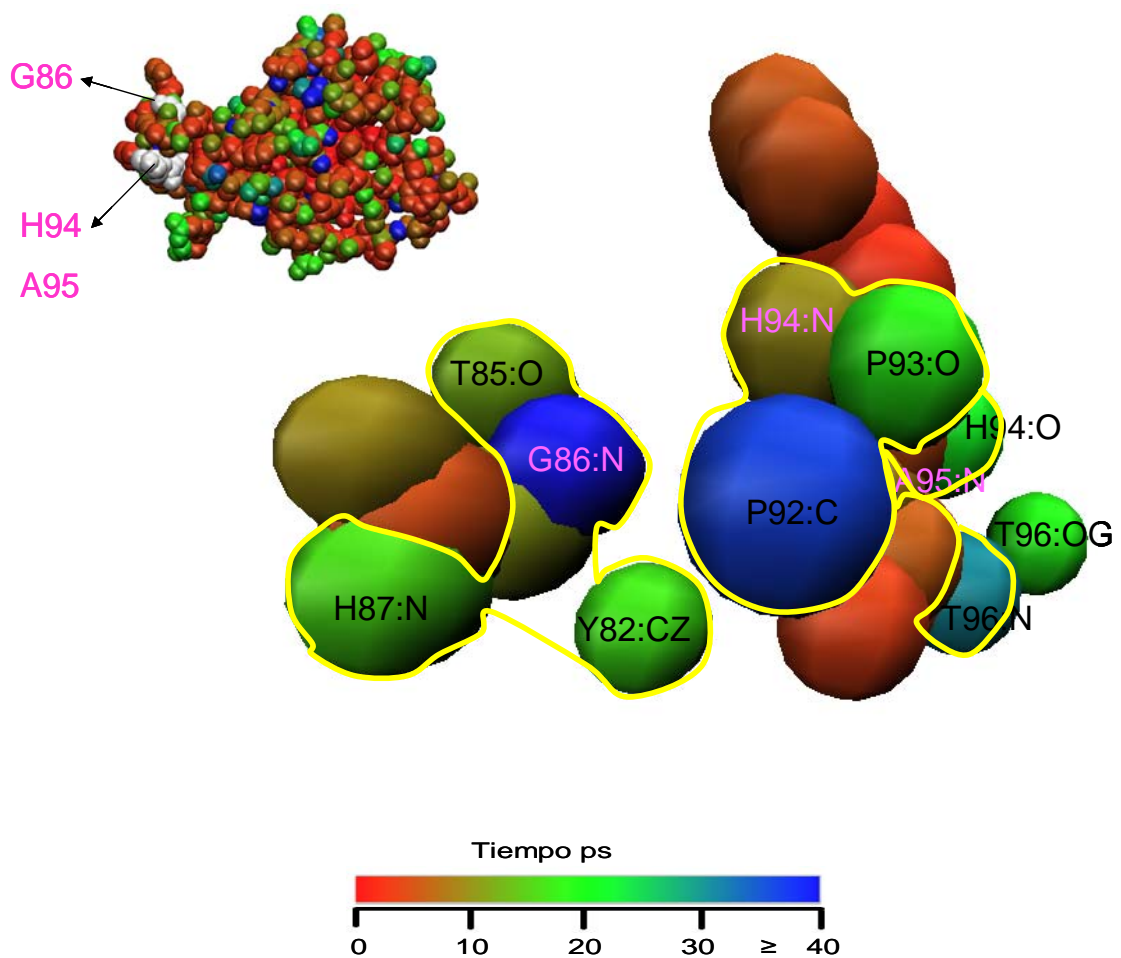


Figura 4.13. Identificación de los átomos donde se ubica la molécula de agua N°1 en el mapa de colores de la proteína FKBP12, considerando  $\tau_2$  de decaimiento. En amarillo se encuentran encerrados los átomos que están a menos de 5 Å de los N de la cadena principal de Gly86, His94 y Ala95. La escala de colores es un indicador del tiempo de residencia de las moléculas de agua en la primera esfera de solvatación. Esta escala va desde el rojo que corresponde a las moléculas de agua que se intercambian rápidamente con el seno de la solución, al azul que indica altos tiempos de residencia de las moléculas de agua.

Ninguno de los nitrógenos de la cadena principal de los aminoácidos Tyr82 y Gly83 relacionados con la molécula de agua N°2, presentan tiempos de residencia superiores al tiempo promedio (tiempos de residencia de 14 ps); sin embargo, los átomos de su entorno tienen tiempos de residencia muy significativos ver figura 4.14.

De los 4 N de la cadena principal de los aminoácidos Lys52, Met49, Glu54 y Val55 reportados por RMN (molécula de agua N°4) los 3 últimos presentan tiempos de residencia mayores a los 30 ps, mientras que el N de Lys52 no tiene un tiempo de residencia importante, pero a 2.4 Å de distancia su CB tiene un tiempo de residencia dentro del promedio (14 ps) al igual que el oxígeno de la Gly51 y el nitrógeno de la Gln53 que se encuentran a menos de 5 Å de distancia ver figura 4.15.

En el espectro de RMN de las cadenas laterales, aparece que el CD de la Phe48 participa en la ubicación del agua N°4, en nuestro caso se aprecia el CA, que está a 2.43 Å del N de la Met49. El N de Met49 también está muy cerca de su oxígeno que presenta coloración azul, y a su vez éste está próximo al CB de la Lys52 con un tiempo de residencia superior al promedio de las moléculas de agua. Los nitrógenos del Glu54 y de la Val55 relacionados con la molécula de agua N°4, están muy cerca uno del otro, de hecho hay varios átomos en su entorno de coloración relevante que comparten ver figura 4.15.

Es evidente que los sitios en los que se ubican las 3 moléculas de agua reportadas por RMN-NOE 3D coinciden con las zonas de la proteína que presentan tiempos de residencia altos. Hay algunos casos donde se aprecia que los átomos reportados por RMN no coinciden con los observados en los mapas de colores, pero sus átomos más cercanos sí lo hacen. Esta consideración es válida ya que los experimentos NOESY y ROESY implican acoplamientos espaciales o bipolares en un rango de 5 Å.

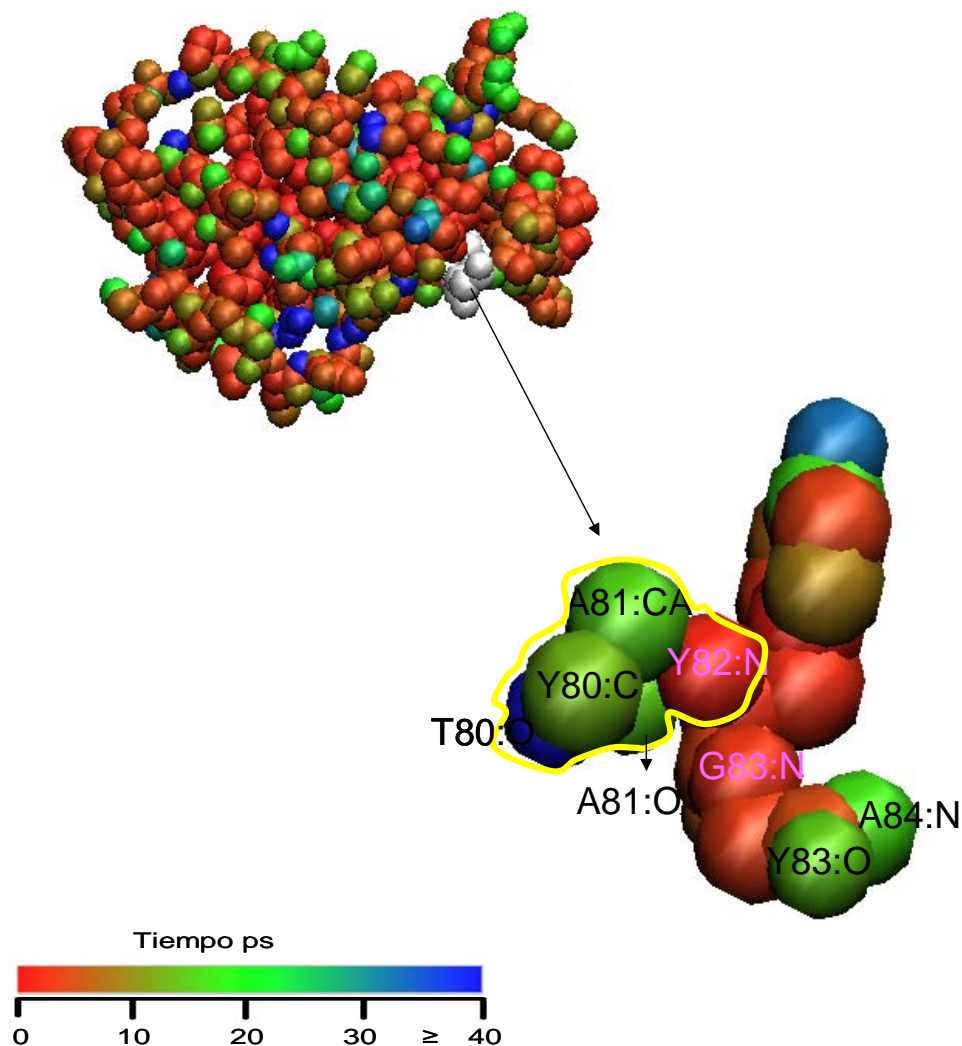


Figura 4.14. Identificación de los átomos donde se ubica la molécula de agua N°2 en el mapa de colores de la proteína FKBP12, considerando  $\tau_2$  de decaimiento. En amarillo se encuentran encerrados los átomos que están a menos de 5 Å del N de la cadena principal de la Tyr82. Los átomos de color verde sin encerrar están próximos al N de la cadena principal de la Gly83, al igual que el C $\alpha$  de la Ala81 y el C del grupo carbonilo de la Tyr80. La escala de colores es un indicador del tiempo de residencia de las moléculas de agua en la primera esfera de solvatación. Esta escala va desde el rojo que corresponde a las moléculas de agua que se intercambian rápidamente con el seno de la solución, al azul que indica altos tiempos de residencia de las moléculas de agua.



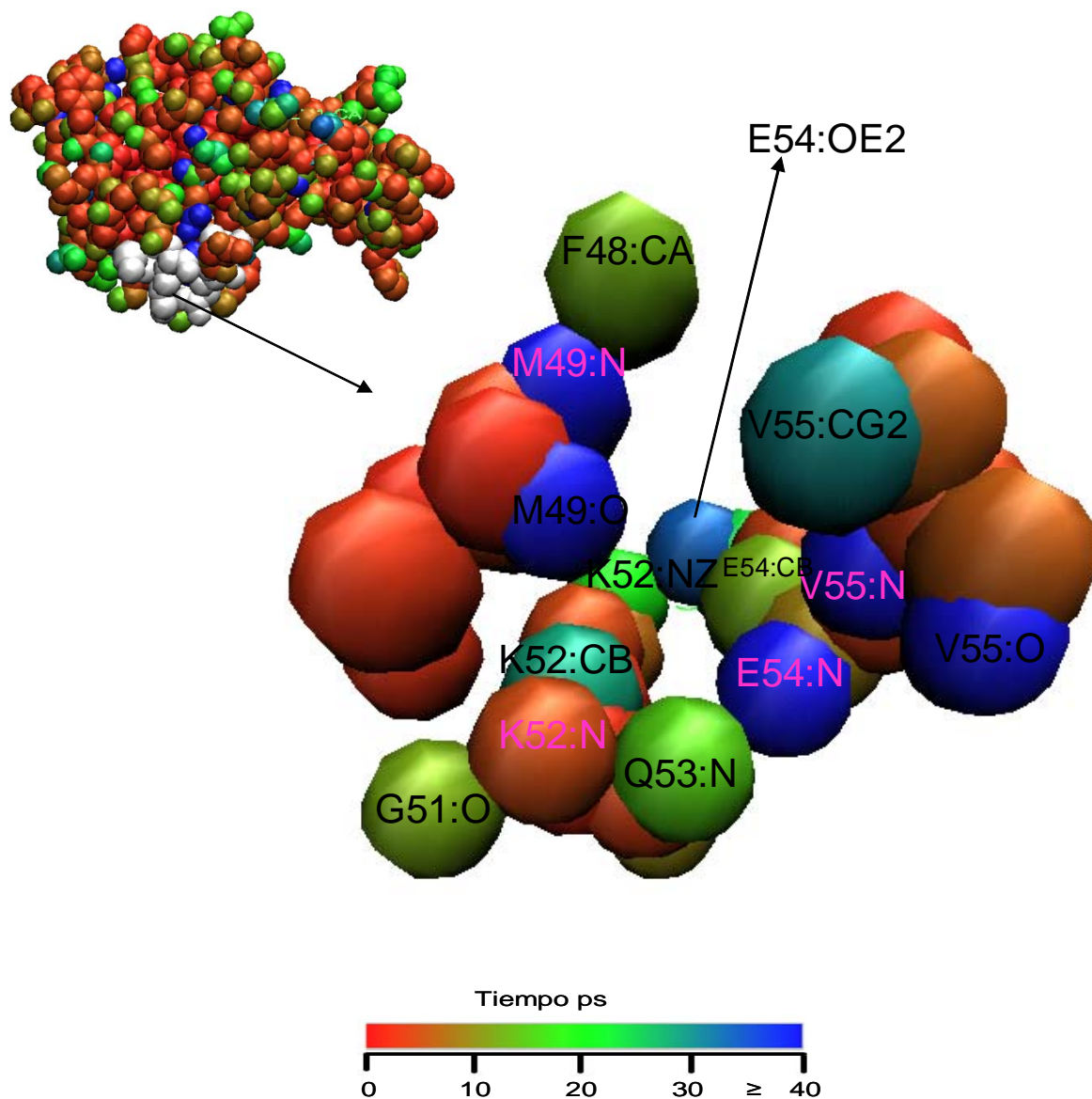


Figura 4.15. Identificación de los átomos donde se ubica la molécula de agua N°4 en el mapa de colores de la proteína FKBP12, considerando  $\tau_2$  de decaimiento. En fucsia aparecen los N de la cadena principal de los residuos Met49, Lys52, Glu54 y Val55. La escala de colores es un indicador del tiempo de residencia de las moléculas de agua en la primera esfera de solvatación. Esta escala va desde el rojo que corresponde a las moléculas de agua que se intercambian rápidamente con el seno de la solución, al azul que indica altos tiempos de residencia de las moléculas de agua.

Aquí vale la pena aclarar que no se sabe si exactamente estos sitios corresponden a una o más moléculas de agua, solo se puede afirmar que son zonas de alto tiempo de residencia de las moléculas de agua.

En los estudios que se hicieron de la FKBP con RMN 3D, también se identificaron las moléculas de agua 11 y 13, que no aparecen en la estructura de rayos X FKBP/ascomicina. La molécula de agua 11 está asociada al grupo amido Lys47 y la molécula 13 al grupo amido Glu31. Al analizar estas moléculas de agua en los mapas de colores encontramos que los átomos tanto de la Lys47 como de la Glu31 y sus alrededores, (menor a 5 Å) en su mayoría presentan tiempos de residencia de 14 ps, sólo hay un átomo de tiempo de residencia superior a los 30 ps la Met49:N cercano a la Lys47:N. De acuerdo a esto se puede decir que en esta zona se ubican moléculas de agua de la superficie, las cuales se intercambian en un tiempo promedio con el seno de la solución, pero que se lograron identificar en RMN debido a la proximidad de los protones de las amidas con las moléculas de agua.

Las diferentes zonas de coloración azul en los mapas de colores de la FKBP12 abarca un total de 41 átomos que presentan un tiempo de residencia mayor a los 30 ps ver figura 4.7. De este grupo sólo Gln53 y Gln54 corresponden al sitio activo de la proteína.

#### **4.1.7.Comparación de los resultados de la simulación de la dinámica molecular con los de RMN de la proteína tendamistat**

La idea de esta sección es confrontar los resultados de la simulación de dinámica molecular con los resultados experimentales de la proteína tendamistat. Para ello primero hacemos una descripción del experimento, luego cada uno de los casos reportados experimentalmente y que son de nuestro interés, los analizamos de acuerdo a los resultados de la simulación de dinámica molecular y posteriormente comparamos los resultados experimentales con los obtenidos teóricamente.

Scarselli y colaboradores analizaron la accesibilidad de los átomos superficiales de la proteína tendamistat con el co-soluto paramagnético TEMPOL, con la finalidad de ubicar experimentalmente los sitios superficiales de elevada interacción agua-proteína. La idea de estos experimentos es que los residuos que están expuestos al solvente, pero que no tienen contacto con el co-soluto paramagnético, corresponden a casos donde las moléculas de agua sirven de “escudo”, pues éstas, al estar fuertemente unidas a la superficie, impiden la transferencia de magnetización por parte del co-soluto paramagnético a los átomos de la proteína. En este estudio se reportaron 4 casos en los cuales se observan residuos altamente expuestos que no interactúan con el co-soluto paramagnético estos son: Pro50 (conectividades de los hidrógenos  $\gamma$ - $\delta$ ), Thr46 (conectividades de los hidrógenos  $\delta$ - $\epsilon$ ), Arg49 (conectividades de los hidrógenos  $\alpha$ - $\beta$ ), Thr3 (conectividades de los hidrógenos  $\alpha$ - $\gamma$ ) y Thr54 (conectividades de los hidrógenos  $\alpha$ - $\gamma$ ). Como estos casos corresponden a los sitios superficiales de elevada interacción agua-proteína, la idea aquí es analizar estos casos para la simulación de dinámica molecular y posteriormente confrontarlos para distinguir si en efecto la simulación de DM predice la existencia de moléculas de agua fuertemente adheridas a la proteína en dichas zonas y a las que se les pueda atribuir el efecto de “escudo”. Para comparar los sitios de los resultados de Scacelli y cols. con los resultados de la función de supervivencia, se tuvieron en cuenta solo los átomos que se encuentran a menos de 5 Å de distancia de los hidrógenos correspondientes a las conectividades anteriormente mencionadas. Esta distancia se consideró porque los experimentos de transferencia de magnetización implican acoplamientos espaciales, los cuales se dan a menos de 5 Å de distancia.

En la figura 4.16 se muestran los posibles puentes de hidrógeno que pueden formar los HE1 y los HD1 de la Tyr46. En los dos casos la formación de los puentes de hidrógeno se da entre átomos que presentan altos tiempos de residencia. Este comportamiento también se representó en una gráfica de barras, figura 4.17, en la cual

se muestran los diferentes átomos que pueden formar los puentes de hidrógeno con los HE1 y los HD1 de la Tyr46; la línea punteada indica el tiempo promedio  $\langle \tau_2 \rangle$  de permanencia de las moléculas de agua dentro de la esfera de solvatación de esta proteína, considerando  $P_\alpha(t=0) > 0.15$ .

Al hacer el mismo análisis para Pro50, Thr3 y Thr54 se encontró que los diferentes protones de estos aminoácidos tienden a enlazarse con átomos que superan el tiempo de residencia promedio de las moléculas de agua, ver figuras 4.18-4.20. En el caso de la Arg49, no se encontraron átomos cercanos que presenten altos tiempo de residencia y con los cuales pueda formar puentes de hidrógeno; pero contrario a todas las demás situaciones, la Arg49 tiene un átomo de alto tiempo de residencia. El análisis de la función de supervivencia para los residuos Thr46, Pro50, Thr3, Thr54 y Arg49 nos permite concluir que los aminoácidos Thr46, Thr3 y Arg49 corresponden a zonas donde las moléculas de agua tienden a permanecer largo tiempo, mientras que en los aminoácidos Thr54 y Pro50 se ubican moléculas de agua que se intercambian rápidamente con el seno de la solución, pero mucho menos rápido que una molécula de agua del bulto.

Al comparar los resultados experimentales de RMN de la proteína tendamistat con los obtenidos teóricamente, se aprecia una buena concordancia, pues es evidente que la mayoría de los residuos anteriormente citados corresponden a zonas donde las moléculas de agua no se intercambian rápidamente con el seno de la solución.

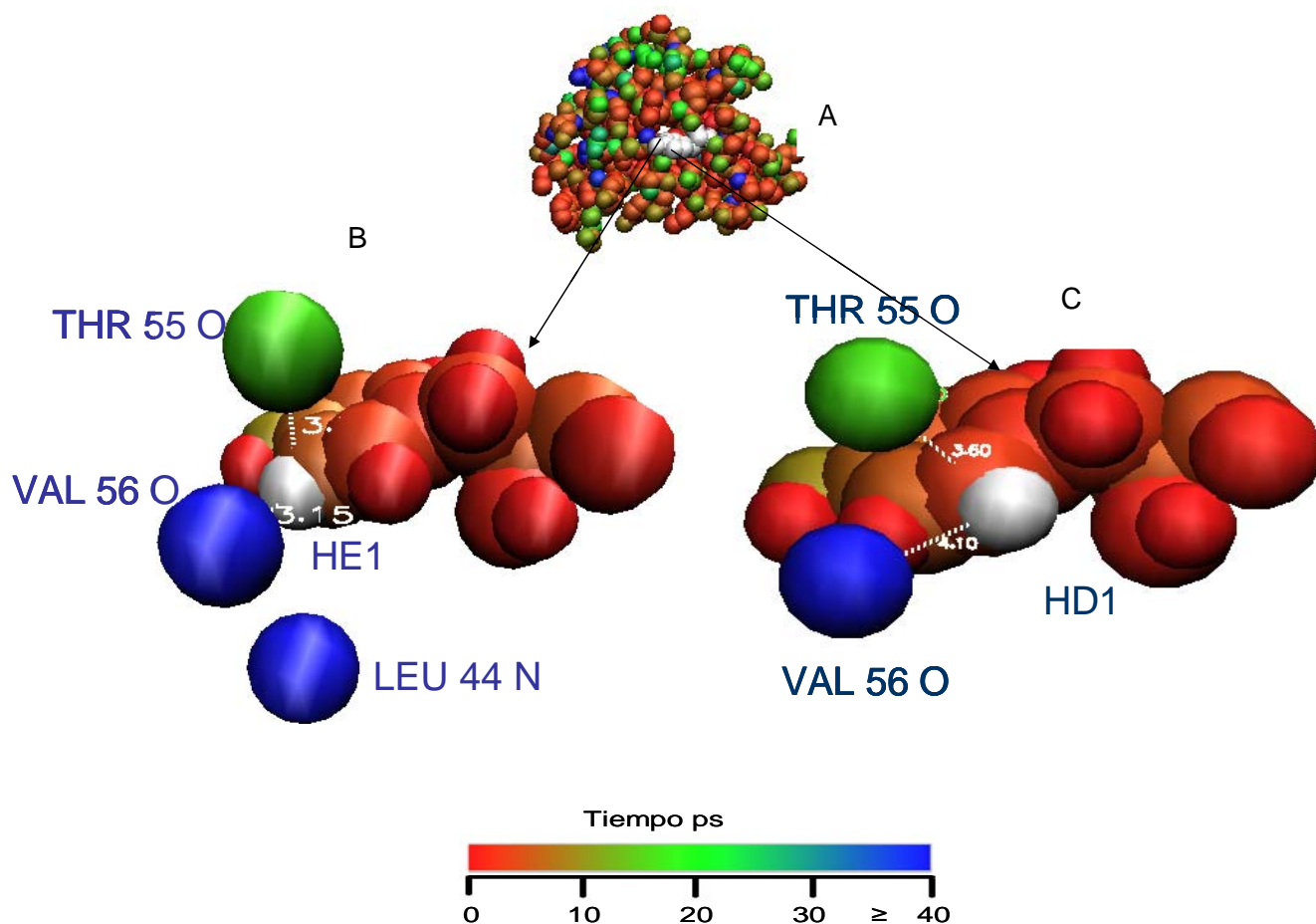


Figura 4.16. Comparación de la conectividad  $\delta - \varepsilon$  de la Tyr46 reportada por scarselli con los resultados de la función de supervivencia, considerando  $\tau_2$  de decaimiento. En **A** se identifica la Tyr46 dentro de la proteína tendamistat. En **B** se muestran los posibles puentes de hidrógenos que puede formar el HE1 (protón blanco) de la Tyr46 y en **C** los posibles puentes de hidrógeno que puede formar el HD1 (protón blanco) de la Tyr46. La escala de colores es un indicador del tiempo de residencia de las moléculas de agua en la primera esfera de solvatación. Esta escala va desde el rojo que corresponde a las moléculas de agua que se intercambian rápidamente con el seno de la solución, al azul que indica altos tiempos de residencia de las moléculas de agua.

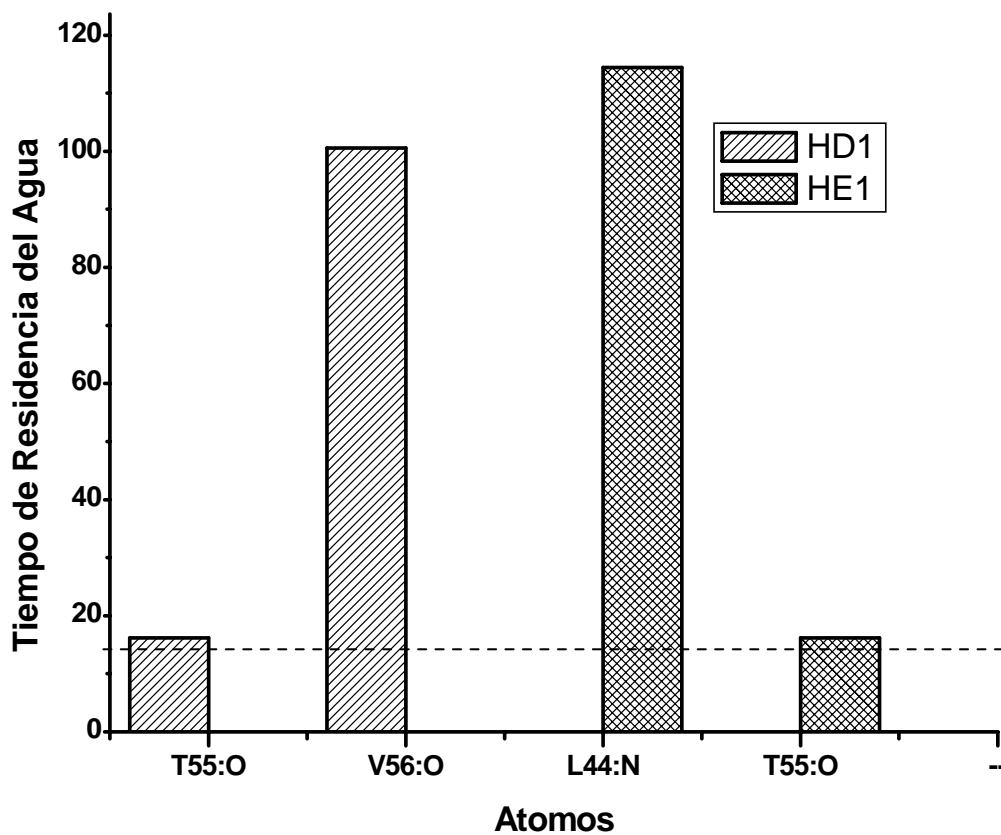


Figura 4.17. Diagrama de barras en el que se muestran los átomos que pueden formar puentes de hidrógeno con los HD1 y HE1 de la Tyr46 de la proteína tendamistat a una distancia inferior a los 5 Å, considerando  $\tau_2$  de decaimiento. La línea punteada indica el tiempo promedio de permanencia de las moléculas de agua dentro de la primera esfera de solvatación, considerando  $P_\alpha(t=0) > 0.15$ .

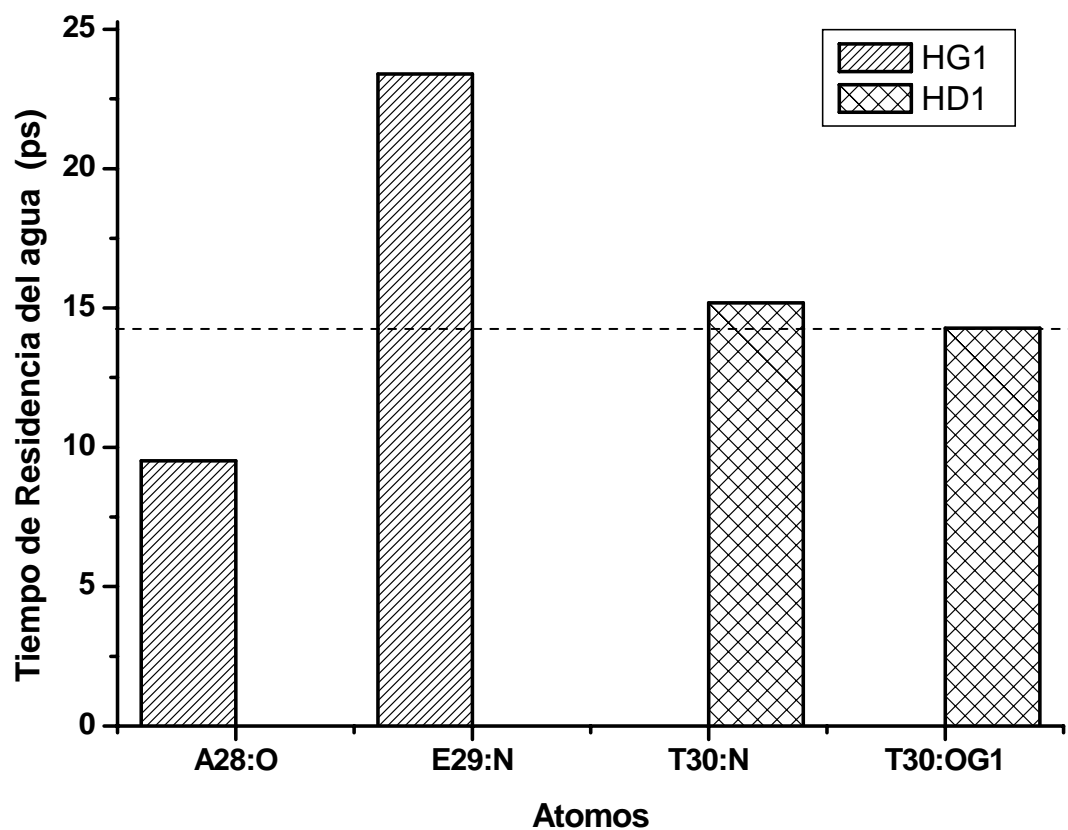


Figura 4.18. Diagrama de barras en el que se muestran los átomos que pueden formar puentes de hidrógeno con los HG1 y HD1 de la Pro50 de la proteína tendamistat a una distancia inferior a los 5 Å, considerando  $\tau_2$  de decaimiento. La línea punteada indica el tiempo promedio de permanencia de las moléculas de agua dentro de la primera esfera de solvatación, considerando  $P_\alpha(t=0) > 0.15$ .

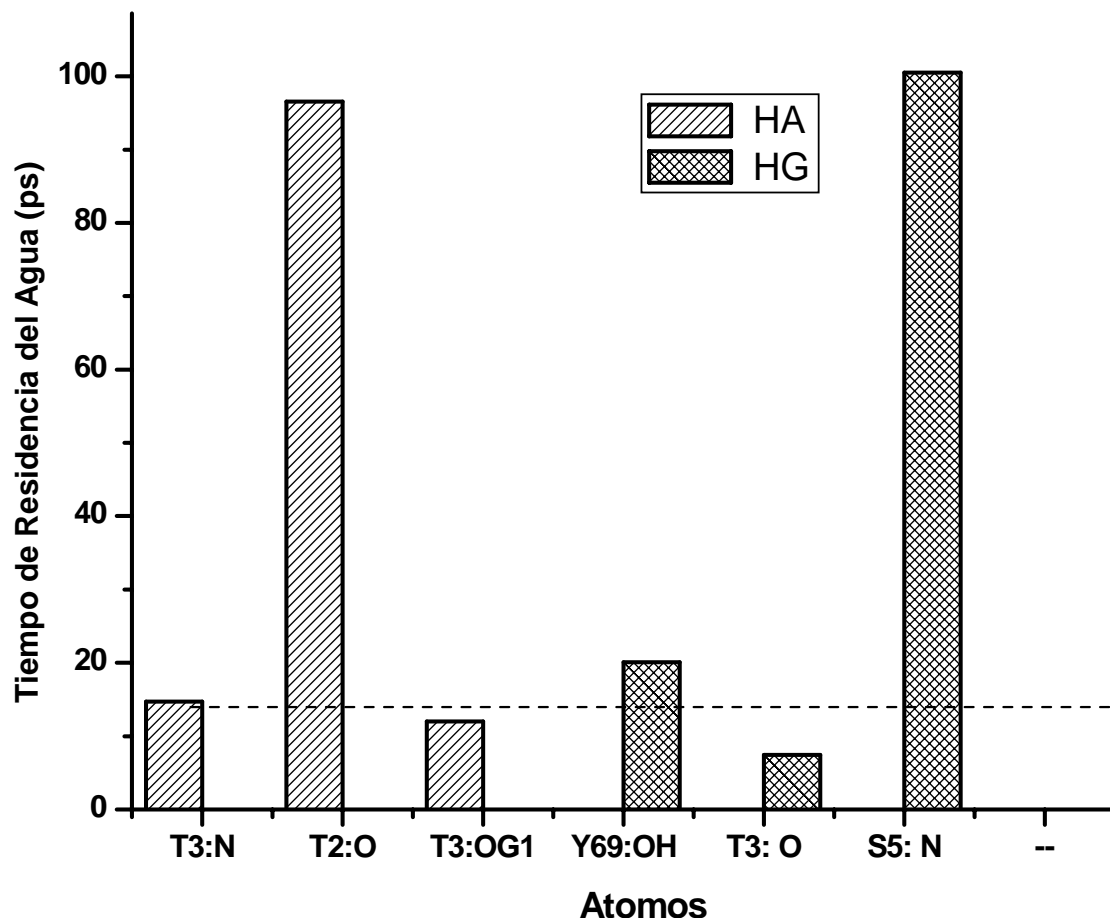


Figura 4.19. Diagrama de barras en el que se muestran los átomos que pueden formar puentes de hidrógeno con los HA y HG de la Thr3 de la proteína tendamistat a una distancia inferior a los 5 Å, considerando  $\tau_2$  de decaimiento. La línea punteada indica el tiempo promedio de permanencia de las moléculas de agua dentro de la primera esfera de solvatación, considerando  $P_\alpha(t=0) > 0.15$ .



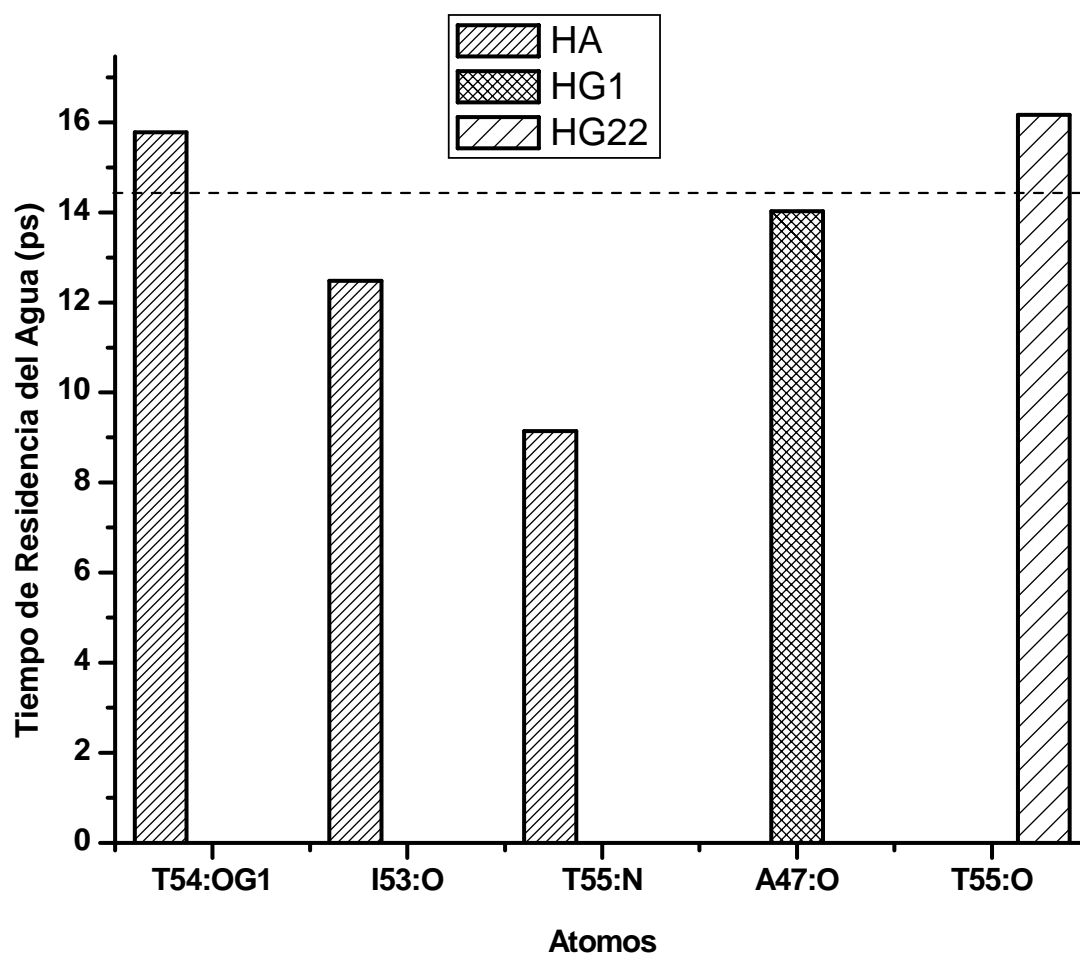


Figura 4.20. Diagrama de barras en el que se muestran los átomos que pueden formar puentes de hidrógeno con los HA, HG1 y HG22 de la Thr54 de la proteína tendamistat a una distancia inferior a los 5 Å, considerando  $\tau_2$  de decaimiento. La línea punteada indica el tiempo promedio de permanencia de las moléculas de agua dentro de la primera esfera de solvatación, considerando  $P_\alpha(t=0) > 0.15$ .

Existe un total de 31 átomos de la proteína tendamistat, en su mayoría N y O con funciones  $P_{\alpha}(t)$  de decaimiento lento  $\tau_2 > 30$  ps. El sitio de enlace de la tendamistat está mediado por cuatro segmentos los cuales involucran un total de 15 aminoácidos. El primer segmento lo integran Trp18, Arg19, Tyr20 y Tyr15. De estos cuatro aminoácidos solo los tres últimos presentan un átomo, en este caso el oxígeno del grupo carbonilo, con alto tiempo de residencia, cercanos a los 40 ps (coloración azul) y el aminoácido restante Trp18 tiene un átomo de nitrógeno (de la cadena principal) cuyo tiempo de residencia es de 19 ps (coloración verde). Como se puede notar estos 4 aminoácidos representan un sitio de alto tiempo de residencia para las moléculas de agua y además se sabe de ellos, que juegan un papel importante en las interacciones fuertes entre la tendamistat y la amilasa de mamífero<sup>43</sup>.

#### **4.1.8. Comparación de los resultados de la simulación de la dinámica molecular del complejo barnasa-barstar con los resultados de la simulación de la dinámica molecular de las proteínas libres**

El complejo barnasa-barstar es un sistema de particular interés debido a que la interacción entre las dos proteínas está mediada por moléculas de agua, de hecho aparece en la literatura que el grado de complementariedad entre las dos proteínas aumenta por la presencia de las moléculas de agua<sup>37</sup>. Es de nuestro interés identificar las moléculas de agua que participan en el acoplamiento de las dos macromoléculas, al igual que identificar aquellas moléculas de agua que se mantienen asociadas a una u otra proteína desde antes de que se forme el complejo. Esta es la razón por la cual a continuación se comparan los resultados del análisis de las trayectorias del complejo y de las proteínas libres entre sí, e igualmente se confrontan con los resultados cristalográficos existentes.

Para el análisis de la interfase entre barnasa-barstar; se tomaron de los mapas de colores (ver figura 4.8) los átomos que presentan coloración azul ( $\tau_2 > 30$  ps) y que se encuentran entre las dos proteínas (el borde de cada una), obteniéndose de esta

forma un total de 42 átomos para barnasa y 25 átomos para barstar, ver tablas 2 y 3. El mismo procedimiento se hizo con las proteínas libres de barnasa, barstar y se obtuvieron 22 sitios de hidratación para barnasa y un sitio de hidratación (Glu46:O) para barstar. Al confrontar los átomos de la interfase que presentan altos tiempos de residencia, obtenidos de la simulación de DM del complejo, de la simulación de DM de barnasa libre y de la simulación de DM de barstar libre; se encontró que 13 átomos son comunes para barnasa mientras que para barstar no hay ningún átomo que coincida.

Posteriormente estos datos se compararon con las moléculas de agua cristalográficas reportadas por Buckle y colaboradores, ver tabla 4. La comparación entre las moléculas de agua reportadas por Buckle con las obtenidas de la trayectoria de simulación de DM del complejo indicó que hay 10 átomos en común en los que se ubican moléculas de agua de alto tiempo de residencia en barnasa y 5 átomos en barstar. Al hacer lo mismo con las trayectorias de las proteínas libres, se hallaron 3 sitios de hidratación para barnasa, correspondientes a los átomos Arg83:N, Lys62:N y Asn58:OD1 y ninguno para barstar. Los átomos Arg83:N, Lys62:N y Gln58:OD1 de barnasa se identificaron tanto en la simulación de DM del complejo, como en la simulación de DM de la proteína libre y en los sitios cristalográficos reportados por Buckle.

La literatura<sup>59</sup> reporta que hay tres residuos de barnasa que se conservan en toda la familia de ribonucleasas, éstos son: Arg87, Glu73 y His102 y se caracterizan por formar parte del sitio catalítico de la proteína. Estos aminoácidos en el mapa de colores del complejo tienen átomos cuyos tiempos de residencia son superiores a los 30 ps, pero en el mapa de colores de la proteína libre, sólo presentan esta particularidad los dos últimos. La Lys27 es un aminoácido que forma parte del sitio activo de la proteína y se caracteriza por estabilizar al grupo fosfato en el acoplamiento proteína-sustrato. Este residuo en el mapa de colores tiene dos átomos de color azul,

la Lys27:N con un tiempo de residencia de 730 ps y Lys27:NZ con un tiempo de residencia de 54 ps, de acuerdo a lo anterior, podemos inferir de estos 4 residuos que su actividad catalítica se ve favorecida por la presencia de moléculas de agua.

**Tabla 2:** Átomos de la interfase de la proteína barnasa que presentan altos tiempos de residencia, obtenidos de la simulación de DM del complejo, considerando  $\tau_2$ . de decaimiento. En gris aparecen los átomos de alto tiempo de residencia que coinciden con los resultados de la simulación de DM de barnasa libre.

Índice del átomo	Tipo de átomo	Residuo	N° de residuo	Tiempo (ps)
869	CD	ARG	59	36.05
872	NE	ARG	59	114.82
875	NH1	ARG	59	148.54
1571	NE2	HSD	102	368.17
1572	CD2	HSD	102	33.68
1575	O	HSD	102	42.59
436	N	SER	28	76.08
1240	CB	ARG	83	78.69
1243	CG	ARG	83	36.04
1255	NH2	ARG	83	100.94
575	N	SER	38	230.76
565	N	ALA	37	123.99
574	O	ALA	37	91.96
1589	OH	TYR	103	104.81
1596	O	TYR	103	80.63
844	O	SER	57	503.92
1103	CD	GLU	73	577.89
1104	OE1	GLU	73	350.66
1105	OE2	GLU	73	544.05
430	NZ	LYS	27	53.66
790	CG	ASP	54	77.21
792	OD2	ASP	54	134.40
1100	CG	GLU	73	51.08
1236	N	ARG	83	72.06
1238	CA	ARG	83	53.41
535	NE1	TRP	35	773.02
845	N	ASN	58	47.84
414	N	LYS	27	730.73
402	CA	THR	26	425.04
404	CB	THR	26	36.89
412	C	THR	26	47.40
813	O	ILE	55	40.74
905	N	LYS	62	73.86
853	OD1	ASN	58	239.53
883	N	GLU	60	86.77
894	OE1	GLU	60	36.19
895	OE2	GLU	60	32.10
834	N	SER	57	48.09
1274	N	SER	85	47.18
1281	OG	SER	85	31.43
1259	O	ARG	83	37.11
1260	N	ASN	84	635.58
1297	N	ARG	87	306.89
1320	O	ARG	87	54.84
816	CA	PHE	56	98.73
833	O	PHE	52	≥ 800
921	NZ	LYS	62	≥ 800
1268	OD1	ASN	84	46.02

**Tabla 3:** Átomos de la interfase de la proteína barstar que presentan altos tiempos de residencia, obtenidos de la simulación de DM del complejo, considerando  $\tau_2$ . de decaimiento.

Índice del átomo	Tipo de átomo	Residuo	N° de residuo	Tiempo (ps)
2335	CG	TRP	38	52.34
2272	CB	LEU	34	101.54
2270	CA	LEU	34	117.94
2287	N	ASP	35	98.56
2289	CA	ASP	35	68.19
2294	CG	ASP	35	186.37
2298	O	ASP	35	673.36
2297	C	ASP	35	30.30
2291	CB	ASP	35	86.80
2356	CB	ASP	39	89.28
2359	CG	ASP	39	300.63
2361	OD2	ASP	39	728.92
2414	N	TRP	44	333.40
2436	C	TRP	44	39.83
2438	N	VAL	45	326.27
2407	N	GLY	43	32.77
2409	CA	GLY	43	41.10
2969	OE2	GLU	76	37.47
2308	O	ALA	36	95.97
2336	CD1	TRP	38	41.09
2453	O	VAL	45	227.17
2469	N	TYR	47	33.85
2482	OH	TYR	47	39.73
2465	OE1	GLU	46	32.02
2467	C	GLU	46	42.18
2308	O	ALA	36	95.97
2336	CD1	TRP	38	41.09
2268	N	LEU	34	≥ 800
2295	OD1	ASP	35	≥ 800
2296	OD2	ASP	35	≥ 800
2468	O	GLU	46	≥ 800

**Tabla 4:** Moléculas de agua que median el acoplamiento entre barnasa y barstar, reportadas por Buckle<sup>53</sup>. En negrilla aparecen los átomos de barstar y en claro los átomos de barnasa. En color gris se resaltan los átomos que coinciden con los resultados de la simulación de DM del complejo.

Solvente	residuos	D....A (Å)	solvente ASA(Å <sup>2</sup> )
agua 14	<b>O Asp35</b>	2.9	0.0
	OE2 Glu 73	3.2	
agua 22	<b>OD2 Asp35</b>	2.8	0.0
	N Lys62	3.2	
	OD1 Asn58	2.8	
agua 29	<b>OD1 Asp35</b>	2.8	0.0
	NE Arg59	2.9	
agua 33	<b>OD1 Asp39</b>	2.6	0.0
	NZ Lys27	2.5	
	OE2 Glu 73	3.0	
agua 36	<b>O Val 45</b>	2.6	0.0
	OG Ser38	2.9	
	N Ser38	2.8	
agua 48	<b>NE1 Trp38</b>	3.2	3.0
	O Ile55	2.7	
	OE2 Glu 73	2.7	
agua 93	<b>N Tyr 47</b>	3.0	17.0
	OG Ser38	2.8	
agua 28	<b>OD2 Asp35</b>	2.7	0.0
	NZ Lys62	3.1	
	O Tyr103	2.7	
agua 155	<b>O Gly43</b>	3.0	1.0
	N Arg83	3.2	

Los sitios de hidratación de la interfase del complejo también se compararon con la interfase de las proteínas libres de los PDB con código de acceso 1A2P<sup>60</sup> para barnasa y 2HXX<sup>61</sup> para barstar. Ambas proteínas se resolvieron por cristalografía de rayos X, la primera con una resolución de 1.5 Å y la segunda con una resolución de 2.0 Å.

Del PDB 1A2P correspondiente a la proteína barnasa se identificaron los átomos de su interfase que están a menos de 3.5 Å de distancia con respecto a las moléculas de agua cristalográficas y se obtuvieron 43 sitios de hidratación, de los cuales 6 (Glu73:OE2, Lys27:NZ, Ile55:O, Asn58:OD1 Lys62:N y Lys62:NZ) coinciden con los reportados por Buckle y colaboradores. De estos 6 sitios de hidratación cristalográficos de la interfase de barnasa, todos concuerdan con los sitios de hidratación identificados a través de las simulaciones de DM del complejo. El mismo procedimiento se hizo con la proteína barstar de código de acceso 2HXX y se encontraron 3 sitios de hidratación (Asp39:OD1, Trp38:NE1 y Val45:O) cristalográficos, de los cuales el último concuerda con los reportados por las simulaciones de DM del complejo.

Al comparar los sitios de hidratación de la interfase del complejo con los de las proteínas libres se observa que hay mayor cantidad de estos sitios en el complejo que en las macromoléculas libres, lo que indica que en la interacción barnasa-barstar participan muchas moléculas de agua. Esta proliferación de moléculas de agua en la interfase del complejo se ve favorecida porque la interfase proteína-proteína es principalmente de naturaleza electrostática<sup>48</sup>, donde el sitio activo de barnasa presenta un potencial electrostático positivo y el de barstar un potencial negativo. La región con mayor carga negativa de la macromolécula barstar<sup>62</sup> está constituida por los residuos Asp35, Asp39, Glu76 y Glu80, de éstos los 3 primeros tienen átomos con tiempos de residencia superiores a los 30 ps.



De los resultados de la simulación de DM del complejo se observa que barnasa conserva varios sitios de hidratación con respecto a la proteína libre, mientras que barstar sólo conserva uno, con esto se puede inferir que barstar tiende a enlazarse al sitio activo de barnasa<sup>63</sup>. La interacción de un inhibidor al sitio activo de una enzima se ha visto en muchos inhibidores de proteasas, así como para el inhibidor de la ribonucleasa placentaria<sup>64</sup>.

En la interfase entre barnasa-barstar se identifican dos zonas en las cuales participa la hélice  $\alpha 2$  de barstar<sup>53</sup>. Una de estas zonas comprende el residuo Asp39 de barstar, el cual interactúa con los residuos Arg83, Arg87 e His102 del sitio activo de barnasa. La interacción entre estos aminoácidos es de tal magnitud, que las dos superficies, por este lado, presentan buena complementariedad; con muy pocos espacios en la interfase. La otra zona implica el lado de la hélice  $\alpha 2$  de barstar que colinda con el asa de barnasa, el cual participa en el reconocimiento de la guanina. La complementariedad en esta zona es bastante pobre, de hecho los espacios existentes entre las dos superficies están completamente llenos de agua (ver figura 4.21). De acuerdo con nuestros resultados de la función de supervivencia, encontramos que los tres residuos de barnasa de la zona de alta complementariedad, presentan átomos con tiempos de residencia superiores a los 40 ps, además de que éstos están poco expuestos, pues sus valores de  $P_{\alpha}(t)$  a  $t = 0$  no superan a 3 moléculas de agua coordinando a un átomo. Esto es una prueba de que en la zona de alta complementariedad hay moléculas de agua de alto tiempo de residencia y probablemente se encuentren enterradas en alguna cavidad. En cuanto a la zona de poca complementariedad, encontramos que los seis aminoácidos correspondientes al asa que interactúa con la guanina, presentan por lo menos un átomo o más átomos con alto tiempo de residencia, e igual que en la zona de poca complementariedad, la mayoría de ellos están poco expuestos, lo que también nos conduce a inferir que en esta zona hay moléculas de agua de alto tiempo de residencia.

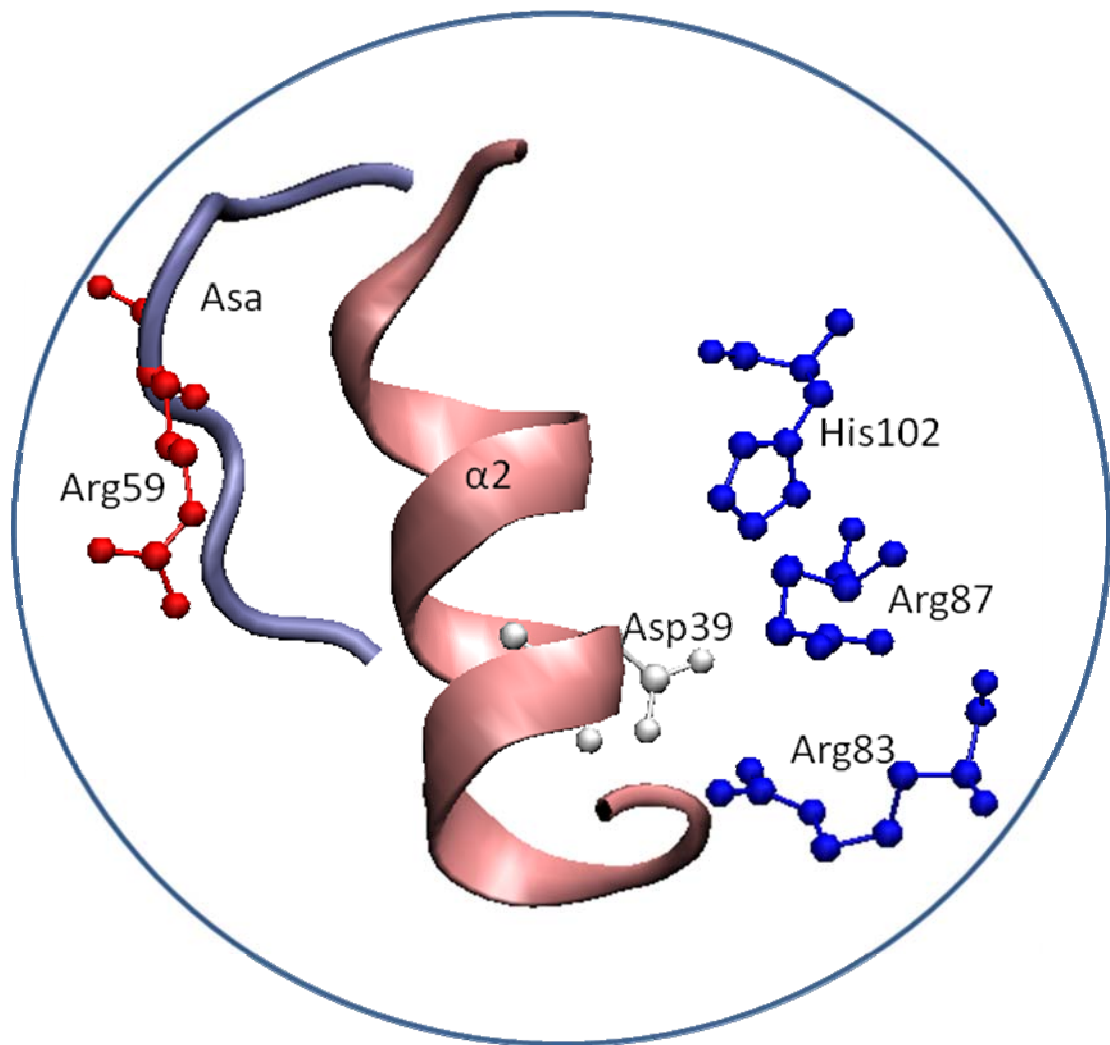


Figura 4.21. Interfase entre las proteínas barnasa-barstar, mediada por la hélice  $\alpha 2$  (color rosado) de barstar. Esta hélice divide a la interfase en dos zonas: una de alta complementaridad y otra de poca complementariedad. La primera la comprenden el Asp39 de barstar (color blanco) y los aminoácidos Arg83, Arg87 e His107 (color azul) de barnasa. La segunda zona corresponde al lado de la hélice  $\alpha 2$  que colinda con el asa (color azul claro) de barnasa que interactúa con la guanina.

#### **4.2. Resultados de la función de distribución radial de dos centros**

Hasta el momento, con la función de supervivencia hemos logrado identificar las moléculas de agua que permanecen unidas a la proteína durante largos tiempos. Una forma de complementar esta información es a través de la función de distribución radial de dos centros, la cual como se indicó en la sección de antecedentes, permite identificar moléculas de agua enlazadas por puentes de hidrógeno simultáneamente a dos átomos de una misma o de diferentes proteínas. En nuestro caso esta función puede ser una guía para identificar las moléculas de agua que median la interacción proteína-proteína incluso cuando la estructura en la zona de reconocimiento sufre de muchas fluctuaciones estructurales.

Buckle y colaboradores han reportado 14 casos en los que las moléculas de agua median la interacción entre barnasa-barstar ver tabla 4. Inicialmente para analizar la interfase del complejo, nosotros estudiamos los casos reportados por Buckle, con la función de distribución radial de dos centros y encontramos que 8 casos coinciden con los reportados por Buckle y colaboradores (ver figuras 2.4, 4.22, 4.23, 4.24 y 4.25). En la figura 2.4, en la figura 4.22 (gráfica A), en la figura 4.23 (gráficas E, G, H), y en la figura 4.24 (gráficas J, K y L) se observa que los átomos correspondientes a estas gráficas, se coordinan simultáneamente a una misma molécula de agua, pues la mayor densidad del agua, se ubica en estos casos, con respecto a cada uno de los átomos, a una distancia característica de enlace puente de hidrógeno (2.8-3.0 Å). En las otras figuras es evidente que las moléculas de agua, precisamente, no están coordinando por puentes de hidrógeno a los 2 átomos de cada gráfica (B, C, D, F, I y 4.25), pero sí se puede apreciar que en la mayoría de ellas hay una alta densidad de agua que se inclina por un átomo en particular de la pareja correspondiente de cada gráfica, esto es una muestra de que en la zona de la interfase barnasa-barstar hay un alto porcentaje de aguas estructurales.

En la gráfica A se observa que el átomo OD2 del residuo Asp35 de barstar y el átomo N de la Lys62 de barnasa coordinan simultáneamente a una molécula de agua, al analizar el átomo N de barnasa libre, también observamos que este átomo coordina simultáneamente a una molécula de agua con el átomo OD1 de Asn58 de barnasa libre, lo que nos permite suponer que esta molécula de agua contribuye en el reconocimiento de barnasa por barstar, puesto que se conserva en barnasa del complejo, otros ejemplos similares al anterior se advierten en las representaciones G y K, donde el oxígeno de la Ile55 y el oxígeno de la Tyr103 respectivamente, también coordinan simultáneamente por puente de hidrógeno a una molécula de agua tanto en barnasa del complejo como en barnasa libre.

En el esquema D no se aprecia una molécula de agua coordinada por dos átomos, pero sí se ve una alta densidad de agua cerca al átomo NE de la Arg59 de barnasa. Al evaluar este mismo átomo en barnasa libre, no se contempla una alta densidad de agua próxima a éste, lo que sugiere el ingreso de agua en la interfase del complejo para contribuir en la interacción barnasa-barstar. El mismo comportamiento se observa en el nitrógeno de la Ser38 de la gráfica 2.4, ya que en barnasa libre no se percibe una alta densidad de agua cerca de este átomo, pero en el complejo es evidente una molécula de agua coordinada al átomo de nitrógeno de la Ser38 de barnasa con el oxígeno de la Val45 de barstar. En la gráfica F la alta densidad de agua para los dos átomos, se localiza más allá de la distancia de un enlace puente de hidrógeno, mientras que en barnasa libre se percibe una alta densidad de agua en el átomo oxígeno gama de la Ser38 con otro átomo, lo que sugiere un desplazamiento de la molécula de agua en el momento de la formación del complejo.

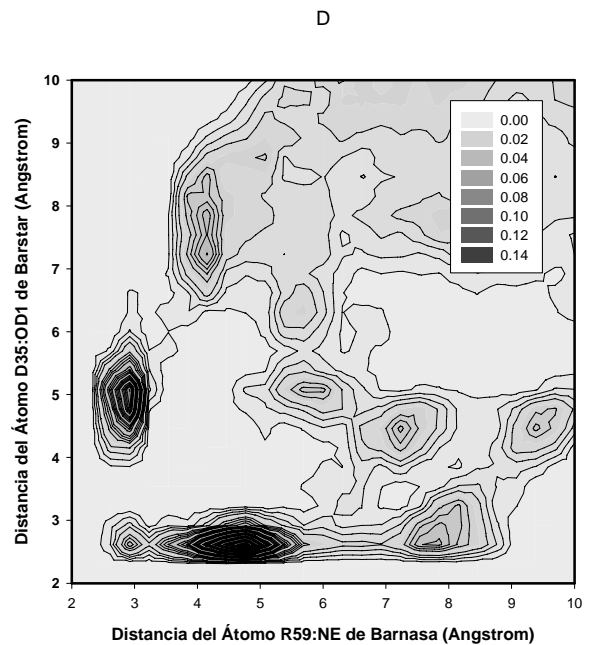
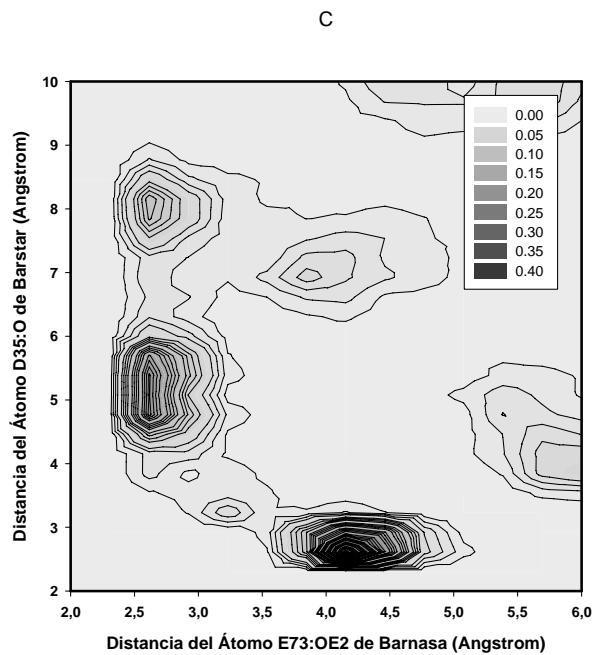
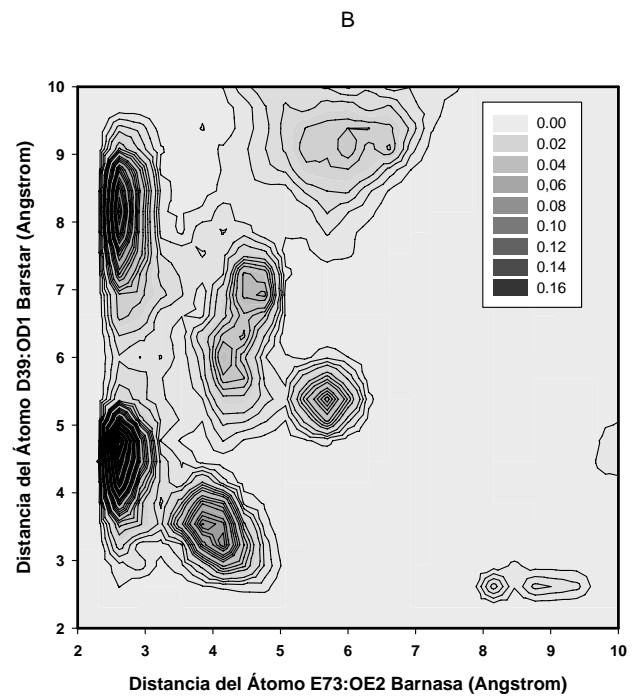
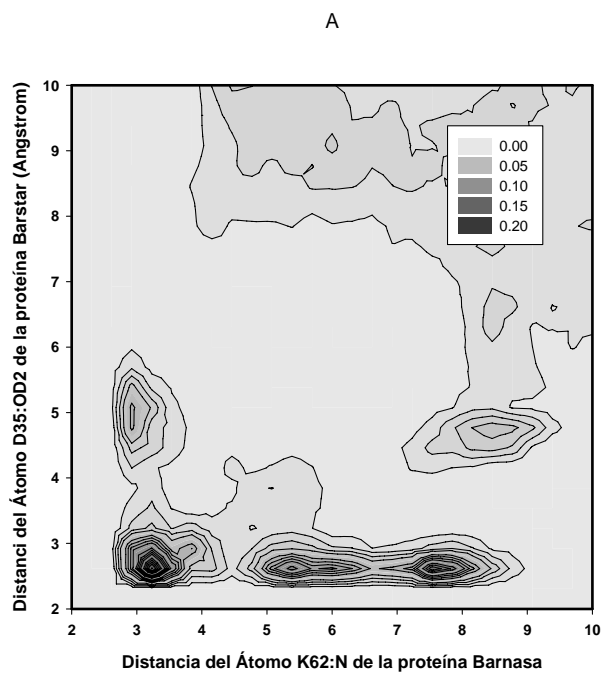


Figura 4.22. Evaluación de la función radial de dos centros para átomos de la interfase entre barnasa y barstar, con la finalidad de identificar moléculas de agua que medien dichas interacciones.

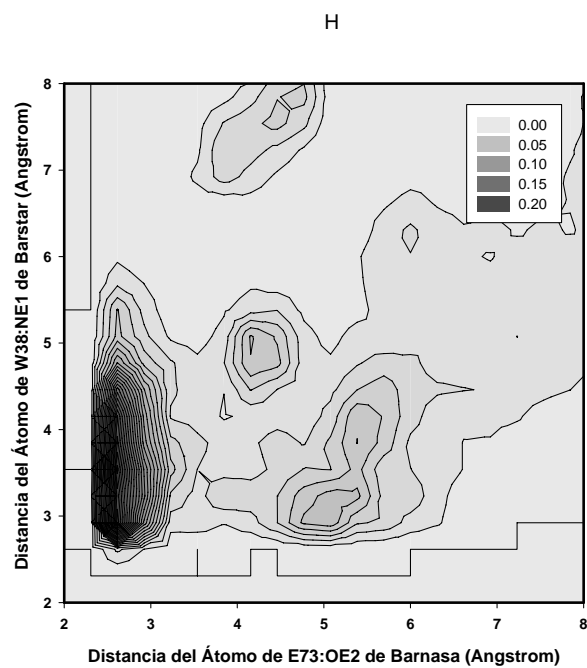
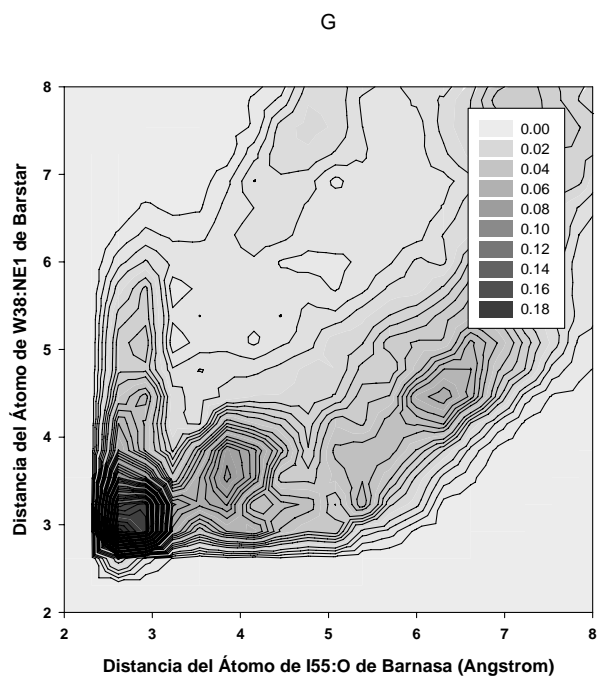
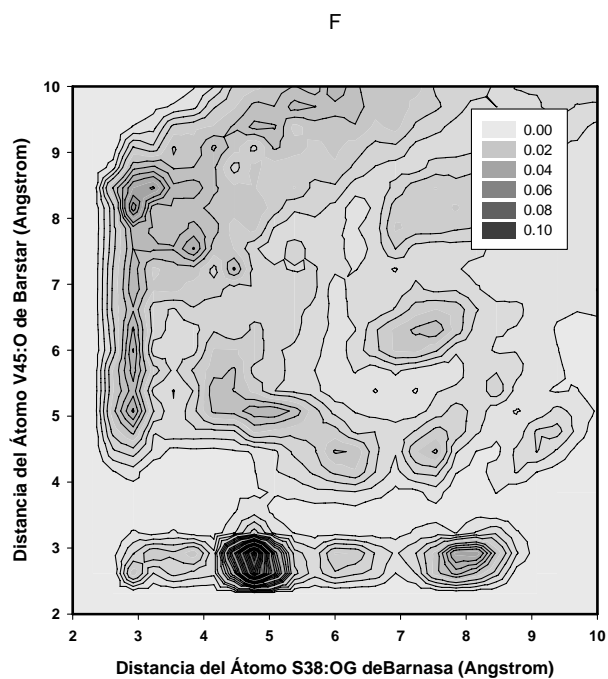
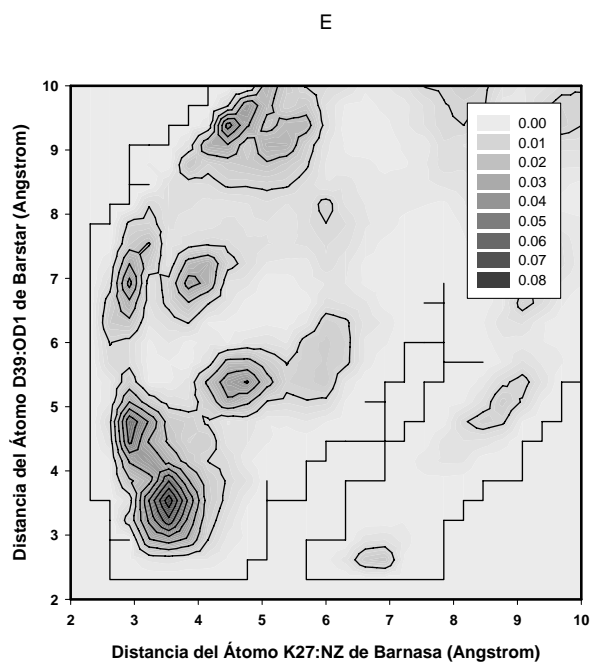


Figura 4.23. Evaluación de la función radial de dos centros para átomos de la interfase entre barnasa y barstar, con la finalidad de identificar moléculas de agua que medien dichas interacciones.

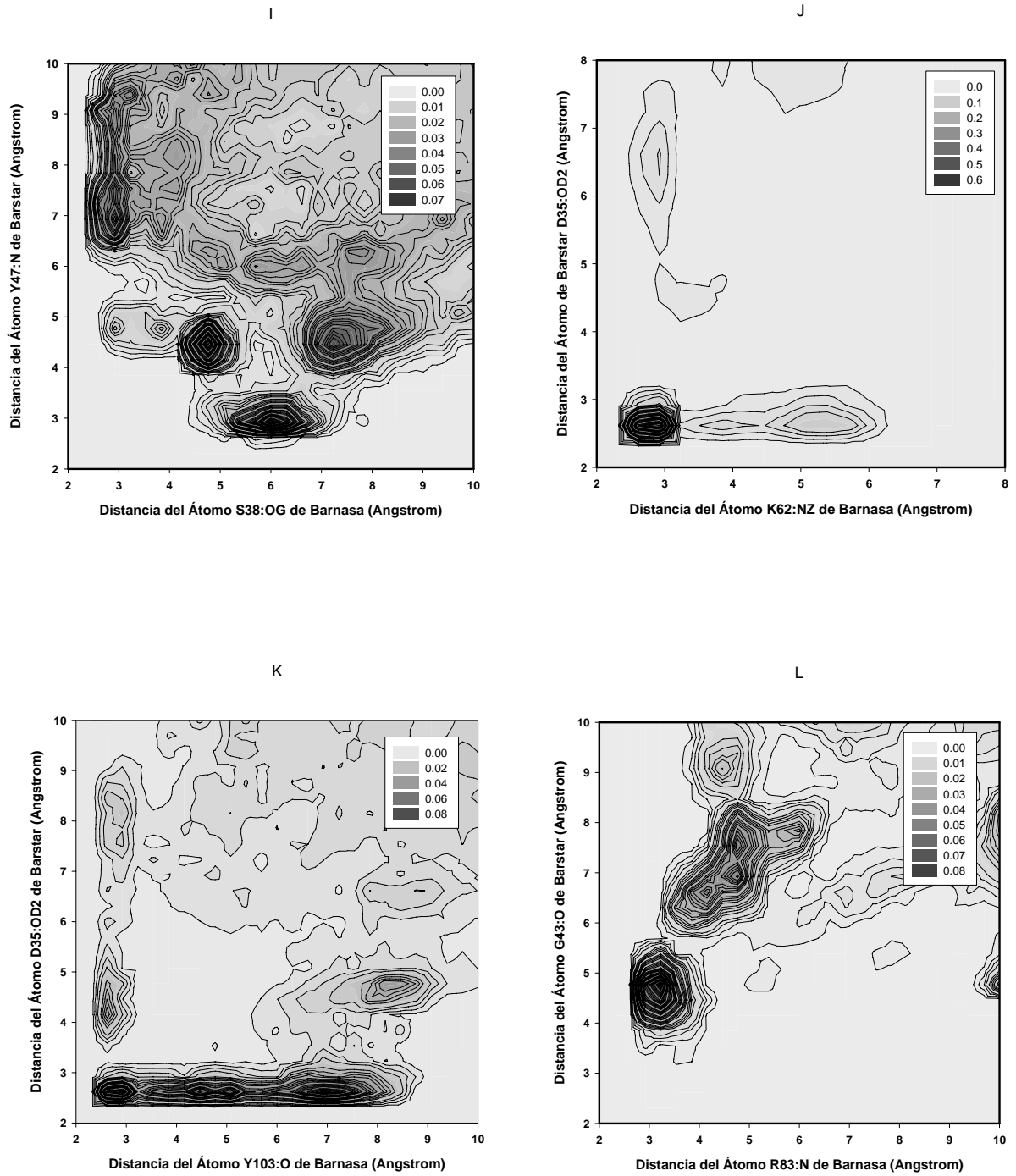


Figura 4.24. Evaluación de la función radial de dos centros para átomos de la interfase entre barnasa y barstar, con la finalidad de identificar moléculas de agua que medien dichas interacciones.

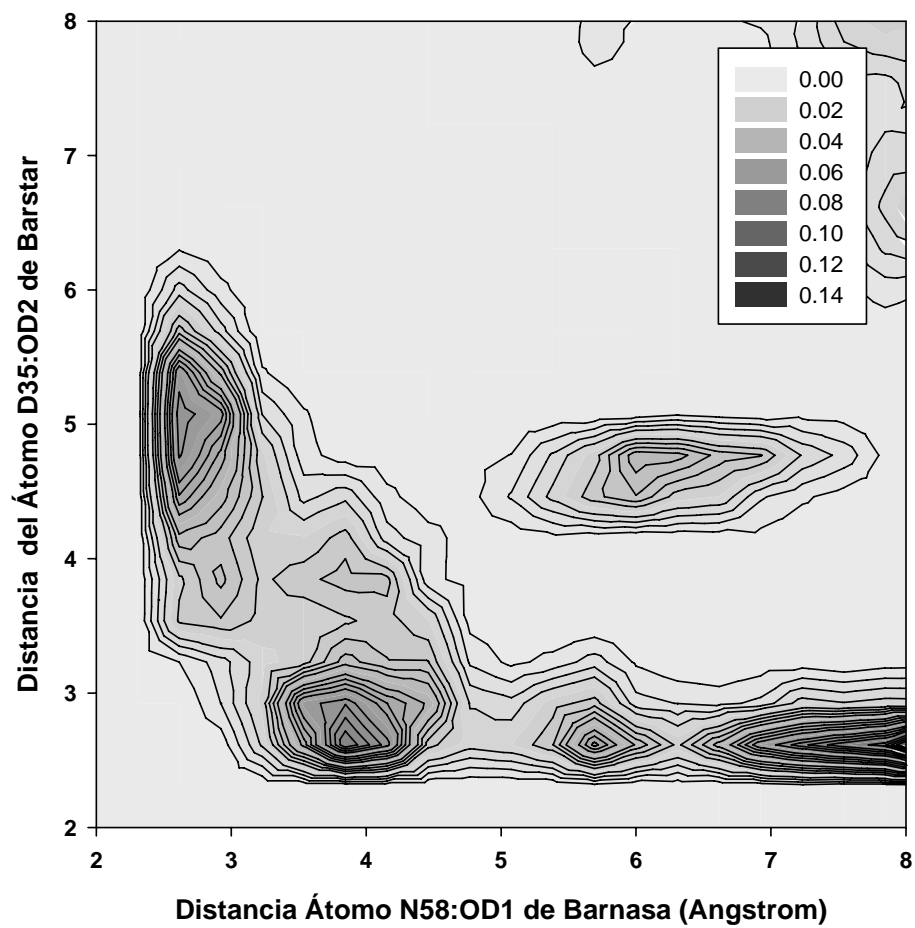


Figura 4.25. Evaluación de la función radial de dos centros para átomos de la interfase entre barnasa y barstar, con la finalidad de identificar moléculas de agua que medien dichas interacciones.



## 5. Conclusiones

Las simulaciones de dinámica molecular representan una buena alternativa para estudiar la esfera de solvatación de un átomo  $\alpha$  de la proteína, debido a que esta herramienta computacional proporciona información a nivel atómico de la interacción agua-proteína. Además ésta brinda la posibilidad de relacionar las diferentes configuraciones de los sistemas de estudio con el tiempo, de tal forma que para el análisis de los resultados de la dinámica molecular se puedan emplear funciones de correlación espacial y temporal, como fue el caso de la función de supervivencia, con la cual hicimos un estudio detallado de la distribución de las velocidades con que las moléculas de agua de la primera esfera de solvatación de un átomo  $\alpha$  de la proteína, se intercambian con las del seno de la solución.

Las diferentes simulaciones de dinámica molecular realizadas en esta tesis se efectuaron con el paquete NAMD, lo cual resultó ventajoso, porque el cálculo de 1 ns para proteínas pequeñas se efectuó aproximadamente en 12 horas y el de proteínas grandes en 24 horas, gracias a que éste es un programa optimizado para su ejecución en paralelo. El análisis de las simulaciones de DM se realizó también en poco tiempo, pues las proteínas pequeñas se analizaron en 12 horas y las grandes en 24 horas, esto gracias a la optimización del programa TRAPro y a su automatización mediante el script. En conjunto, el paquete NAMD, el programa de la función de supervivencia (TRAPro) y el script, permiten hacer un mapeo de las aguas superficiales de una proteína en poco tiempo.

El estudio de las propiedades dinámicas del agua en torno a las proteínas, con el programa de la función de supervivencia, nos permitió concluir que en una esfera de solvatación de un átomo  $\alpha$  de una proteína, las diferentes moléculas de agua se intercambian a diferentes tiempos con las moléculas de agua del seno de la solución; de hecho hay moléculas de agua de estas esferas que se comportan como si se

tratase de moléculas de agua del bulto (tiempos menores a 10 ps), y las otras, que precisamente no se comportan como moléculas de agua del seno de las solución, presentan tiempos de intercambio en el rango de 15 ps hasta 800 ps (ver figuras 4.2 y 4.3). Con esta información se puede confirmar que las moléculas de agua que se encuentran alrededor de una proteína, efectivamente no forman un cascarón semi-rígido entorno a ella.

La mayoría de las moléculas de agua de una esfera de solvatación (98%), salen de ésta en un tiempo inferior a los 10 ps, otras (pocas) antes de los 30 ps y muy pocas permanecen en la esfera de solvatación unidas por enlaces puentes de hidrógeno, más de 40 ps (ver figuras 4.4 y 4.5).

Una ventaja del uso del script en la ejecución del programa de la función de supervivencia, es la formación de los archivos PDB con los tiempos de decaimiento de la función de supervivencia, pues estos archivos al ser observados con el visualizador VMD, dan información inmediata del comportamiento de las moléculas de agua entorno a los átomos de la proteína, mediante una escala de colores (ver figuras 4.6, 4.7 y 4.8), lo que facilita la localización de las moléculas de agua de acuerdo a su tiempo de residencia en los átomos de una proteína; esto hace que la metodología aquí implementada para el estudio del comportamiento de una esfera de solvatación de un átomo alfa de una proteína, sea una técnica, sencilla, rápida, práctica y fácil de manipular.

La función de supervivencia también arroja información sobre el número de aguas coordinadas a un átomo  $\alpha$  de la proteína (ver figura 4.9). En la mayoría de los casos se puede observar que en el instante  $t = 0$  muchos de los átomos de las proteínas están coordinados a una molécula de agua solo la mitad del tiempo y muy pocos átomos a 6 ó 7 moléculas de agua (ver figuras 4.10 y 4.11). Esta cantidad de aguas coordinadas a los átomos de las proteínas es un referente de que algunas

partes de ésta son de alta exposición al solvente y otras partes de la macromolécula presentan un menor grado de exposición. Al confrontar la información del número de moléculas de agua coordinadas a un átomo de la proteína, con el tiempo de residencia de las moléculas de agua, se puede concluir que la mayoría de los átomos que presentan altos tiempos de residencia, corresponden a zonas de la proteína de poca exposición, posiblemente debido a que estas moléculas de agua se encuentran atrapadas en grietas o cavidades profundas de las proteínas.

De los 84 sitios de hidratación identificados en el consenso de sitios de hidratación, de la superposición de 8 estructuras cristalográficas de la proteína FKBP12, se encontró que 61 sitios concuerdan con los reportados mediante análisis de las simulaciones de dinámica molecular con la función de supervivencia. De estos sitios cristalográficos, los que presentan una mayor concordancia con nuestros estudios teóricos, son los correspondientes a las clases 8/8 y 7/8 (ver figura 4.12). Dentro de la clase 8/8 hay tres sitios hidratación que coinciden con los reportados por RMN, de éstos, todos coinciden con nuestros resultados teóricos. Con esto podemos concluir que las moléculas de agua identificadas por rayos X no siempre corresponden a moléculas de aguas estructurales.

Al evaluar los sitios de hidratación de la proteína tendamistat, con la función de supervivencia, correspondiente a las zonas de alta área de superficie expuesta y poca atenuación, identificados por RMN, mediante estudio con la sonda TEMPOL, se encontró que de los seis casos analizados (Tyr46, Pro50, Thr3, Thr54 y Arg49) ninguno está relacionado con moléculas de agua que se intercambian muy rápido con el seno de la solución, como si se tratase moléculas de agua del bulto. Los casos correspondientes a la Tyr46, Thr3 y Arg49, se caracterizaron por estar cerca de átomos que tienen tiempos de residencia mayores al promedio, mientras la Pro50 y la Thr54 están cerca a átomos con tiempos de residencia dentro del promedio.

Al confrontar los resultados de las simulaciones de dinámica molecular del complejo con los de las proteínas libres de barnasa y barstar, se encontró que 13 sitios de hidratación son comunes entre barnasa del complejo y barnasa libre, mientras que ninguno entre barstar del complejo y barstar libre, lo que sugiere que barstar se enlaza a barnasa y que las moléculas de agua de barnasa participan en este reconocimiento. De los 19 átomos del complejo reportados por Buckle, los cuales forman puentes de hidrógenos con moléculas de agua, 15 presentan altos tiempo de residencia superiores a los 30 ps. Estos 19 átomos de la tabla número 4, están organizados en 14 parejas de átomos de barnasa-barstar mediadas por moléculas de agua. De estas 14 parejas solo 8 coinciden con el análisis que se efectuó con la función de distribución radial de dos centros. De las moléculas de agua que median dichas interacciones, hay algunas que aparecen por primera vez en el complejo, otras que están en barnasa libre y se conservan en el complejo y otras que se reacomodan durante la formación del complejo.

Al considerar los mapas de colores del complejo barnasa-barstar se encontró que de los 23 aminoácidos de la interfase barnasa, solo 4 (Gln31, Val36, Phe38, Phe82) no forman parte de la zona de coloración azul, los demás tienen al menos un átomo cuyo tiempo de residencia es superior a los 30 ps. Mientras que de los 18 residuos de la interfase barstar sólo 6 residuos (Tyr29, Gly43, Glu32, Asn33, Thr42, Val73 y Glu80) tienen tiempos de residencia inferiores a los 30 ps y los demás hacen parte de la zona de coloración azul.

La metodología propuesta en esta tesis puede ser una herramienta complementaria a la determinación de la estructura cristalina de las proteínas, debido a que el conocimiento de las posiciones de las moléculas de agua estructurales, permite realizar refinamientos de la estructura cristalina de las proteínas. Como la RMN se encarga de identificar las diferentes posiciones de los átomos de la macromolécula y no da un informe global acerca de la estructura, sino que se vale de

simulaciones de DM para conseguir esta última, entonces saber la posición de estas moléculas en una posterior simulación de DM, puede contribuir, igual que en cristalografía de rayos X, a un refinamiento de la estructura de las proteínas.

## Apéndice A

En este apéndice se indican las funciones de los programas, de la función de supervivencia y la función radial de dos centros, que se utilizaron en el análisis de la dinámica molecular, al igual que el script que se empleó para ejecutar el programa de la función de supervivencia. También se muestran los programas de cada una de las funciones y el script.

### A.1 Funciones del programa de la función de supervivencia

Con la función de supervivencia se elaboró el programa TRAPro, el cual se:

- I. Usa para cualquier proteína, con la condición de que el número de átomos de agua, sea mayor que el número de átomos de la proteína; esta condición es modificable.
- II. Utiliza para leer las coordenadas de cada uno de los átomos que se encuentran en los archivos de configuraciones, tanto del agua como de las proteínas con todas sus variantes, arrojados por las simulaciones de DM.
- III. Emplea para hacer un arreglo tridimensional de las coordenadas del agua y de sus configuraciones (formación del vector agua).apartando memoria en la computadora (alocación dinámica).
- IV. Usa para levantar las coordenadas de toda la proteína en todas sus configuraciones en el vector agua, con la finalidad de minimizar el uso de memoria.
- V. Encarga de usar las condiciones periódicas para calcular la distancia entre las moléculas de agua y el átomo  $\alpha$  de la proteína
- VI. Emplea para organizar la matriz bidimensional que da información sobre la presencia o no de las moléculas de agua en la esfera de solvatación, este registro el programa lo hace así: en la matriz coloca valores de uno si la molécula de agua está presente en la esfera, o de cero si no está presente. En el programa esta matriz está denotada como la matriz "A"

- VII. Utiliza para calcular los tiempos de residencia de las moléculas de agua con  $t = m \times dt$ , donde  $m$  es el índice de la configuración y  $dt$  es el tiempo entre configuración y configuración.
- VIII. Usa para calcular los valores de  $P_\alpha$ , número de aguas coordinadas al átomo  $\alpha$  de la proteína, a partir de la matriz "A".
- IX. Emplea para construir un archivo para cada átomo de la proteína con los tiempos de residencia y los valores de  $P_\alpha$ .
- X. Utiliza para determinar los tiempos de decaimiento del 90% al 10% de la función de correlación temporal.
- XI. Ejecuta sobre átomos que se encuentran coordinados por lo menos a una o más moléculas de agua.

## **A.2 Funciones del programa de la función radial de dos centros**

Los ítems de I a V del programa de la función de supervivencia también se cumplen para la función radial de dos centros. A continuación se explican otras funciones de este programa. El programa de la función radial de dos centros se encarga de:

- I. Delimitar 65 esferas en torno a cada uno de los dos átomos de una misma proteína o de diferentes proteínas, previamente seleccionados.
- II. Contar el número de moléculas de agua que están presentes en la intersección de las esferas de solvatación de los átomos A y B de las proteínas en todas las configuraciones.
- III. Calcular los volúmenes de la intersección entre la esfera de solvatación del átomo A y del átomo B en todas sus configuraciones.
- IV. Determinar las densidades de la zona de intersección de los átomos A y B, empleando los datos calculados en los incisos II y III, en todas sus configuraciones.

- V. Definir la densidad promedio de cada una de las intersecciones de las esferas de los átomos A y B, considerando el número total de configuraciones.

### **A.3 Funciones del script (TRAPro)**

El programa de la función de supervivencia (TRAPro) se puede ejecutar sin emplear un script, claro está que sin éste, la ejecución del programa sobre cada uno de los átomos de la proteína puede demandar mucho tiempo, de ahí la necesidad de elaborar el script TRAPro. El script TRAPro se encarga de:

- XII. Formar un archivo inicial (input) para cada átomo de la proteína.
- XIII. Evaluar el programa TRAPro para cada átomo de la proteína en forma continúa.
- XIV. Asignar valores  $P_\alpha = 0$  y  $t = 0$  a los hidrógenos.
- XV. Organizar los 4 archivos generados por el programa TRAPro, para cada átomo de la proteína, en una carpeta
- XVI. Construir 9 archivos PDB, sustituyendo los valores de la columna del factor de temperatura por los tiempos característicos de decaimiento desde el 90% al 10%.
- XVII. Construir un archivo PDB, sustituyendo los valores de la columna del factor de temperatura por los valores de  $P_\alpha$  a  $t = 0$

### **A.4 Programa de la función de supervivencia**

A continuación se muestra el programa de la función de supervivencia, elaborado en lenguaje C para la plataforma de Linux, con el cual es posible estudiar las velocidades con que moléculas de agua de la primera esfera de solvatación de un átomo  $\alpha$  se intercambian con el seno de la solución.

```
// Programas de calculo de las funciones de correlacion
```



```

// para el analisis de la esfera de solvatacion de las proteinas
// Por:Eddy Francis Plaza Medina y el Dr. Jorge Peon Peralta
// Instituto de Quimica UNAM
// Mayo, 20007
// CALCULO DE FUNCION DE SUPERVIVENCIA (TIEMPOS DE RESIDENCIA)
// Algunos parametros de la medicion se indican con defines:

#include <stdio.h> //libreria estandar
#include <stdlib.h> //libreria para alocaacion dinamica
#include <math.h> //libreria de funciones seno y coseno
int main(void)
{
    //Declaracion de variables
    FILE *ofp, *off, *outf,*Pesout;
    float *wc;
    //float wc[nfr][3][nwats];
    float *atm;
    //float atm[nfr][3];
    short *A;
    short p;
    //short A[nfr][nwats]={0};
    float Palfa[4000]={0};
    float tiempo[4000]={0};
    float distx, disty, distz, psum, dist,distpre;
    float wcx,wcy,wcz, atmx, atmy, atmz;
    float boxsizeX, boxsizeY, boxsizeZ;
    int nats, nwats, nfr;
    float mindist,dt;
    int i,j,m,n,k,nmem,ppp,wcsize,index,atmsize,indice;
    char dummy[100], waterfile[100], protfile[100], nameout[100],
    poft[100];
    char Pesname[100];//Para el nombre del archivo que contiene pes y
tes
    float bb, mm;
    float Pes[11], Tes[11]={0.0};
    int checador[11]={0};
    scanf("%d \n",&indice);//lectura del indice de la proteina
    scanf("%s \n",dummy);
    scanf("%s \n",waterfile);//lectura del dcd del agua
    scanf("%s \n",dummy);
    scanf("%s \n",protfile);//lectura del dcd de la proteina
    scanf("%s \n",dummy);
    scanf("%d \n",&nfr);//lectura del numero de configuraciones en dcds
    scanf("%s \n",dummy);
    scanf("%d \n",&nats);//lectura del numero de atomos en dcd de
proteinas
    scanf("%s \n",dummy);
    scanf("%d \n",&nwats);//lectura del numero de atomos en dcd de agua
(oxigenos)
    scanf("%s \n",dummy);
    scanf("%f \n",&mindist);//lectura de la distancia que define el
radio de solvatacion del sitio
    scanf("%s \n",dummy);
    scanf("%f \n",&boxsizeX);//lectura de las dimensiones de la caja
    scanf("%s \n",dummy);
    scanf("%f \n",&boxsizeY);
    scanf("%s \n",dummy);
    scanf("%f \n",&boxsizeZ);
    scanf("%s \n",dummy);
    scanf("%f \n",&dt); //Lectura del delta del tiempo entre
configuraciones en picosegundos

```

```

    sprintf(nameout,"OUTPUT_%d",indice); //creacion de los archivos de
salida
    sprintf(poft,"surv_%d.sur",indice);
    sprintf(Pesname,"Pes_y_tes%d.tes", indice);
    outf=fopen(nameout,"wb"); //Despliegue del archivo OUTPUT para
escritura
    printf("-----
--\n");
    printf("PROGRAMA DE CALCULO DE LA FUNCION DE RESIDENCIA DE \n");
    printf("MOLECULAS DE AGUA EN LOS SITIOS DE HIDRATAACION DE UNA
PROTEINA \n");
    printf("-----
--\n");
    fprintf(outf,"-----
-----\n");
    fprintf(outf,"PROGRAMA DE CALCULO DE LA FUNCION DE RESIDENCIA DE
\n");
    fprintf(outf,"MOLECULAS DE AGUA EN LOS SITIOS DE HIDRATAACION DE UNA
PROTEINA \n");
    fprintf(outf,"-----
-----\n");
    fprintf(outf,"Indice (base 1) del sitio de hidratacion %d
\n",indice);
    fprintf(outf,"Nombre del dcd de las coordenadas del solvente= %s
\n",waterfile);
    fprintf(outf,"Nombre del dcd de las coordenadas de la proteina= %s
\n",protfile);
    fprintf(outf,"Nombre del archivo de salida con P(t) para este sitio=
%s \n",poft);
    fprintf(outf,"Este archivo de salida %s \n",nameout);
    fprintf(outf,"Delta tiempo entre configuraciones %f \n", dt);
    fprintf(outf,"Radio de la esfera de solvatacion considerada %f
\n",mindist);
    fprintf(outf,"Lado de X de la caja %f \n",boxsizeX);
    fprintf(outf,"Lado de Y de la caja %f \n",boxsizeY);
    fprintf(outf,"Lado de Z de la caja %f \n",boxsizeZ);
    fprintf(outf,"Numero de atomos en dcd de la proteina %d \n",nats);
    fprintf(outf,"Numero de atomos en dcd del agua %d \n",nwats);
    fprintf(outf,"Numero de frames en dcd %d \n",nfr);
    //Primero leer las coordenadas del sitio de la proteina,
    //se ocupa transitoriamente la matriz tridim de las aguas.
    wcsiz=nfr*3*nwats; //alocacion dinamica (se aparta memoria en el
computador
    wc=calloc(wcsiz,sizeof(float)); //para la matriz tridimensional del
agua)
    readccd (wc,protfile,nats,outf,nfr,nats);//Funcion auxiliar para la
lectura del dcd de la proteina

    atmsiz=nfr*3;
    atm=calloc(atmsiz,sizeof(float));
    //Copiar las coords del sitio a su respectivo array
    fprintf(outf,"Concluida la lectura de coords del sitio de
hidratacion \n");
    for (i=0;i<nfr;i++)
    {
        //Seleccion del atomo de la proteina
        //imax=nfr      jmax=3      kmax=nwats
        //wc[i][0][indice-1] //Guia para la lectura de la matriz
tridimensional de la proteina
        //atm[i][0] //Guia para formar la matriz bidimensional de la
proteina

```

```

        index=(3*nats*i+nats*0+(indice-1)); //ubicacion en el vector de
la proteina
        (atm[3*i+0])=*(&wc[index]); //valor identificado en el vector de
la proteina
        //wc[i][1][indice-1]
        //atm[i][1]
        index=(3*nats*i+nats*1+(indice-1));
        (atm[3*i+1])=*(&wc[index]);
        //wc[i][2][indice-1]
        //atm[i][2]
        index=(3*nats*i+nats*2+(indice-1));
        (atm[3*i+2])=*(&wc[index]);
    }
    //A continuacion, leer las coordenadas de las aguas
    readdcd (wc,waterfile,nwats,outf,nfr,nats); //funcion auxiliar para
la lectura del dcd del agua
    printf(      "Concluida la lectura de coords de las aguas \n");
    printf(      "Calculando la matriz A \n");
    //Calculo de la matriz A:
    //Identificacion de las aguas entorno al atomo de la proteina
    A=calloc(nwats*nfr,sizeof(short)); //alocacion dinamica para la
matriz A
    for (j=0;j<nwats;j++)
    {
        for (i=0;i<nfr;i++)
        {
            index=(3*nwats*i+nwats*0+j); //wc[i][0][j] //ubicacion en el
vector del agua
            wcx=*(&wc[0]+index); //valor identificado en el vector del
agua
            index=(3*nwats*i+nwats*1+j); //wc[i][1][j]
            wcy=*(&wc[0]+index);
            index=(3*nwats*i+nwats*2+j); //wc[i][2][j]
            wcz=*(&wc[0]+index);
            atm[3*i+0]=atm[3*i+0]; //coordenadas del atomo de la proteina
            atm[3*i+1]=atm[3*i+1];
            atm[3*i+2]=atm[3*i+2];
            distx=(wcx-atm[3*i+0]);
            distx=distx-boxsizeX*rintf(distx/boxsizeX);
            disty=(wcy-atm[3*i+1]);
            disty=disty*-boxsizeY*rintf(disty/boxsizeY);
            distz=(wcz-atm[3*i+2]);
            distz=distz*-boxsizeZ*rintf(distz/boxsizeZ);
            distpre=distx*distx+disty*disty+distz*distz;
            dist=sqrtf(distpre); //distancia del agua al atomo de la
proteina
            if(dist<mindist)
            {
                A[nwats*i+j]=1;
                // else A[i][j]=0; //Implicito
            }
        }
    }
    free(wc);
    free(atm);
    //Calculo de la funcion de supervivencia:
    fprintf (outf,"Terminado el calculo de la matriz A \n");
    fprintf (outf,"Realizandose el calculo de la P(t) \n");
    for (j=0;j<nwats;j++)//Loop sobre las aguas...total de nwats
terminos
    {

```

```

        if ((fmod(j,100)==0)||j==(nwats-1))
        printf ("Calculando el agua # %d \n",j);
        for (m=0;m<nfr;m++)//Loop sobre tiempos de resid t=m*dt ...c
total de nfr casos
    {
        psum=0.0;
        for (n=1;n<=(nfr-m);n++)//Loop sobre posibles tiempos zero
            {
                p=1.0;
                for (k=n;k<=(n+m);k++)//Loop sobre el indice de tiempos
(n a n+m)
                    {
                        p=p * A[(k-1)*nwats+j]; //Nota que el array es zero based
                        if (p==0.0)
                            break;
                    }//fin del loop k=n...
                    //hasta aqui, p=1 si agua @ sitio desde conf n hasta n+m
y cero si se salio
                psum=psum+p;
            }//fin del loop n=1...
        psum=psum/(nfr-m);
        Palfa[m]=Palfa[m]+psum;
        tiempo[m]=m*dt;
    }//fin del loop m=0...
} //fin del loop j=0...
//Calculos terminados, A salvar el asunto..
ofp=fopen(poft,"wb"); //despliegue del archivo surv para escritura
for (m=0;m<=nfr-1;m++)
    {
        fprintf (ofp,"%f %f \n", tiempo[m], Palfa[m]);
    }
fclose(ofp);
Pesout=fopen(Pesname,"wb"); //despliegue del archivo Pes y Tes para
escritura
for (i=1;i<=10;i++)
    {
        Pes[i]=Palfa[0]*(float)i/(float)10;
    }
fprintf (Pesout,"%d Tini 0 %f \n",indice, Palfa[0]);
for (m=1;m<=nfr-1;m++)
    {
        for (i=9;i>=1;i--)
            {
                if (Palfa[m]<Pes[i] && checador[i]==0)
                    {
                        checador[i]=1;
                        // indicemayor= m ;
                        //indicemenor= m-1;
                        //Seccion de interpolacion
                        mm=(Palfa[m]-Palfa[m-1])/(tiempo[m]-tiempo[m-1]); //calculo de
la pendiente
                        bb=Palfa[m]-mm*tiempo[m]; //calculo del punto de corte
                        Tes[i]=(Pes[i]-bb)/mm; //calculo del tiempo
                    }
            }
    }
    for (i=9;i>=1;i--)
        {
            fprintf (Pesout,"%d T%d %f %f \n",indice, i, Tes[i],
Pes[i]);
        }
    }

```

```

    fclose(Pesout);
    fprintf(outf,"Programa de analisis de dinamica del solvente
terminado OK /n");
    return 0;
}
//Funcion auxiliar
//int readdcd(float cw[nfr][3][nwats],char *filein, int numa)
int readdcd(float *cw,char *filein, int numa, FILE *outf, int nfr, int
nats)
{
    char s1, s2, s3, s4;
    int i,j, nset, istart, nsavc, namnf;
    int natoms, newsize, numlines;
    int a, k, lastbyte, numind=2, blocksize,index;
    double delta;
    FILE *ifp;
    if (numa==nats)
        fprintf(outf,"Lectura de coordenadas de sitio en la proteina \n");
    else
        fprintf(outf,"Lectura de coordenadas de oxigenos del agua \n");
    // Abrir archivo y leer varias partes del encabezado
    ifp=fopen(filein,"rb");
    a=fscanf(ifp, "%4c" , &i);
        fseek(ifp,0,2);
    lastbyte=ftell(ifp);
    fseek(ifp,4,0);
    fscanf(ifp,"%c",&s1);
    fscanf(ifp,"%c",&s2);
    fscanf(ifp,"%c",&s3);
    fscanf(ifp,"%c",&s4);
    fscanf(ifp,"%4c",&nset);
    fprintf(outf,"Numero de configuraciones indicadas= %d \n",nfr);
    fprintf(outf,"Numero de configuraciones por leerse (nset)= %d
\n",nset);
    if (nfr != nset)
        {
            fprintf(outf,"!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!! \n");
            fprintf(outf,"Numero de configuraciones por leerse del dcd no
coincide \n");
            fprintf(outf,"con el indicado en el programa...REVISAR \n");
            fprintf(outf,"!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!! \n");
            exit(1);
        }
    fscanf(ifp,"%4c",&istart);
    a=fscanf(ifp,"%4c",&nsavc);
    fseek(ifp,40,0);
    a=fscanf(ifp,"%4c",&namnf);
    //La proxima lectura de i indica si este, esta en formato charmm
    fseek(ifp,84,0);
    a=fscanf(ifp,"%4c",&i);
    fprintf(outf,"Indicador de Charmm (0-> NO) = %d \n",i);
    //siguiente, lectura del delta stuff
    fseek(ifp,44,0);
    a=fscanf(ifp,"%8c",&delta);
    //Nota, este es el delta leido del DCD, no necesariamente coincidira
con
    //El que se indica al programa a sabiendas de que es correcto
    fseek(ifp,92,0);
    a=fscanf(ifp,"%4c",&newsize);
    a=fscanf(ifp,"%4c",&numlines);
    fseek(ifp,numlines*80,1);

```

```

a=fscanf(ifp,"%4c",&newsize);
//Leer un cuatro, luego el numero de atomos , luego otro cuatro
a=fscanf(ifp,"%4c",&i);
a=fscanf(ifp,"%4c",&natoms);
fprintf(outf,"Numero de atomos en dcd= %d \n",natoms);
fprintf(outf,"Numero de atomos indicados al programa= %d \n",numa);
if (natoms != numa)
{
fprintf(outf,"!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!! \n");
fprintf(outf,"Numero de atomos por leerse del dcd no coincide
\n");
fprintf(outf,"con el indicado en el programa...REVISAR \n");
fprintf(outf,"!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!! \n");
exit(1);
}
a=fscanf(ifp,"%4c",&i);
//A Continuacion se leeran las coordenadas de los atomos
for (k=0;k<=nfr-1;k++)
{
a=fscanf(ifp,"%4c",&blocksize);
if ((blocksize/4)!=numa)
{
fprintf(outf,"Problema con blocksize");
exit(1);
}
for (j=0; j<=natoms-1;j++)
{
index=(3*natoms*k+natoms*0+j);
a=fscanf(ifp,"%4c",&cw[index]);
}
a=fscanf(ifp,"%4c",&blocksize);
if ((blocksize/4)!=numa)
{
fprintf(outf,"Problema con blocksize");
exit(1);
}
}
//Leyendo coordenadas y
a=fscanf(ifp,"%4c",&blocksize);
if ((blocksize/4)!=numa)
{
fprintf(outf,"Problema con blocksize");
exit(1);
}
for (j=0; j<=natoms-1;j++)
{
index=(3*natoms*k+natoms*1+j);
a=fscanf(ifp,"%4c",&cw[index]);
}
a=fscanf(ifp,"%4c",&blocksize);
if ((blocksize/4)!=numa)
{
fprintf(outf,"Problema con blocksize");
exit(1);
}
}
//Leyendo coordenadas z
a=fscanf(ifp,"%4c",&blocksize);
if ((blocksize/4)!=numa)
{
fprintf(outf,"Problema con blocksize");
exit(1);
}
}

```

```

    for (j=0; j<=natoms-1;j++)
    {
        index=(3*natoms*k+natoms*2+j);
        a=fscanf(ifp,"%4c",&cw[index]);
    }
    a=fscanf(ifp,"%4c",&blocksize);
    if ((blocksize/4)!=numa)
    {
        printf("Problema con blocksize");
        exit(1);
    }
}
fprintf(outf,"Saliendo del loop K sobre frames \n");
fclose(ifp);
fprintf(outf,"Archivo de lectura cerrado ok \n");
return 0;
}

```

## A.5 Programa de la función de distribución radial de dos centros

A continuación se muestra el programa de la función radial de dos centros, el cual permite identificar aguas coordinadas simultáneamente a dos átomos de una misma o de diferentes proteínas, elaborado en lenguaje C para la plataforma de Linux.

```

// Programas de calculo de las funciones de correlacion
// Aqui: Funcion de distribucion radial de DOS CENTROS
// para el analisis de la esfera de solvatacion de proteinas
// Por: Eddy Francis Plaza Medina y el Dr. Jorge Peon Peralta
// Instituto de Quimica UNAM
// Mayo, 2007
// CALCULO DE FUNCION DE DISTRIBUCION RADIAL DE DOS CENTROS

// Algunos parametros de la medicion se indican con defines:

#include <stdio.h> //Libreria estandar
#include <stdlib.h> //Libreria para alocaacion dinamica
#include <math.h> //Libreria de las funciones seno y coseno
#define nats 3159 //Numero de atomos en dcd de proteina
#define nwats 6788 //Numero de atomos en dcd de agua (oxigenos)
#define nfr 800 //Numero de configuraciones en dcDs
#define numbins 65 //Numero de bins para el calculo de g(r1,r2)
#define boxsizeX 57.8086 //Dimensiones de la caja
#define boxsizeY 52.0277
#define boxsizeZ 76.1145
#define radmax 20.0 //Maximo radio de calculo para g(r1,r2) en
angstroms
#define pi 3.1415926535897932384626433832795
//Densidad del agua en el seno de la solucion 0.0338
int main(void)
{
    //Declaracion de variables
    FILE *ofp, *off, *outf, *ofr;
    float *wc;
    //float wc[nfr][3][nwats];
    float *atmA;
    //float atmA[nfr][3];

```

```

float *atmB;
//float atmB[nfr][3];
short p;
int count[numbins][numbins][nfr]={0}; // contador del numero de ceros
en cada bin
float dr;
float distAx, distAy, distAz, distBx, distBy, distBz, psum,
distA,distpreA, distB,distpreB;
float wcx,wcy,wcz, atmAx, atmAy, atmAz, atmBx, atmBy, atmBz;
float distABx, distABy, distABz, distpreAB, R, rA, RA, rB, RB, V1,
V2, V3, V4;
float dvolpre,dvol=0.0;
float xii, xio, xoi, xoo;
int i,j,m,n,k,nmem,ppp,wcsize,index,indexA,indexB;
int atmAsize,atmBsize,indiceA,indiceB,l,binA,binB;
int rodol,rodolpre;
float helpA,help1,gr[numbins][numbins]={0}; //Matriz bidimensional de
contadores
char dummy[100], waterfile[100], protfile[100], nameout[100],
pofr[100], pofr[100];
float vl(float, float, float, float);
int acumula=0;

scanf("%s \n",dummy);
scanf("%d \n",&indiceA); //Lectura del indice del atomo A de la
proteina
scanf("%s \n",dummy);
scanf("%d \n",&indiceB); //Lectura del indice del atomo B de la
proteina
scanf("%s \n",dummy);
scanf("%s \n",waterfile); //Lectura del dcd de aguas
scanf("%s \n",dummy);
scanf("%s \n",protfile); //Lectura del dcd de la proteina
printf("%d \n", indiceA);
printf("%d \n", indiceB);
sprintf(nameout,"OUTPUT_%d_%d",indiceA, indiceB); //Creacion de los
archivos de salida
sprintf(poft,"gder_%d_%d.gr",indiceA, indiceB);
sprintf(pofr,"radiator_%d_%d.gr",indiceA,indiceB);
outf=fopen(nameout,"wb"); //Despliegue del archivo OUTPUT para
escritura
printf("-----
--\n");
printf("PROGRAMA DE CALCULO DE LA FUNCION DE DISTRIBUCION RADIAL
BIDIMENSIONAL DE \n");
printf("MOLECULAS DE AGUA EN LOS SITIOS DE HIDRATAION DE UN SOLUTO
\n");
printf("-----
--\n");
fprintf(outf,"-----
-----\n");
fprintf(outf,"PROGRAMA DE CALCULO DE LA FUNCION DE DISTRIBUCION
RADIAL BIDIMENSIONAL DE \n");
fprintf(outf,"MOLECULAS DE AGUA EN LOS SITIOS DE HIDRATAION DE UN
SOLUTO \n");
fprintf(outf,"-----
-----\n");
fprintf(outf,"Indice (base 1) del sitio de hidratacion de A %d
\n",indiceA);
fprintf(outf,"Indice (base 1) del sitio de hidratacion de B %d
\n",indiceB);

```



```

    fprintf(outf,"Nombre del dcd de las coordenadas del solvente= %s
\n",waterfile);
    fprintf(outf,"Nombre del dcd de las coordenadas de la proteina= %s
\n",protfile);
    fprintf(outf,"Nombre del archivo de salida con g(r) para este sitio=
%s \n",pofst);
    fprintf(outf,"Este archivo de salida %s \n",nameout);

    //Primero leer las coordenadas del sitio de la proteina,
    //se ocupa transitoriamente la matriz tridimensional de las aguas.
    wcsizewc=nfr*3*nwats*3 ; // Depende de la simulacion definido en el
define
    wc=calloc(wcsizewc,sizeof(float)); //Alocacion dinamica (se aparta
memoria en el computador)
    readdcd (wc,protfile,nats,outf); //Funcion auxiliar para la lectura
del dcd de la proteina
    atmASize=nfr*3;
    atmA=calloc(atmASize,sizeof(float));
    atmBSize=nfr*3;
    atmB=calloc(atmBSize,sizeof(float));
    //Copiar las coordenadas del sitio a su respectivo array
    fprintf(outf,"Concluida la lectura de coordenadas del sitio de
hidratacion \n");
    printf("Concluida la lectura de coordenadas del sitio de hidratacion
\n");
    for (i=0;i<nfr;i++)
    {
        //Seleccion del atomo de la proteina
        //imax=nfr      jmax=3      kmax=nwats
        //wc[i][0][indice-1] //Guia para la lectura de la matriz
tridimensional de la proteina
        //atmA[i][0] //Guia para formar la matriz bidimensional del
atomo de la proteina
        indexA=(3*nats*i+nats*0+(indiceA-1)); //Ubicacion en el vector
de la proteina
        (atmA[3*i+0])=(&wc[indexA]); //Valor identificado en el vector
de la proteina
        //wc[i][0][indice-1]
        //atmB[i][0]
        indexB=(3*nats*i+nats*0+(indiceB-1));
        (atmB[3*i+0])=(&wc[indexB]);
        //wc[i][1][indice-1]
        //atmA[i][1]
        indexA=(3*nats*i+nats*1+(indiceA-1));
        (atmA[3*i+1])=(&wc[indexA]);
        //wc[i][1][indice-1]
        //atmB[i][1]
        indexB=(3*nats*i+nats*1+(indiceB-1));
        (atmB[3*i+1])=(&wc[indexB]);
        //wc[i][2][indice-1]
        //atmA[i][2]
        indexA=(3*nats*i+nats*2+(indiceA-1));
        (atmA[3*i+2])=(&wc[indexA]);
        //wc[i][2][indice-1]
        //atmB[i][2]
        indexB=(3*nats*i+nats*2+(indiceB-1));
        (atmB[3*i+2])=(&wc[indexB]);
    }

    //A continuacion, leer las coordenadas de las aguas

```

```

    readdcd (wc,waterfile,nwats,outf); //Funcion auxiliar para la
lectura del dcd del agua
    fprintf(outf,"Concluida la lectura de coordenadas de las aguas \n");
    fprintf(outf,"Calculando la funcion de distribucion radial
bidimensional \n");
    //Calculos preparatorios para la g(r1,r2)
    dr=radmax/(float)numbins;
    //A continuacion se realizara el conteo de aguas
    for (i=0;i<nfr;i++)
    {
        for (j=0;j<nwats;j++)
        {
            index=(3*nwats*i+nwats*0+j); //wc[i][0][j] //Ubicacion en el
vector del agua
            wcx=*(&wc[0]+index); //Valor identificado en el vector
del agua
            index=(3*nwats*i+nwats*1+j); //wc[i][1][j]
            wcy=*(&wc[0]+index);
            index=(3*nwats*i+nwats*2+j); //wc[i][2][j]
            wcz=*(&wc[0]+index);
            atmAx=(atmA[3*i+0]); //coordenadas de los atomos A y B de la
proteina
            atmAy=(atmA[3*i+1]);
            atmAz=(atmA[3*i+2]);
            atmBx=(atmB[3*i+0]);
            atmBy=(atmB[3*i+1]);
            atmBz=(atmB[3*i+2]);
            distAx=(wcx-atmAx); //wc[i][0][j]
            distAx=distAx-boxsizeX*rintf(distAx/boxsizeX);
            distAy=(wcy-atmAy); //wc[i][1][j]
            distAy=distAy-boxsizeY*rintf(distAy/boxsizeY);
            distAz=(wcz-atmAz); //wc[i][2][j]
            distAz=distAz-boxsizeZ*rintf(distAz/boxsizeZ);
            distpreA=distAx*distAx+distAy*distAy+distAz*distAz;
            distA=sqrtf(distpreA); //distancia del agua al atomo A
            distBx=(wcx-atmBx); //wc[i][0][j]
            distBx=distBx-boxsizeX*rintf(distBx/boxsizeX);
            distBy=(wcy-atmBy); //wc[i][1][j]
            distBy=distBy-boxsizeY*rintf(distBy/boxsizeY);
            distBz=(wcz-atmBz); //wc[i][2][j]
            distBz=distBz-boxsizeZ*rintf(distBz/boxsizeZ);
            distpreB=distBx*distBx+distBy*distBy+distBz*distBz;
            distB=sqrtf(distpreB); //distancia del agua al atomo B

            //se calcula el bin al que corresponde esta mol de agua
            //en esta configuracion, y se acumula para el resp histograma
            if (distA<radmax && distB<radmax)
            {
                binA=floor(distA/dr);
                binB=floor(distB/dr);
                count[binA][binB][i]=count[binA][binB][i]+1; //Contador de
las agua en binA y binB
            }
        }
    }
}
    fprintf (outf,"Terminados los calculos sobre configuraciones \n");
    fprintf (outf,"Realizandose el calculo de la normalizacion de
g(r1,r2) \n");
    ofp=fopen(poft,"wb"); //se despliega el archivo gder para escritura
    ofr=fopen(pofr,"wb"); //se despliega el archivo radior para
escritura

```

```

//Calculos de la funcion de distribucion radial de dos centros
//En esta parte se realiza la normalizacion de los datos.
// Notese que para calcularlo, se necesita el valor de R
// Distancia entre los dos centros para poder hacer la
normalizacion
// Esto se debe a que el elemento de volumen (que depende de R)
// Depende de la configuracion del sistema
for (binA=0;binA<numbins;binA++)
{
for (binB=0;binB<numbins;binB++)
{
//Se resetea el acumulador que cuenta
//en cuantas configuraciones resultose en un dvol posit.
acumula=0;
for (i=0;i<nfr;i++)
{
atmAx=(atmA[3*i+0]); //coordenadas de los atomos A y B de
atmAy=(atmA[3*i+1]); //de la proteina
atmAz=(atmA[3*i+2]);
atmBx=(atmB[3*i+0]);
atmBy=(atmB[3*i+1]);
atmBz=(atmB[3*i+2]);
distABx=(atmBx-atmAx);
distABx=distABx-boxsizeX*rintf(distABx/boxsizeX);
distABY=(atmBy-atmAy);
distABY=distABY-boxsizeY*rintf(distABY/boxsizeY);
distABz=(atmBz-atmAz);
distABz=distABz-boxsizeZ*rintf(distABz/boxsizeZ);
distpreAB=distABx*distABx+distABY*distABY+distABz*distABz;
R=sqrtf(distpreAB); //distancia entre los dos Sitios
dvol=0.0;
rodol=0;
rodolpre=0;
rA=binA*dr; //radio interno de A
RA=rA+dr; //radio externo de A
rB=binB*dr; //radio interno de B
RB=rB+dr; //radio externo de B
xii=((rA*rA)+(R*R)-(rB*rB))/(2.0*R);
xio=((rA*rA)+(R*R)-(RB*RB))/(2.0*R);
xoi=((RA*RA)+(R*R)-(rB*rB))/(2.0*R);
xoo=((RA*RA)+(R*R)-(RB*RB))/(2.0*R);

//V(x1,x2,r,R)
//Calculo de los diferentes volmenes de interseccion
if (-RA<(R-rB) && (R-rB)<-rA) //caso II
{
V1=vl(-RA,xoi,RA,0.0);
V2=vl(R-rB,xoi,rB,R);
dvol=V1-V2;
rodol=2;
rodolpre=2;
}
if (-RA<=(R-RB) && (R-RB)<=-rA) //caso III
{
V1=vl(R-RB,xoo,RB,R);
V2=vl(xoo,xoi,RA,0.0);
V3=vl(-rA,xii,rA,0.0);
V4=vl(xii,xoi,rB,R);
dvol=V1+V2-V3-V4;
rodol=3;
// rodolpre=3;
}
}
}
}

```

```

    }

if (-rA<(R-RB) && (R-rB)<rA) //caso IV
{
V1=vl(xio,xoo,RB,R);
V2=vl(xoo,xoi,RA,0.0);
V3=vl(xio,xii,rA,0.0);
V4=vl(xii,xoi,rB,R);
dvol=V1+V2-V3-V4;
dvolpre=dvol;
rodol=4;
//    rodolpre=4;
}

if (rA<=(R-rB) && (R-rB)<=RA) //caso V
{
V1=vl(xio,xoo,RB,R);
V2=vl(xoo,xoi,RA,0.0);
V3=vl(xio,rA,rA,0.0);
V4=vl(R-rB,xoi,rB,R);
dvol=V1+V2-V3-V4;
rodol=5;
//    rodolpre=5;
}

if (RA<=(R-rB) && (R-RB)<=RA) //caso VI
{
V1=vl(R-RB,xoo,RB,R);
V2=vl(xoo,RA,RA,0.0);
dvol=V1+V2;
rodol=6;
//    rodolpre=6;
}

if ((RA >= R+RB) && ( rA<=R+RB ) ) //caso VII
{
V1=vl(xio,R+RB,RB,R);
V2=vl(xio,rA,rA,0);

dvol=V1-V2;
rodol=333;
//    rodolpre=333;
}

if (((R+rB) < RA) && ( (R+RB)>RA ) ) //caso VIII
//Ojo, este caso entra inicialmente en el caso IV
//pero aqui se re calcula correctamente el volumen
{
V1=vl(xoo,RA,RA,0.0);
V2=vl(xio,xoo,RB,R);
V3=vl(xio,xii,rA,0.0);
V4=vl(xii,rB+R,rB,R);
dvol=V1+V2-V3-V4;
rodol=1000;
//rodolpre=1000;
}

//    fprintf (ofp, "%f ",dvol); //para ver en el archivo gder
// si los volúmenes son (+)

if (dvol<0.0)
{
printf ("ABUSADO, SALIO UN VOL NEGATIVO %d\n",rodol);
printf ("Caso %d\n",rodol);
printf ("Caso pre-entro %d\n",rodolpre);
}

```

```

printf ("Valor de R  %f\n",R);
printf ("Valor de RA  %f\n",RA);
printf ("Valor de rA  %f\n",rA);
printf ("Valor de RB  %f\n",RB);
printf ("Valor de rB  %f\n",rB);
printf ("Valor de xoo  %f\n",xoo);
printf ("Valor de xio  %f\n",xio);
printf ("Valor de xii  %f\n",xii);
printf ("Valor de xoi  %f\n",xoi);
printf ("Valor de V1  %f\n",V1);
printf ("Valor de V2  %f\n",V2);
printf ("Valor de V3  %f\n",V3);
printf ("Valor de V4  %f\n",V4);
return 0;
}
//Corresponde al caso en que
//los segmentos de esfera no se
//traslapan (su intersec es cero)

if (dvol==0.0)
    gr[binA][binB]=gr[binA][binB]+0.0;
else

    //Ojo, solo cuando el dvol para una configuracion sale
posit,
    //se actualiza un acumulador, para asi solo dividir por
los
    //casos en los que en efecto se pudo calcular un volumen
    {
    acumula=acumula+1;
    //calculo de la densidad

gr[binA][binB]=gr[binA][binB]+(((float)count[binA][binB][i])/dvol;
    }
    if (i==(nfr-1))
    {
    if(acumula==0)
        help1=0;
    else
        //calculo de la densidad promedio
        help1=gr[binA][binB]/(float)nfr; //no esta dividido por la
densidad
    }
    } //Se cierra el for de nframes
    fprintf (ofp,"%f ",help1);
    } //Se cierra el for de binB
fprintf (ofp,"\n");
//Se da un retorno del renglon pa comenzar otro
// Notese que la columna determina el binB (for interno)
// y el renglon el binA (for externo)
helpA=((float)binA)*dr+(dr*0.5);
fprintf (ofr,"%f \n",helpA);
} //Se cierra el for de binA
free(wc);
free(atmA);
free(atmB);
//Calculos terminados, A salvar el asunto..
fclose(ofp);
fclose(ofr);
fprintf(outf,"Programa de analisis de la funcion de distribucion del
solvente terminado OK /n"); return 0;

```

```

}
float vl(float lim1, float lim2, float radio, float orig)
{
    float vol1, vo2, vo3, volu;
//V(lim1,lim2,radio,orig)
    vol1=(((lim1*lim1*lim1)-(lim2*lim2*lim2))*pi/3.0);
    vo2=((pi*orig)*((lim2*lim2)-(lim1*lim1)));
    vo3=(pi*(lim2-lim1)*((radio*radio)-(orig*orig)));
    volu=vol1+vo2+vo3;
    return volu;
}
//Funcion de lectura de los dcds
int readdcd(float *cw,char *filein, int numa, FILE *outf)
{
    char s1, s2, s3, s4;
    int i,j, nset, istart, nsavc, namnf;
    int natoms, newsize, numlines;
    int a, k, lastbyte, numind=2, blocksize,index;
    double delta;
    FILE *ifp;
    if (numa==nats)
        fprintf(outf,"Lectura de coordenadas de sitio en la proteina \n");
    else
        fprintf(outf,"Lectura de coordenadas de oxigenos del agua \n");
    // Abrir archivo y leer varias partes del encabezado
    ifp=fopen(filein,"rb");
    a=fscanf(ifp, "%4c" , &i);
    fseek(ifp,0,2);
    lastbyte=ftell(ifp);
    fseek(ifp,4,0);
    fscanf(ifp,"%c",&s1);
    fscanf(ifp,"%c",&s2);
    fscanf(ifp,"%c",&s3);
    fscanf(ifp,"%c",&s4);
    fscanf(ifp,"%4c",&nset);
    fprintf(outf,"Numero de configuraciones indicadas= %d \n",nfr);
    fprintf(outf,"Numero de configuraciones por leerse (nset)= %d
\n",nset);
    if (nfr != nset)
        {
            fprintf(outf,"!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!! \n");
            fprintf(outf,"Numero de configuraciones por leerse del dcd no
coincide \n");
            fprintf(outf,"con el indicado en el programa...REVISAR \n");
            fprintf(outf,"!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!! \n");
            exit(1);
        }
    fscanf(ifp,"%4c",&istart);
    a=fscanf(ifp,"%4c",&nsavc);
    fseek(ifp,40,0);
    a=fscanf(ifp,"%4c",&namnf);
    // namnf es el numero de atomos libres
    // La siguiente lectura de i indica si este esta en formato Charmm
    fseek(ifp,84,0);
    a=fscanf(ifp,"%4c",&i);
    fprintf(outf,"Indicador de Charmm (0-> NO) = %d \n",i);
    //Siguiente, lectura del delta stuff
    fseek(ifp,44,0);
    a=fscanf(ifp,"%8c",&delta);
    //Nota, este es el delta leído del DCD, no necesariamente coincidira
con

```

```

//El que se indica al programa a sabiendas de que es correcto
fseek(ifp,92,0);
a=fscanf(ifp,"%4c",&newsize);
a=fscanf(ifp,"%4c",&numlines);
fseek(ifp,numlines*80,1);
a=fscanf(ifp,"%4c",&newsize);
//Leer un cuatro, luego el numero de atomos , luego otro cuatro
a=fscanf(ifp,"%4c",&i);
a=fscanf(ifp,"%4c",&natoms);
fprintf(outf,"Numero de atomos en dcd= %d \n",natoms);
fprintf(outf,"Numero de atomos indicados al programa= %d \n",numa);
if (natoms != numa)
{
    fprintf(outf,"!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!! \n");
    fprintf(outf,"Numero de atomos por leerse del dcd no coincide
\n");
    fprintf(outf,"con el indicado en el programa...REVISAR \n");
    fprintf(outf,"!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!! \n");
    exit(1);
}
a=fscanf(ifp,"%4c",&i);
//A Continuacion se leeran las coordenadas de los atomos
for (k=0;k<=nfr-1;k++)
{
//Leyendo coordenadas x
a=fscanf(ifp,"%4c",&blocksize);
if ((blocksize/4)!=numa)
{
    fprintf(outf,"Problema con blocksize");
    exit(1);
}
for (j=0; j<=natoms-1;j++)

{
    index=(3*natoms*k+natoms*0+j);
    a=fscanf(ifp,"%4c",&cw[index]);
}
a=fscanf(ifp,"%4c",&blocksize);
if ((blocksize/4)!=numa)
{
    fprintf(outf,"Problema con blocksize");
    exit(1);
}
//Leyendo coordenadas y
a=fscanf(ifp,"%4c",&blocksize);
if ((blocksize/4)!=numa)
{
    fprintf(outf,"Problema con blocksize");
    exit(1);
}
for (j=0; j<=natoms-1;j++)
{
    index=(3*natoms*k+natoms*1+j);
    a=fscanf(ifp,"%4c",&cw[index]);
}
a=fscanf(ifp,"%4c",&blocksize);
if ((blocksize/4)!=numa)
{
    fprintf(outf,"Problema con blocksize");
    exit(1);
}
}

```

```

//Leyendo coordenadas z
a=fscanf(ifp,"%4c",&blocksize);
if ((blocksize/4)!=numa)
{
    fprintf(outf,"Problema con blocksize");
    exit(1);
}
for (j=0; j<=natoms-1;j++)
{
    index=(3*natoms*k+natoms*2+j);
    a=fscanf(ifp,"%4c",&cw[index]);
}
a=fscanf(ifp,"%4c",&blocksize);
if ((blocksize/4)!=numa)
{
    printf("Problema con blocksize");
    exit(1);
}
}
fprintf(outf,"Saliendo del loop K sobre frames \n");
fclose(ifp);
fprintf(outf,"Archivo de lectura cerrado ok \n");
return 0;
}

```

## A.6 Script que se emplea para la ejecución del programa de la función de supervivencia

Este script se elaboró con la finalidad de agilizar el cálculo de la función de supervivencia, éste se elaboró en lenguaje bash, para la plataforma de Linux.

```

#script para la ejecución del programa TRAPro#

#!/bin/bash

#introduce el nombre del archivo#

read Y

echo "nombre de archivo de salida"

#crea una carpeta con el nombre del archivo de salida#

mkdir /home/eddy/EjercEddyTend/$Y

A=home

B=eddy

C=EjercEddyTend

#construyendo el archivo input y ejecutando el programa TRAPro para
cada uno de los átomos de la proteína#

a=1

b=1076

```



```

w=_
m=0.0

grep ATOM /$A/jpeon/Tendamistat/run1/dcads/tend_md1_wb_run1C.pdb >
/$A/$B/$C/$Y/A1

cut -b 7-20 /$A/$B/$C/$Y/A1 >> /$A/$B/$C/$Y/A2

file=/$A/$B/$C/$Y/A2

{
while [ $a -le $b ]
do

    read a p z

    echo $a, $p, $z

        mkdir /$A/$B/$C/$Y/$a$p$w$z

    echo $a > /$A/$B/$C/$Y/$a$p$w$z/$p

    cat /$A/$B/$C/input_1.0ps >> /$A/$B/$C/$Y/$a$p$w$z/$p

    if [ /$A/$B/$C/$Y/$a$p$w$z/$p = /$A/$B/$C/$Y/$a$p$w$z/H* ]
    then
        cat /$A/$B/$C/Pes_y_tes.tes >> /$A/$B/$C/Pes_y_tes$a.tes
    else
        ./a.out < /$A/$B/$C/$Y/$a$p$w$z/$p
    fi

#Trasladando los archivos del programa TRAPro, a las carpetas de cada
átomo#

    mv /$A/$B/$C/surv_$a.sur /$A/$B/$C/$Y/$a$p$w$z

    mv /$A/$B/$C/OUTPUT_$a /$A/$B/$C/$Y/$a$p$w$z

    mv /$A/$B/$C/Pes_y_tes$a.tes /$A/$B/$C/$Y/$a$p$w$z

    y=1

    t=9

#Formando los archivos PDB#

    grep Tini /$A/$B/$C/$Y/$a$p$w$z/Pes_y_tes$a.tes
>>/$A/$B/$C/$Y/Tini

    while [ $y -le $t ]
    do

        grep T$y /$A/$B/$C/$Y/$a$p$w$z/Pes_y_tes$a.tes
>>/$A/$B/$C/$Y/T$y

        y=$((y+1))

```

```
done
a=$((a+1))
done
}</$A/$B/$C/$Y/A2
mkdir /$A/$B/$C/$Y/resulpdb
x=1
w=10
while [ $x -le $w ]
do
    cut -b 1-60 <
/$A/jpeon/Tendamistat/run1/dcads/tend_md1_wb_run1C.pdb >
/$A/$B/$C/$Y/sinfactem$x

    grep ATOM /$A/$B/$C/$Y/sinfactem$x > /$A/$B/$C/$Y/sinfa$x

    cut -d ' ' -f5,6 /$A/$B/$C/$Y/T$x | paste /$A/$B/$C/$Y/sinfa$x ->
/$A/$B/$C/$Y/T$x.pdb

    mv /$A/$B/$C/$Y/T$x.pdb /$A/$B/$C/$Y/resulpdb

    rm /$A/$B/$C/$Y/sinfactem$x

    rm /$A/$B/$C/$Y/sinfa$x

    x=$((x+1))

    if [ $x = $w ]

    then

        break

    fi

done
```

## Referencias bibliográficas

1. Petsko, G., Ringer, D., Protein Structure and Function (primers in Biology). New Science Press, Ltd. 2004.
2. Takano, K., Yamagata, Y., Yutani, K., *Protein Engineering* **2003**, 16: .5–9.
3. Mikol, V., Kallen, J., Pflügl, G., Walkinshaw, M. D., *J. Mol. Biol.* **1993**, 234: 1119-1130.
4. Huang, K., Lu, W., Anderson, S., Laskowski, M., R. J., James, M. N., *Protein Sci.* **1995**, 4: 1985-1997.
5. Kríz, Z., Otyepka, M., Bártová, I., Koca, J., *PROTEINS: Structure, Function, and Bioinformatics*, **2004**, 55: 258-274.
6. Tuena de Gómez-Puyou, M., and Gómez Puyou, A., *Critical Reviews in Biochemistry and Molecular Biology* **1998**, 33: 53-89.
7. Fernández-Velasco, D.A., Rodríguez, R. Vargas. S., Tuena de Gómez-Puyou, M., and Gómez-Puyou, A., *J. Colloid. Interface. Sci.* **1998**, 127: 29-35.
8. Derreumaux, P., Schlick, T., *Biophys. J.* **1998**, 74: 72-81.
9. Gregory, R., B., Protein-Solvent interactions, New York. Marcel Dekker, 1995.
10. Levitt, M., Park, B. H., *Structure* **1993**, 4: 223-226.
11. Otting, G., WüTrich, K., *J. Am. Chem. Soc.* **1989**, 111: 1871-1875.
12. Karplus, P., Faerman, C., *Curr Opin Struct Biol.* **1994**, 4: 770-776.
13. Pal, S. K.; Zewail, A. H., *Chem. ReV.* **2004**, 104: 2099-2123.
14. Otting, G., Liepinsh, E., WüTrich, K., *Science* **1991**, 254: 974-980.
15. Finer-Moore J. S., Kossiakoff, A. A., Hurley, J. H., Earnest, T., Stroud, R. M., *Proteins.* **1992**, 12: 203-222.
16. Faerman, C., Karplus, A., *Protein Struct. Funct. Genet.* **1995**, 23: 1-11.
17. Pal, S. K., Peón, J., Bagchi, B., Zewail, A. H., *J. Phys. Chem. B* **2002**, 106: 12376-12395.
18. Meyer E., *Protein Sci.* **1992**, 1: 1543-1562.

19. Denisov, V., Venu, K., Peters, J., Hörlein, H. D., Halle, Bertil Halle., *J. Phys. Chem. B* **1997**, 101: 9380-9389.
20. Williams, M. A., Goodfellow, J. M., Thornton, J. M., *Protein Sci.* **1994**, 3: 1224-1235.
21. Li, Z., Lazaridis, T., *Phys. Chem. Chem. Phys.* **2006**, 9: 573-581.
22. Halle, B., *Phil. Trans. R. Soc. Lond. B* **2004**, 359: 1207-1224.
23. Scarselli, M., Bernini, A., Segoni, C., Molinari, H., Esposito, G., Lesk, A., Laschi, F., Temussi, P., Niccolai, N., *J. of Biomolecular NMR* **1999**, 15: 125-133.
24. Becker, O., Mackerell, A. Roux, B., Watanabe, M. Computational Biochemistry and Biophysics. New York. Marcel Dekker, 2001.
25. Frenkel, D., Smit, B. Understanding Molecular Simulation From algorithms to Applications. San Diego, Academic Press, 2002.
26. Karplus, M., Petsko, G.A., *Nature* **1990**, 347: 631-639.
27. Jensen, F. Introduction to Computational Chemistry. Wiley, 1999, p. 15-23.
28. Cuevas, G., Cortés, F. Introducción a la Química Computacional. México. Fondo de Cultura Económica, 2003.
29. Encyclopedia of Computational Chemistry. Wiley, 1998; Vol. 1, p. 11-3, 2191-2199, Vol. 3, p. 271-277.
30. Allen, M. P., Tildesley, D. J., Computer Simulations of Liquids. Oxford: Clarendon Press, 1987.
31. Phillips, J., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R., Kalé L., Schulten, K., *J. Comput. Chem.* **2005**, 26: 1781-1802.
32. Impey, R. W., Madden, P. A., McDonal, I. R., *J. Phys. Chem.* **1983**, 87: 5071-5083.
33. García, A. E., Stiller, L., *J., Comput. Chem.* **1993**, 14: 1396-1406.
34. Luise, A., Falconi, M., Desideri, A., *PROTEINS: Structure, Function, and Genetics*, **2000**, 39: 56-67.
35. Bizzarri, A. R., Cannistraro, S., *J. Phys. Chem. B* **2002**, 106: 6617-6633.

36. Gu, W., Schoenborn, B., *PROTEINS: Structure, Function, and Genetics*, **1995**, 22: 20-26.
37. Lawrence, M., Colman, P., *J. Mol. Biol.* **1993**, 234: 946-950.
38. Xu, R.X., Meadows, R.P., Fesik, S.W., *Biochemistry* **1993**, 32: 2473-2480.
39. Van Duyne, G., Standaert, R., Karplus, P., Schreiber, S., Clardy, J., *J. Mol. Biol.* **1993**, 229: 105-124.
40. Holt, D., Luengo, J., Yamashita, D., Oh, H., Konialian, A., Yen, H., Rozamus, L., Brandt, M., Bossard, M., Levy, M., Eggleston, D., Liang, J., Schultz, W., Stout, T., Clardy, J., *J. Am. Chem. Soc.* **1993**, 115: 9925-9938.
41. Main, E., Fulton, K., Daggett, V., Jackson, S., *J. Biol. Phys.* **2001**, 27: 99-117.
42. DeCenzo, M., Park, S., Jarrett, B., Aldape, R., Futer, O., Murcko, M., Livingston, D., *Protein Eng.* **1996**, 9: 173-180.
43. Wiegand, G., Epp, O., Huber, R., *J. Mol. Biol.*, **1995**, 247: 99-110.
44. Köenig, V., Vértesy, L., Schneider, T., *Acta Cryst. Sect. D* **2003**, 59: 1737-1743.
45. Ikura, T., Urakubo, Y., Ito, N., *Chem. Phys.* **2004**, 307: 111-119.
46. Baudet, S., Janin, J., *J. Mol. Biol.* **1991**, 219: 123-132.
47. Wong, K. B., Daggett V., *Biochemistry* **1998**, 37: 11182-11192.
48. <sup>1</sup> Vaughan, C. K., Buckle, A. M., Fersht, A. R., *J. Mol. Biol.* **1999**, 286: 1487-1506.
49. Bhandarkar, M., Brunner, R., Chipot, C., Dalke, A., Dixit, S., Grayson, P., Gullingsrud, J., Gursoy, A., Hardy, D., Humphrey, W., Hurwitz, D., Krawetz, N., Nelson, M., Phillips, J., Shinozaki, A., Zheng, G., Zhu, F. NAMD User's Guide. Theoretical Biophysics Group, University of Illinois and Beckman Institute. 2003.
50. Phillips, J., Isgro, T., Sotomayor, m., Villa, E. NAMD TUTORIAL. University of Illinois. 2003.
51. Caddigan, E., Cohen, J., Gullingsrud, J., Stone, J. VMD User's Guide. Theoretical Biophysics Group, University of Illinois and Beckman Institute. 2003.
52. Michnick, S.W., Rosen, M.K., Wandless, T.J., Karplus, M., Schreiber, S.L. *Science* **1991**, 252: 836-839.

53. Buckle, A., Schreiber, G., Fersht, A., *Biochemistry* **1994**, 33: 8878-8889.
54. Ramachandran, G. N., Ramakrishnan, C., Sasisekharan, V., *J. Mol. Biol.* **1963**, 7: 95-99.
55. Havel, T. F., *Prog. Biophys. Mol. Biol.* **1991**, 56: 43-78.
56. Kelley, A., Pohl I., C By Dissection the Essentials of C Programming. California. Addison-Wesley Publishing Company, Inc. 1996.
57. Evans, A., Matthew, N., Stone, R. Instant Unix. Canada. Wrox Press Ltd. 1995.
58. Fiamingo, F., Debula, L., Condron, L. Introduction to Unix. University Technology Services, The Ohio State University. 1998.
59. Hartley, R. W., *Biochemistry* **1997**, 32: 5978-5984.
60. Mauguen, Y., Hartley, R. W., Dodson, E. J., Dodson, G. G., Bricogne, G., Chothia, C., Jack, A., *Nature* **1982**, 297: 162-164.
61. Rubini, M., Lepthien, S., Golbik, R., Budisa, N. B., *Biochim. Biophys. Acta* **2006**, 1764: 1147-1158.
62. Lubienski, M., Bycroft, M., Freund, S., Fersht, A., *Biochemistry* **1994**, 33: 8866-8877.
63. Schreiber, G., Fersht, A., *Biochemistry* **1993**, 32: 5145-5150.
64. Janin, J., Cothia, C., *J. Biol. Chem.* **1990**, 265: 16207-16030.