



UNIVERSIDAD NACIONAL AUTÓNOMA DE  
MÉXICO

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

**LA TRIOSAFOSFATO ISOMERASA DE *TRIPANOSOMA CRUZI* EN MEZCLAS AGUA-DECANO  
ESTUDIADA POR MEDIO DE DINÁMICA MOLECULAR**

TESIS  
PARA OPTAR POR EL GRADO DE  
**MAESTRO EN CIENCIAS**

PRESENTA  
NORMA DELIA DÍAZ VERGARA



TUTOR: DR. ÁNGEL PIÑEIRO GUILLÉN 2006



Universidad Nacional  
Autónoma de México



**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.

## DEDICATORIA

*A la memoria de mi padre,*

*a mi madre y hermanas.*

*A mi familia y amigos.*

## AGRADECIMIENTOS

*Al Doctor Ángel Piñeiro Guillén por su apoyo, confianza y amistad.*

*A los Doctores Miguel Costas Basín y Silvia Pérez Casas.*

*A los Doctores Carlos Villarreal Luján y Leonor Huerta Hernández.*

*Por su contribución al manuscrito agradezco a los Doctores miembros del jurado:*

*Rosario A. Muñoz Clares, Marieta Tuena de Gómez Puyou, María Eugenia Costas Basín, Arturo Rojo Domínguez y Jorge Peón Peralta.*

*A la Doctora Pilar Brocos.*

*A Jorge e Isaac por la ayuda mutua en el trabajo. Y a quienes compartieron conmigo el laboratorio: Edgar, Jimena, Francisco, Omar, Enrique, Abel, Aidee, Ariel, Mayeli, Xelhua.*

*Al Consejo Nacional para la Ciencia y la Tecnología (CONACYT) de México y la Dirección General de Estudios de Posgrado por las becas otorgadas.*

*Al Consejo Nacional para la Ciencia y la Tecnología (CONACYT) de México (Proyectos número J49811-Q y número 41328-Q).*

*A la Dirección General de Apoyo al Personal Académico (DGAPA) de la Universidad Nacional Autónoma de México (Proyecto número IN113302).*

*A la Dirección General de Servicios de Cómputo Académico (DGESCA) por el tiempo de cómputo utilizado en las simulaciones cuyos resultados se muestran en este trabajo.*

LUGAR DONDE SE DESARROLLÓ EL TEMA

*Laboratorio 102 Termofísica, Departamento de Fisicoquímica, Facultad de Química,  
Universidad Nacional Autónoma De México, México, D. F.*

# Índice general

<b>1</b>	<b>Motivación</b>	<b>1</b>
<b>2</b>	<b>Antecedentes</b>	<b>7</b>
2.1	Enfermedades parasitarias . . . . .	7
2.2	El Mal de Chagas . . . . .	7
2.3	Estructura y función de la triosafosfato isomerasa (TIM) . . . . .	9
2.4	Inhibidores enzimáticos de la TcTIM . . . . .	16
<b>3</b>	<b>Metodología</b>	<b>17</b>
3.1	Simulaciones moleculares . . . . .	17
3.2	Campo de fuerza . . . . .	18
3.3	Comportamiento dinámico de la TcTIM . . . . .	21
3.4	Parámetros de simulación . . . . .	21
3.4.1	Preparación de las simulaciones en agua y en mezclas agua-decano	22
3.4.2	Simulación en decano . . . . .	25
3.5	Métodos de análisis . . . . .	26
3.5.1	Visualización de las trayectorias . . . . .	26
3.5.2	Desviaciones cuadráticas medias (RMSD) . . . . .	26
3.5.3	Fluctuaciones cuadráticas medias (RMSF) . . . . .	27
3.5.4	Estructura secundaria . . . . .	27
3.5.5	Interacción de la TcTIM con el disolvente . . . . .	28

3.5.6	Dinámica de las moléculas de agua en el sistema . . . . .	28
3.5.7	Puentes de hidrógeno . . . . .	29
<b>4</b>	<b>Resultados y discusión</b>	<b>30</b>
4.1	Análisis visual de las trayectorias . . . . .	30
4.2	Análisis cuantitativo de la dinámica de la proteína . . . . .	31
4.2.1	Desviaciones cuadráticas medias (RMSD) . . . . .	33
4.2.2	Fluctuaciones cuadráticas medias (RMSF) . . . . .	36
4.2.3	Estructura secundaria . . . . .	42
4.3	Interacción de la TcTIM con el disolvente . . . . .	49
4.3.1	Agua . . . . .	50
4.3.2	Decano . . . . .	53
4.4	Dinámica de las moléculas de agua en el sistema . . . . .	55
4.5	Puentes de hidrógeno . . . . .	63
<b>5</b>	<b>Conclusiones</b>	<b>65</b>
<b>A</b>	<b>Dinámica molecular</b>	<b>67</b>
A.1	Simulaciones de dinámica molecular . . . . .	68
A.2	Métodos de búsqueda y minimización de energía . . . . .	72
A.3	Definiciones y unidades . . . . .	73
A.4	Campo de fuerza . . . . .	75
A.4.1	Interacciones entre átomos no enlazados . . . . .	76
A.4.2	Interacciones entre átomos enlazados . . . . .	79
A.5	Algoritmos . . . . .	82
A.5.1	Condiciones periódicas . . . . .	82
A.5.2	El concepto de grupo . . . . .	83
A.5.3	Algoritmos para la presión y temperatura . . . . .	84

# Índice de figuras

1.1	Representación esquemática de la relación entre estabilidad, flexibilidad y función de una proteína. . . . .	2
1.2	Superactividad enzimática en micelas invertidas. . . . .	3
2.1	Niveles de organización de las proteínas . . . . .	10
2.2	Ciclo de la glucólisis . . . . .	11
2.3	Estructura de la TcTIM . . . . .	12
2.4	Secuencia de estructura secundaria de la TcTIM . . . . .	13
2.5	Estructura de un barril de la TcTIM . . . . .	14
3.1	Conformación inicial del sistema D8 . . . . .	23
3.2	Conformación inicial del sistema D6 . . . . .	24
3.3	Conformación inicial del sistema D4 . . . . .	24
3.4	Esquema del diseño de las simulaciones . . . . .	25
4.1	Conformaciones inicial y final de la proteína en la simulación D8 . . . . .	32
4.2	Conformaciones inicial y final de la proteína en la simulación D6 . . . . .	33
4.3	Conformaciones inicial y final de la proteína en la simulación D4 . . . . .	34
4.4	RMSD de todos los átomos pertenecientes a la TcTIM. . . . .	35
4.5	RMSD de los átomos pertenecientes a la cadena central de la TcTIM . . . . .	36
4.6	RMSD de los átomos pertenecientes a las cadenas laterales de la TcTIM . . . . .	37
4.7	RMSF por residuo de la TcTIM. . . . .	39

4.8	RMSF de los átomos pertenecientes a las cadenas laterales de la TcTIM . . . . .	40
4.9	RMSF acumulado para la proteína y sus cadenas laterales . . . . .	42
4.10	Estructura secundaria de la TcTIM a lo largo las trayectorias . . . . .	43
4.11	RMSD de las hélices $\alpha$ en la TcTIM. . . . .	47
4.12	RMSD de las hebras $\beta$ en la TcTIM. . . . .	48
4.13	Número de residuos que se encuentran formando hélices $\alpha$ . . . . .	50
4.14	Representación gráfica de la hélice $\alpha_1$ del monómero A de la TcTIM. . . . .	51
4.15	Moléculas de agua a 4 Å de la TcTIM. . . . .	52
4.16	Moléculas de agua acumuladas por monómero a 4 Å de la TcTIM. . . . .	53
4.17	Número de átomos de carbono a 4 Å de la TcTIM. . . . .	54
4.18	Número de átomos de carbono acumulados por monómero a 4 Å de la TcTIM. . . . .	55
4.19	Incidencia de moléculas de agua a menos de 2 Å de la TcTIM, frecuencia entre 2 y 4 ns . . . . .	57
4.20	Incidencia de moléculas de agua a menos de 2 Å de la TcTIM, frecuencia entre 4 y 20 ns . . . . .	58
4.21	Incidencia acumulada de moléculas de agua a menos de 2 Å de la TcTIM. . . . .	59
4.22	Incidencia de moléculas de agua a menos de 4 Å de la TcTIM, frecuencia entre 2 y 4 ns . . . . .	60
4.23	Incidencia de moléculas de agua a menos de 4 Å de la TcTIM, frecuencia entre 4 y 20 ns . . . . .	61
4.24	Incidencia acumulada de moléculas de agua a menos de 4 Å de la TcTIM . . . . .	62
4.25	Puentes de hidrógeno intramoleculares en las trayectorias R, D8, D6 y D4. . . . .	63
4.26	Puentes de hidrógeno intramoleculares en las trayectorias R, D8, D6, D4 y D. . . . .	64
A.1	Fuerzas de Lennard-Jones. . . . .	76
A.2	Potencial de Coulomb. . . . .	78
A.3	Enlaces. . . . .	80

A.4	Ángulos simples. . . . .	80
A.5	Ángulos diédros propios. . . . .	81
A.6	Ángulos diédros impropios. . . . .	81

## Indice de Tablas

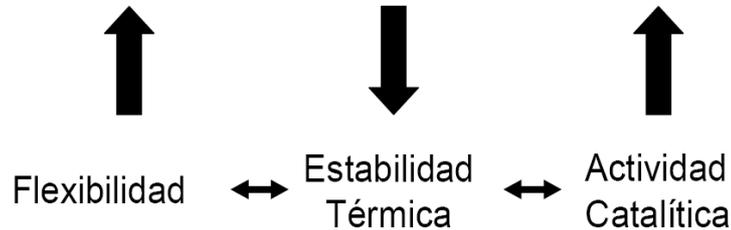
3.1	Diseño de las simulaciones. . . . .	21
A.1	Periodos de oscilación de movimientos atómicos . . . . .	70
A.2	Unidades básicas . . . . .	74
A.3	Unidades derivadas . . . . .	74

# Capítulo 1

## Motivación

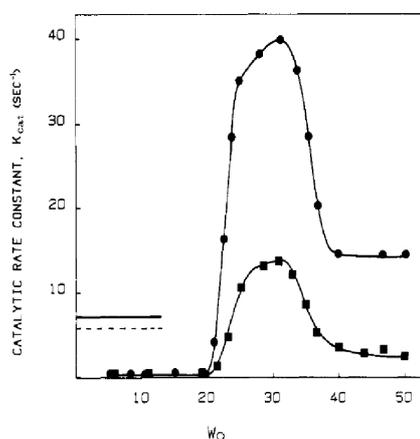
El agua juega un papel muy importante en la estructura, estabilidad, dinámica y por tanto en la propia función de las biomoléculas. En particular, se sabe que las moléculas de agua que están cercanas a un soluto en una disolución acuosa presentan diferencias tanto en su organización como en sus propiedades termodinámicas y dieléctricas con respecto a las moléculas de agua que forman parte del seno de la disolución. Las llamadas interacciones hidrofóbicas, denominadas así precisamente en referencia al agua como solvente, son las principales responsables de la mayoría de los niveles de organización que se observan en los organismos vivos y que incluyen la formación de micelas, vesículas, membranas celulares y el propio reconocimiento molecular. A nivel intramolecular, el ejemplo más representativo del efecto de las interacciones hidrofóbicas es el llamado plegamiento de proteínas y otros biopolímeros.

El comportamiento de las moléculas de agua en el entorno de una proteína ha despertado desde hace varios años mucho interés. Trabajos recientes han derivado en el desarrollo de nuevas técnicas experimentales destinadas a su estudio [1, 2]. Durante las últimas dos décadas se han invertido grandes esfuerzos en investigar el efecto del disolvente en la actividad de diversas enzimas. Principalmente, se trata de entender la dependencia de éstas con el medio acuoso. En algunos de sus trabajos, Zaks y Klivanov [3, 4, 5, 6] encontraron que cuando se suspende una enzima en un medio orgánico con bajas concentraciones de



**Figura 1.1:** Representación esquemática de la relación estabilidad, flexibilidad y función de una enzima.

agua, la actividad catalítica es comparable a la que se puede medir en disolución acuosa y que, en general, dicha actividad aumenta con el porcentaje de agua en la mezcla. Estos autores demostraron que cuanto más hidrofóbico es un disolvente, la cantidad de agua necesaria para que la enzima sea capaz de realizar su función es menor, y explicaron este fenómeno en términos del reparto de las moléculas de agua entre el entorno de la proteína y el propio solvente orgánico. A pesar de no disponer de evidencias a nivel microscópico, los mismos autores argumentaron que las enzimas en un medio no polar forman puentes de hidrógeno intramoleculares que las vuelven muy rígidas. Sin embargo, al rodear a las proteínas de una fina capa de moléculas de agua (del orden de 4 Å de espesor), los puentes de hidrógeno intramoleculares se rompen para interactuar con las moléculas del entorno. De esta manera las enzimas adquieren la flexibilidad necesaria para recuperar su capacidad catalítica. Esta capa de agua necesaria para que la proteína *funcione* se denominó posteriormente agua biológica [7]. Basándose en experimentos de otros autores, realizados con proteínas suspendidas en medios orgánicos, o encapsuladas en micelas invertidas con distintas concentraciones de agua, Gómez-Puyou y Tuena de Gómez-Puyou discuten en uno de sus trabajos [8] una interesante relación entre la flexibilidad de una enzima, su estabilidad térmica y su actividad catalítica. El agua biológica dota a la proteína de la flexibilidad necesaria para experimentar los cambios conformacionales que requiere en



**Figura 1.2:** *Actividad enzimática de la fosfatasa alcalina en función de la concentración de agua (ver Referencia [9]).*

el desempeño de su función. A su vez, cuanto más flexible es la molécula más fácil es desestabilizarla térmicamente (ver Figura 1.1).

Así, las proteínas en su evolución buscan un balance entre flexibilidad y estabilidad térmica que les permita realizar su función y al mismo tiempo mantener su estructura. En este trabajo los mismos autores llaman la atención sobre el sorprendente fenómeno de la superactividad enzimática, actividad catalítica en medios orgánicos con baja concentración de agua que para muchas proteínas es varias veces superior a la que muestran en solución acuosa [9] (ver Figura 1.2).

A pesar de los avances significativos y de la cantidad de trabajo realizado hasta el momento, no existe ninguna técnica experimental que proporcione información a nivel atómico, tanto de la proteína como de cada molécula del solvente, en escalas de tiempo que vayan desde los femtosegundos hasta los microsegundos. El estado del arte en el que se encuentran actualmente las simulaciones de dinámica molecular, gracias al desarrollo de campos de fuerzas más precisos, de algoritmos que aceleran los cálculos sin pérdida de exactitud y de procesadores de alta frecuencia, nos permite abordar de manera más realista este tipo de estudios con mucho detalle. Sin embargo, en los pocos trabajos basados en simulaciones de dinámica molecular en los que se ataca esta cuestión, las escalas

de tiempo que se han utilizado son relativamente limitadas, variando desde unas pocas decenas de picosegundos hasta los 5 nanosegundos [10, 11]. Estos tiempos, aunque se pueden considerar suficientes para el equilibrio local de las cadenas laterales de los residuos, son muy cortos en comparación con lo que duran los periodos de oscilación de bloques de estructura secundaria de proteínas e incluso con lo que tarda en equilibrarse la propia macromolécula en solución sometida al campo de fuerzas con el que se realizan las simulaciones.

Basándonos en los anteriores antecedentes, en este trabajo nos proponemos realizar simulaciones de dinámica molecular de la triosafosfato isomerasa de *Trypanosoma cruzi* (TcTIM) en agua, en decano, y en mezclas agua-decano con diferentes concentraciones de cada uno de los disolventes.

Para la selección de la enzima, fue decisiva su relación con la enfermedad denominada Mal de Chagas y el parásito que la produce. Hoy en día, el Mal de Chagas es considerada una de las enfermedades parasitarias más serias y con más repercusiones socioeconómicas en América Latina, en donde se estima que actualmente hay más de 20 millones de personas infectadas [12]. Dada su importancia, numerosos grupos de investigación dedican esfuerzos para tratar de combatir y erradicar dicha enfermedad. Muchas líneas de investigación apuntan sus esfuerzos a la búsqueda de un fármaco que sea capaz de inhibir la actividad enzimática de la TcTIM, que es una enzima clave para llevar a cabo el ciclo glucolítico de donde el parásito obtiene la energía necesaria para su existencia [13, 14, 15, 16, 17, 18]. El diseño de inhibidores está basado principalmente en el conocimiento estructural de la enzima, aunque suele obviarse la importancia del comportamiento dinámico de la misma porque esta información no es accesible experimentalmente de manera directa. Así, este trabajo tiene como uno de sus objetivos proporcionar información fundamental a nivel microscópico, de utilidad en el desarrollo de las investigaciones dedicadas al diseño de fármacos inhibidores de la TcTIM.

Existe un grupo grande de proteínas, en su mayoría enzimas, que tienen como característica común un patrón de plegamiento denominado barril  $\alpha/\beta$ . Aproximadamente, un

10 % de todas las enzimas presentan esta conformación estructural y debido a que participan en la catálisis de una gran diversidad de reacciones se han convertido en centro de atención. La enzima TcTIM pertenece a esta familia. Su estructura es homodimérica compuesta de dos subunidades caracterizadas por un barril formado por ocho hebras  $\beta$  y ocho hélices  $\alpha$  en su estructura terciaria. Se espera que los resultados que de aquí se obtengan contribuirán al conocimiento del comportamiento dinámico de las enzimas que siguen este patrón de plegamiento.

Cuando la proteína que se estudia es relativamente grande, como en el caso de la TcTIM, se requiere de un importante número de moléculas de agua para solvatar a la enzima dentro de la caja de simulación, lo cual se refleja en un incremento del tiempo de cómputo. La elección de un solvente no polar en lugar de agua conlleva a una disminución significativa en el costo computacional debido a que el tiempo involucrado en el cálculo de interacciones electrostáticas es nulo en el primer caso. En este trabajo se utilizará decano como disolvente principal, ya que la parametrización del campo de fuerzas que utilizamos para describir las interacciones de estas moléculas no incluye cargas parciales netas asociadas a los átomos del decano como sucede en el caso del agua. Esto último permite aumentar la escala de tiempo de las simulaciones de dinámica molecular con un ahorro de los recursos de cómputo hasta en un 70 %.

En resumen, las cuatro motivaciones principales que en nuestra opinión justificarían de manera individual, cada una por separado, el planteamiento de este trabajo y que serán cubiertas a lo largo del mismo son:

- i. Describir a nivel atómico el comportamiento dinámico que muestran las proteínas en solventes no polares con bajas concentraciones de agua.
- ii. Profundizar en el conocimiento de la estructura y comportamiento dinámico de la TcTIM, con el objetivo de proporcionar elementos útiles para combatir así el Mal de Chagas mediante el diseño de inhibidores selectivos de la enzima.

- iii. Extrapolar los resultados del análisis del comportamiento dinámico de la TcTIM a otras enzimas que compartan el mismo patrón de plegamiento (barril  $\alpha/\beta$ ).
- iv. Proponer una metodología útil para el estudio de macromoléculas por medio de dinámica molecular, que permita aprovechar al máximo los recursos computacionales gracias al uso de sustancias apolares como disolvente principal, junto con bajas concentraciones de agua.

## Capítulo 2

### Antecedentes

#### 2.1 Enfermedades parasitarias

Las enfermedades parasitarias han estado presentes en la vida del hombre a lo largo de su evolución sobre la tierra. El número de especies parasitarias a las que el ser humano ha estado expuesto durante ese lapso de tiempo es impresionante. Actualmente existen al menos 300 especies de helmintos y más de 70 especies de parásitos protozoarios, de las cuales aproximadamente 90 son causantes de las enfermedades parasitarias más graves que afectan al ser humano. Algunos de los parásitos con más repercusión en la salud humana son: amibas (amibiasis), *Giardia* (Giardiasis), *Trypanosoma brucei gambiense* y *Trypanosoma brucei rhodesiense* (Enfermedad del sueño), *Leishmania* (Leishmaniasis), diversas variedades de *Plasmodium* (P.), *falciparium*, *vivax*, *ovale*, y *malariae* (malaria), *Toxoplasma gondii* (Toxoplasmosis), microespiridinas, y *Trypanosoma cruzi* (Mal de Chagas) [19].

#### 2.2 El Mal de Chagas

El parásito protozoario *Trypanosoma cruzi* (*T. cruzi*) que provoca el Mal de Chagas fue descrito por primera vez en 1909 por el Dr. Carlos Chagas [20]. El parásito se transmite

al ser humano a través de la picadura de un insecto hematófago comunmente llamado chinche besucona.<sup>1</sup> El insecto o vector de transmisión adquiere el parásito después de succionar su alimento de un reservorio infectado, que suelen ser mamíferos salvajes que habitan las zonas tropicales. El parásito se reproduce dentro de los intestinos del insecto. Cuando el vector de transmisión se alimenta de un nuevo huésped sano, defeca cerca de la lesión las heces infectadas con los tripanosomas, que el mismo huésped hará llegar al punto de picadura por medio de frotamiento. Posteriormente, el parásito alcanzará el torrente sanguíneo donde se gestará el crecimiento del tripanosoma que dará origen al Mal de Chagas. Ésta resulta ser la principal vía de contagio, aunque se ha encontrado posible la transmisión a través de transfusiones sanguíneas en bancos de sangre no controlados y por transmisiones congénitas de madre a hijo.

El vector de transmisión es un insecto hematófago perteneciente a la familia de los triatómidos [21] y geográficamente se encuentra distribuido en casi toda América, principalmente en América Latina. Las manifestaciones más usuales del Mal de Chagas son las llamadas cardiopatías chagásicas que se presentan en forma de paros cardíacos, arritmias, y tromboembolias [22]. Sin embargo, en la mayoría de los casos la enfermedad es asintomática en pacientes infectados, lo que agrava la situación ya que muchas veces no es detectada para su tratamiento.

Con la finalidad de detener el crecimiento en el número de personas infectadas con esta enfermedad parasitaria, por un lado se han implementado campañas de erradicación del vector en las zonas de mayor tendencia a su crecimiento [23] y se busca el tamizaje de la enfermedad en bancos de sangre. Por otro lado, la investigación dirigida al diseño de fármacos que combatan eficazmente la enfermedad ha ido evolucionando. Los fármacos que actualmente se encuentran disponibles para el combate del Mal de Chagas son nitrofuranos, como el nifurtimox, y nitroimidazoles, como el benzinidazol, que fueron descubiertos empíricamente hace al menos tres décadas. Ambos fármacos funcionan de

---

<sup>1</sup>Llamado así por su habilidad para picar en las zonas del rostro y labios, en inglés kissing bug.

manera efectiva en la fase aguda<sup>2</sup> hasta los primeros años de la fase crónica, aunque su eficiencia varia de un individuo a otro. Su principal inconveniente son los efectos colaterales que provoca su uso, como anorexia, vómitos, neuropatías y dermatopatías, entre otros [24].

## 2.3 Estructura y función de la triosafosfato isomerasa (TIM)

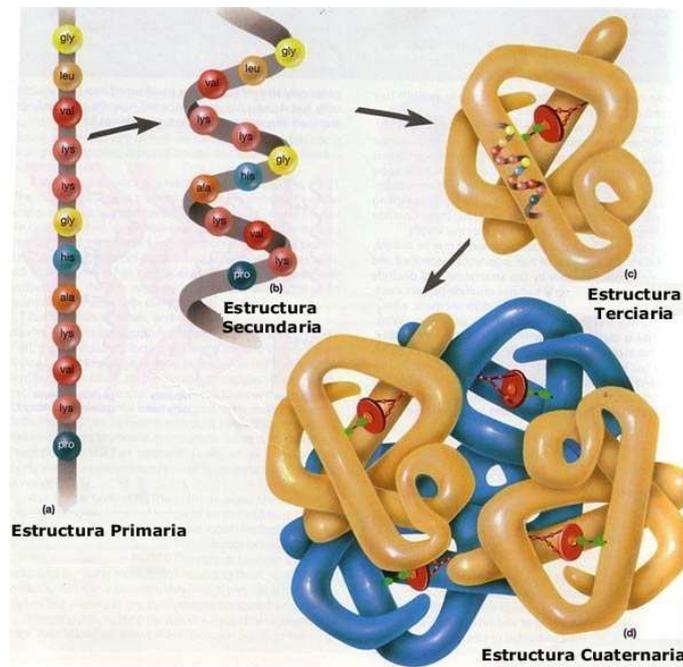
Las proteínas son polímeros formados por cadenas de aminoácidos unidos entre sí por medio de enlaces peptídicos, con la capacidad de arreglarse en varios niveles de organización de diferente complejidad. Poseen una conformación promedio característica, en torno a la cual fluctúan, en el medio en el que desempeñan su función biológica. Al experimentar cambios en la presión, temperatura o pH, entre otros parámetros, la proteína pierde su arreglo estructural y por tanto su función.

El primero de los niveles de complejidad en los que se organiza una proteína se define por la secuencia de los aminoácidos que la componen. La estructura secundaria es la estructura ordenada "local", definida por enlaces de hidrógeno principalmente dentro del esqueleto peptídico. Los elementos más comunes de estructura secundaria en las proteínas son hélices alfa ( $\alpha$ ) y hojas beta ( $\beta$ ), aunque existen otras conformaciones como los giros beta. Cuando grupos de hélices  $\alpha$  y hojas  $\beta$  se unen e interaccionan, dando lugar a la formación de diferentes estructuras más complejas, se obtiene el tercer nivel de organización llamado estructura terciaria. Existe un cuarto nivel (estructura cuaternaria), que es producto de la interacción no covalente de dos o más unidades de estructuras terciarias (ver Figura 2.1).

La estabilidad de una proteína se entiende como la capacidad que ésta tiene para mantener la estructura en sus diferentes niveles. Dicha estructura es estabilizada por enlaces covalentes entre cisteínas, interacciones hidrofóbicas, iónicas y puentes de hidrógeno en-

---

<sup>2</sup>El Mal de Chagas presenta tres fases evolutivas después del contagio, fase aguda, fase intermedia y fase crónica



**Figura 2.1:** Niveles estructurales de la disposición tridimensional de los aminoácidos en las proteínas.

tre cadenas laterales y debido a las interacciones de Van der Waals. La función biológica y la actividad de una proteína depende de su estructura que, a su vez, viene determinada por la secuencia de aminoácidos y por el solvente en que se encuentra disuelta.

La triosafosfato isomerasa (TIM) es una enzima homodimérica que cataliza la interconversión del gliceraldehído 3-fosfato y dihidroxiacetonafosfato, en el ciclo de la glucólisis [25]. La glucólisis es la secuencia metabólica, catalizada por la acción consecutiva de un grupo de 11 enzimas, por la cual se oxida la glucosa dentro de las células del organismo. El balance final de la glucólisis es la producción de dos moléculas de piruvato, dos equivalentes reducidos del NADH y dos moléculas de ATP. La glucólisis es la única vía por la cual los organismos vivos producen ATP en ausencia de oxígeno. En el proceso se distinguen dos etapas principales. En la primera de ellas, la glucosa se prepara para su catabolismo mediante su fosforilación, en el segundo de los pasos la triosafosfato isomerasa cataliza la reacción que produce el gliceraldehído-3-fosfato. En la segunda etapa,

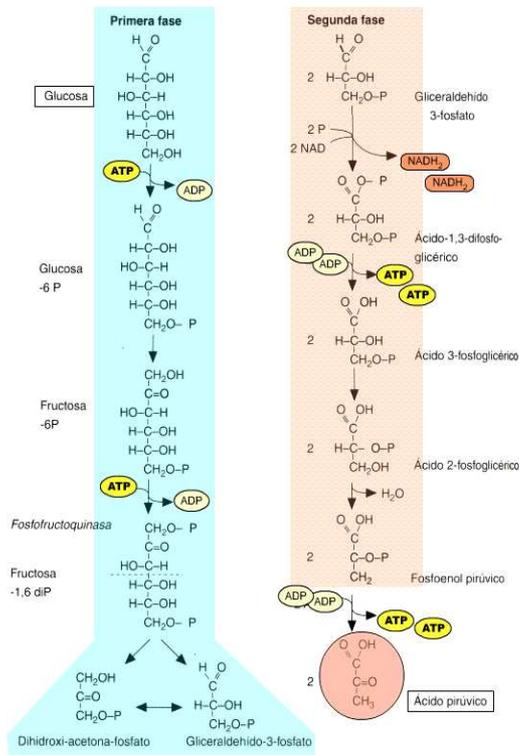
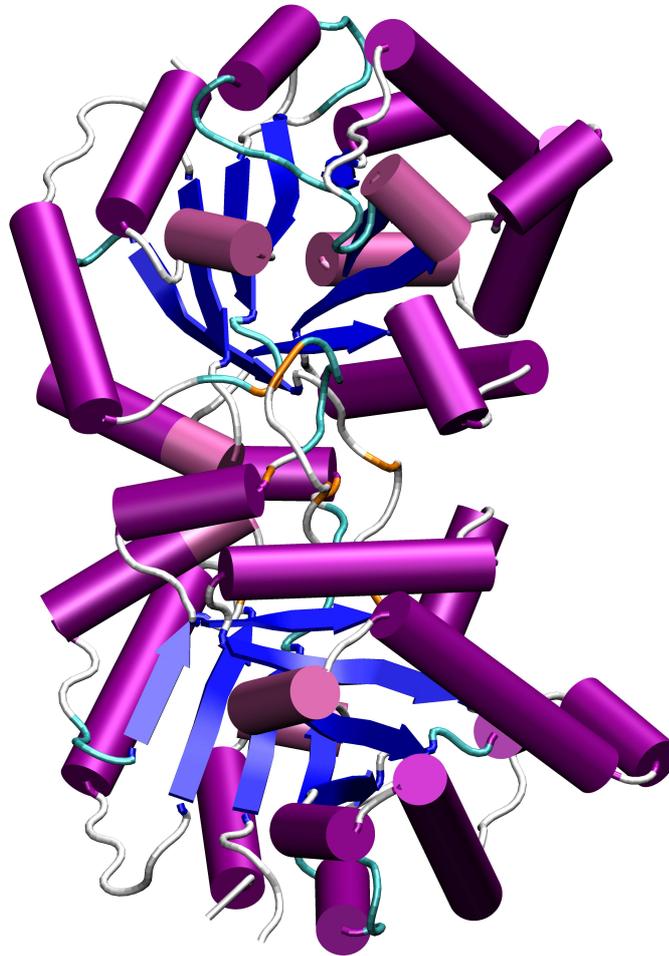


Figura 2.2: Ciclo de la glucólisis.

este último compuesto se convierte en ácido láctico. Muchos organismos como el *T. cruzi* obtienen su energía únicamente mediante este ciclo (Figura 2.2).

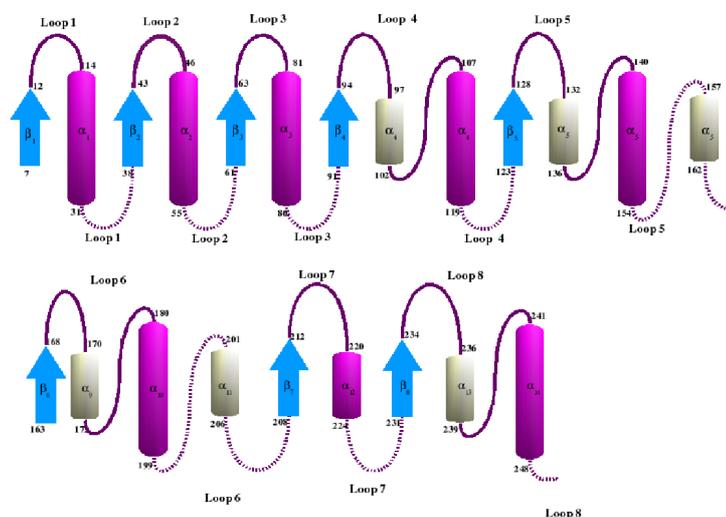
### Triosafofosato isomerasa del parásito *Trypanosoma cruzi* (TcTIM)

Los barriles  $\alpha/\beta$  representan un patrón de plegamiento formado por una secuencia alterada de hélices  $\alpha$  y hebras  $\beta$ . Esta conformación es la estructura terciaria más común de las estructuras cristalinas resueltas que se conocen. Aproximadamente un 10% de todas las enzimas presentan esta conformación estructural [26]. Los miembros de esta larga familia, incluyendo la TcTIM, participan como catalizadores en diferentes tipos de reacciones y precisamente esta diversidad las ha convertido en centro de atención. El nombre de barril  $\beta/\alpha$  se otorga a la topología en la cual la secuencia de aminoácidos comienza con una hebra  $\beta$ , y de forma análoga el barril se denomina  $\alpha/\beta$  si la secuencia inicia con una hélice  $\alpha$ .



**Figura 2.3:** Representación de la TcTIM. En azul se representa a las hebras  $\beta$ , en magenta a las hélices  $\alpha$ . La imagen original se tomó del PDB, código 1TCD [13].

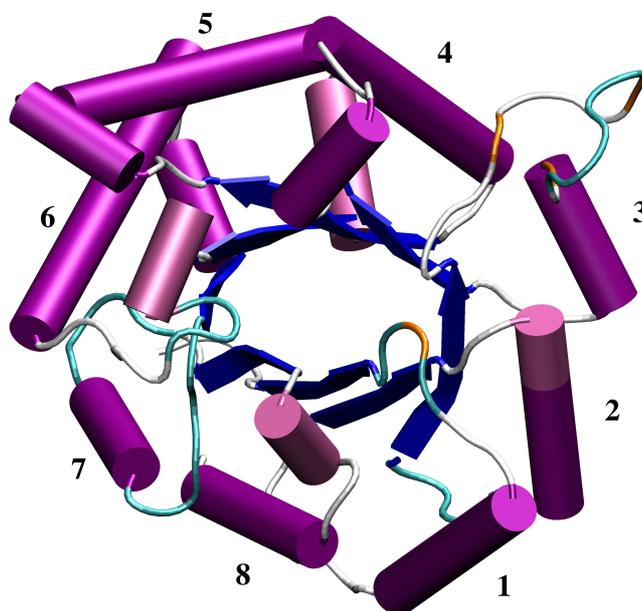
La estructura cristalina de la TIM de *Trypanosoma cruzi* (TcTIM) que hemos utilizado en este trabajo se encuentra disponible en el Protein Data Bank [27] con el código 1TCD, cristalizada por Maldonado y Col. con una resolución de 1.83 Å [13]. La TcTIM está formada por dos cadenas, cada una constituida por un barril( $\beta/\alpha$ )<sub>8</sub>, que constituyen las subunidades A y B de la enzima (Figura 2.3). Cada subunidad cuenta con ocho hebras  $\beta$  que alternan con ocho hélices  $\alpha$  unidas mediante lazos o *loops*. En la Figura 2.4 se esquematiza la secuencia de aminoácidos que forman a la TcTIM así como los bloques



**Figura 2.4:** Secuencia de aminoácidos y estructura secundaria de la triosafosfato isomerasa. En azul se representa a las hebras  $\beta$ , en magenta a las hélices  $\alpha$  y en gris a las hélices  $\alpha$  que forman loops.

de estructura secundaria en que se encuentran distribuidos, una representación de un barril de la TcTIM se puede observar en la Figura 2.5.

Una característica de los barriles de TIM es que la zona más hidrofóbica no es la parte interna del barril, sino el espacio formado entre las hebras  $\beta$  y las hélices  $\alpha$  [28]. El sitio activo de casi todas las TIM, se encuentra en los átomos de carbono terminales de las hebras  $\beta$ , lo que indica que residuos pertenecientes a algunos de los ocho *loops*  $\beta\alpha$  están involucrados en la topología del sitio activo. Los *loops*  $\beta\alpha$  poseen una longitud variable, pero generalmente son más largos que los *loops*  $\alpha\beta$  [29]. También se ha encontrado que los *loops*  $\beta\alpha$  están relacionados directamente con la función de la proteína, mientras que los *loops*  $\alpha\beta$  parecen tener una contribución a la estabilidad del barril [30]. Este mismo patrón topológico se ha encontrado en TIM de otras especies, como en el caso de la TIM de humano (ThTIM), TIM de *T. brucei* (TbTIM) y en otras especies de cuya estructura se tiene conocimiento. Típicamente, el número de residuos por monómero es de 250, aunque existen TIM's con tan sólo 200.



**Figura 2.5:** Representación de un monómero de la TcTIM. En azul se representa a las hebras  $\beta$ , en magenta a las hélices  $\alpha$ . La imagen original se tomó del PDB, código 1TCD [13].

La estructura del sitio activo de la TcTIM ha sido estudiada ampliamente. Se sabe que los residuos que participan de manera directa en el proceso catalítico de esta enzima son Lys-14, His-96 y Glu-168 del mismo monómero. El arreglo de estos residuos observado en la estructura cristalina de la TcTIM es muy similar al que se encuentra en la ThTIM y TbTIM, aunque existe una pequeña diferencia en la orientación del Glu168 [13]. Durante el proceso catalítico, el ligando se aísla del solvente como consecuencia de los cambios conformacionales del *loop* 6 que cubre el centro catalítico [31]. En la TcTIM, el *loop* 6 presenta diferentes conformaciones en los dos monómeros, una conformación cerrada en uno de ellos y abierta en el otro [13]. Se ha sugerido que las interacciones del residuo Cys-15 con el *loop* 3 de la otra subunidad juegan un papel crucial en la estabilización del dímero y en la acción enzimática de la proteína [17, 32, 33]. Los residuos

Asn-12 y Glu-168 participan en la transferencia de protones necesaria para la reacción entre el gliceraldehído 3-fosfato y la dihidroxiacetonafofosfato. El residuo Glu-168 parece tener importantes implicaciones en la catálisis porque al ser sustituido por un aspartato, la cinética de la reacción se ve fuertemente afectada.

Se sabe que las propiedades estructurales que generalmente presenta una enzima en un medio no polar con bajas concentraciones de agua no difieren sustancialmente de las que se observan en un medio acuoso [8]. Con la finalidad de entender el mecanismo de la catálisis en solventes orgánicos, la TcTIM fue cristalizada en presencia de hexano [34]. La estructura cristalina de la TcTIM en hexano fue obtenida con una resolución de 2 Å y se encuentra disponible en el Protein Data Bank con el código 1CI1. En ella se puede apreciar que la enzima mantuvo su conformación dimérica y estructura secundaria, no obstante, a que se encontraron algunas diferencias en la orientación de las cadenas laterales de algunos aminoácidos. En la superficie de la estructura cristalina (1CI1) se encontraron tres moléculas de hexano y en la estructura nativa (1TCD) en esos sitios específicos no se encontró alguna molécula de agua. La ubicación de las moléculas de hexano no fue observada cerca de alguno de los dos sitios catalíticos de la TcTIM [34]. La primera molécula se encontró localizada a una distancia menor a 4 Å de los residuos Arg-135, Thr-140 y Glu 186 de la subunidad B. Estos residuos forman parte del *loop* 5 y la hélice  $\alpha_6$ . La segunda y tercera molécula fueron encontradas a 3.8 Å de distancia entre ellas, situadas en la zona hidrofóbica que se forma entre las dos subunidades. Una de estas moléculas fue localizada a 4 Å de los residuos Ile-69, Tyr-103, Gly-104, Ile-109, y Lys-113 del monómero A; y a 4 Å de los residuos Tyr-102 y Tyr-103 del monómero B. La tercera molécula de hexano se situó a una distancia menor a 4 Å del residuo Phe-75 del monómero A y de los residuos Tyr-102 y Tyr-103 del monómero B [34].

En la estructura nativa de la TcTIM cristalizada en medio acuoso se encontraron 165 moléculas de agua. En la estructura cristalizada en presencia de hexano se encontraron 236 moléculas de agua, de las cuales sólo 127 se encontraron ocupando el mismo sitio que en la estructura nativa, la 38 moléculas de agua restantes pudieron haber sido desplazadas

por moléculas de hexano. Así, 109 moléculas de agua aparecieron en lugares en donde en la estructura nativa no se habían detectado [34].

## **2.4 Inhibidores enzimáticos de la TcTIM**

El comportamiento dinámico y estructural de los residuos que participan directamente en la acción catalítica de la enzima y en la estabilidad dimérica, así como los factores que afectan estas características, han sido punto de partida para el diseño de inhibidores de la TcTIM.

En este contexto se ha encontrado que algunos derivados del benzotiazol inactivan la acción de la TcTIM en concentraciones micromolares. Además, estos mismos compuestos son capaces de inhibir la actividad enzimática de la ThTIM sólo cuando se encuentran en concentraciones mucho mayores [14]. En el trabajo de Téllez-Valencia y cols. [14] se presenta un estudio con 10 diferentes moléculas, todas ellas derivadas del benzotiazol, capaces de inhibir la acción de la TcTIM y otras TIM's. El ácido 3-(2-benzotiazolilitio)-1-propansulfónico (compuesto 8), se caracterizó junto con la TcTIM por medios cristalográficos, calorimétricos y por métodos bioquímicos, para estudiar el efecto en la inactivación de la enzima, encontrándose que afecta la estructura de la interfase dimérica de la TcTIM [15].

Las energías de interacción entre diversos inhibidores de la TcTIM derivados del benzotiazol y la propia enzima se estudiaron también mediante métodos computacionales [35, 36]. Los resultados confirman la hipótesis de que los derivados del benzotiazol de mayor peso molecular son capaces de formar complejos energéticamente más estables con la TcTIM que con la ThTIM, mientras que los que poseen un tamaño menor forman complejos energéticamente similares con ambas enzimas [36].

## Capítulo 3

# Metodología

### 3.1 Simulaciones moleculares

En el Apéndice 1 se describen con detalle los principales algoritmos utilizados en la realización de las simulaciones que se presentan en este trabajo. En este capítulo nos limitaremos a describir brevemente la técnica de simulación utilizada y el nivel de cálculo, además de los valores de los parámetros más importantes que se consideraron.

De manera general, el conocimiento a nivel atómico de aspectos estructurales y dinámicos de sistemas organizados es de gran interés, ya que contribuye a comprender el comportamiento de dichos sistemas. La aplicación de técnicas computacionales en Química puede proporcionar información relevante a nivel atómico acerca de la estructura, estabilidad y comportamiento dinámico del sistema molecular de estudio.

Además, existen algoritmos diseñados para estimar diferencias de energía de Gibbs (entre otras propiedades termodinámicas), entre diferentes estados de un sistema, así como técnicas destinadas a la descripción y caracterización de reacciones químicas. La descripción de sistemas moleculares a temperaturas diferentes de cero, requiere de la generación de un conjunto estadístico representativo de configuraciones del sistema. A este conjunto se le conoce típicamente como ensamble. Las propiedades macroscópicas a determinar se definen a partir de promedios sobre el espacio de configuraciones. La simu-

lación de un sistema finito de partículas está asociada a un número determinado de grados de libertad que definen la precisión y nivel de cálculo con los que se describe el sistema.

Las simulaciones de dinámica molecular que se presentan en este trabajo utilizan las leyes de Newton de la mecánica clásica para predecir las estructuras y las propiedades de las moléculas. La trayectoria de un sistema de  $N$  átomos de masas  $m_i$  ( $i = 1, 2, \dots, N$ ) i definida por los vectores de posición  $\mathbf{r}_i$ , se describe por:

$$m_i \frac{\partial^2 \mathbf{r}_i}{\partial t^2} = \mathbf{F}_i, \quad i = 1, \dots, N. \quad (3.1)$$

La fuerza que se ejerce sobre cada partícula sometida a un potencial  $V_i$  viene dada por:

$$\mathbf{F}_i = -\nabla V_i \quad (3.2)$$

Las simulaciones de dinámica molecular se basan en la resolución simultánea de las ecuaciones 3.1 y 3.2 en pequeños pasos de tiempo. El proceso se repite de manera iterativa a la vez se almacenan con cierta frecuencia las coordenadas y velocidades que resultan del cálculo. Opcionalmente, se pueden grabar otros parámetros como las fuerzas, temperatura, presión o distintas contribuciones energéticas, para su posterior análisis.

## 3.2 Campo de fuerza

Un campo de fuerza es un grupo de ecuaciones que determinan la forma en que la energía potencial de una molécula varía en función de la ubicación de los átomos que la componen, así como de la distribución de los átomos de otras moléculas de su entorno. Además de las ecuaciones, la formulación de un campo de fuerza implica, por un lado, la definición de una serie de tipos de átomos que caracterizan a cada uno de ellos dentro de un contexto específico y, por otro lado, la definición de un conjunto de parámetros que relacionan las ecuaciones con los tipos de átomos.

En este trabajo se utilizó un campo de fuerza *conservativo* y se consideró que las contribuciones al potencial son debidas únicamente a las posiciones de los átomos. Esta suposición se basa en la Aproximación de Born-Oppenheimer, la cual considera que los movimientos electrónico y nuclear están desacoplados (ver Apéndice A). Debido al gran tamaño del sistema a estudiar y a que no se incluyen reacciones químicas, esta aproximación se considera apropiada para describir el sistema.

En la literatura se puede encontrar una gran variedad de campos de fuerza con diferentes niveles de aproximación en el cálculo y en algunas ocasiones específicos de acuerdo al tipo de sistema de interés. La elección de un campo de fuerza particular dependerá directamente de las propiedades del sistema a estudiar, así como de las propiedades que se desean determinar en el sistema.

El campo de fuerza típico para un sistema de  $N$  átomos de masas  $m_i$  ( $i = 1, 2, \dots, N$ ) con posición  $\mathbf{r}_i$  tiene la forma siguiente [37]:

$$V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = \sum_{\text{Enlaces}} \frac{1}{2} K_b [b - b_o]^2 + \quad (3.3)$$

$$\sum_{\text{Angulos}} \frac{1}{2} K_\theta [\theta - \theta_o]^2 + \quad (3.4)$$

$$\sum_{\substack{\text{diédros} \\ \text{impropios}}} \frac{1}{2} K_\xi [\xi - \xi_o]^2 + \quad (3.5)$$

$$\sum_{\substack{\text{diédros} \\ \text{propios}}} K_\varphi [1 + \cos n\varphi - \delta]^2 + \quad (3.6)$$

$$\sum_{\text{pares}(i,j)} [C_{12}(i,j)/r_{ij}^{12} - C_6(i,j)/r_{ij}^6] + \quad (3.7)$$

$$\sum_{\text{pares}(i,j)} \left[ \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} \right] \quad (3.8)$$

El primer término (3.3) representa la energía asociada a las interacciones entre dos átomos unidos entre sí por un enlace covalente expresado mediante un potencial armó-

nico. La constante  $b_o$  representa la longitud de enlace con el mínimo potencial y  $K_b$  la constante de fuerza. Los parámetros  $b_o$  y  $K_b$  dependerán del tipo de átomos en cada enlace. El segundo término del potencial (3.4) describe la energía asociada a la variación del ángulo entre dos enlaces interatómicos contiguos. Este grado de libertad se simulará también mediante un oscilador armónico que utiliza el propio ángulo como coordenada. Las energías asociadas a las oscilaciones entre enlaces covalentes formados por 4 átomos contiguos 1-4 se representan por los términos 3.5 y 3.6. La primera de estas contribuciones se describe nuevamente mediante un oscilador armónico parametrizado por el ángulo diédrico que se puede definir entre los planos generados por los dos grupos de 3 átomos sucesivos 1-2-3 y 2-3-4. Esta contribución se utiliza para definir el movimiento que los átomos que forman anillos experimentan perpendicularmente al plano de los mismos. La contribución 3.6 se refiere a los ángulos diédricos  $\varphi$  definidos de la misma manera que el ángulo  $\xi$  del término 3.5, parametrizados ahora mediante una función sinusoidal que permite giros completos de hasta  $360^\circ$ . Esta última contribución al potencial describe, por ejemplo, las torsiones de los átomos de C en cadenas alifáticas. Los últimos dos términos 3.7 y 3.8 describen la interacción entre todos los pares de átomos que se encuentran a una distancia de más de tres enlaces covalentes. Dicha contribución se expresa mediante un potencial de Lennard-Jones sumado a un potencial electrostático (de Coulomb) entre los átomos  $i$  y  $j$  con cargas  $q_i$  y  $q_j$  separados una distancia  $r_{ij}$ .

El listado de tipos de átomos es más complejo que la clasificación que aparece en la tabla periódica en función del número de protones. Por ejemplo, en un campo de fuerza clásico se consideran de manera distinta los hidrógenos de un grupo hidroxilo a los hidrógenos alifáticos, o los carbonos de un ciclo aromático de los carbonos de grupos ácidos. Para un mismo conjunto de ecuaciones dado dentro de un campo de fuerza es posible utilizar diferentes parámetros, los cuales serán elegidos en función del tipo de átomos de las moléculas a simular.

### 3.3 Comportamiento dinámico de la TcTIM

Por medio de simulaciones de dinámica molecular se presenta un estudio del comportamiento dinámico de la enzima TcTIM con diferentes solventes que son: decano, agua y decano con bajas concentraciones de agua. Las simulaciones que se muestran en este trabajo se realizaron con el paquete computacional GROMACS [38, 39, 40, 41], versión 3.2.1.

Como referencia para el análisis se realizó una simulación de la proteína en agua (R) y otra en decano (D). Para analizar la dinámica de la proteína en mezclas de agua-decano se plantearon tres diferentes sistemas (D8), (D6) y (D4), con diferentes concentraciones de agua y decano. El número de moléculas de agua y decano que se utilizaron en cada una de las simulaciones se presentan en la Tabla 3.1, la metodología utilizada para crear dichos sistemas se describe con detalle en la sección 3.4.1.

Simulación	Agua	decano	Tiempo
Trayectoria (R)	24825	0	60 ns
Trayectoria (D8)	3579	3402	40 ns
Trayectoria (D6)	2336	3208	40 ns
Trayectoria (D4)	1185	2825	40 ns
Trayectoria (D)	0	2587	40 ns

**Tabla 3.1:** *Diseño de las simulaciones.*

### 3.4 Parámetros de simulación

Las simulaciones de dinámica molecular se realizaron con el campo de fuerza GROMOS96 (43a2) para la enzima y el decano. El agua se simuló de manera explícita mediante el modelo SPC (Simple Point Charge) [42]. Para minimizar el número de moléculas de disolvente se tomaron como base para las condiciones periódicas cajas de forma

dodecaédrica. Con esta geometría se optimiza también la isotropía de la periodicidad. La temperatura, presión y el número de moléculas se mantuvieron constantes para todas las simulaciones. Con este fin la caja de simulación se acopló simultáneamente a un termostato y a un barostato de Berendsen con una temperatura y presión de referencia de 298 K y 1 bar [43]. Las constantes de acoplamiento utilizadas fueron de 0.1 y 0.5 ps para cada uno de los dos algoritmos respectivamente y el coeficiente de compresibilidad utilizado se tomó como un promedio entre el del agua y el decano, ponderado con las concentraciones de ambos disolventes para cada simulación.

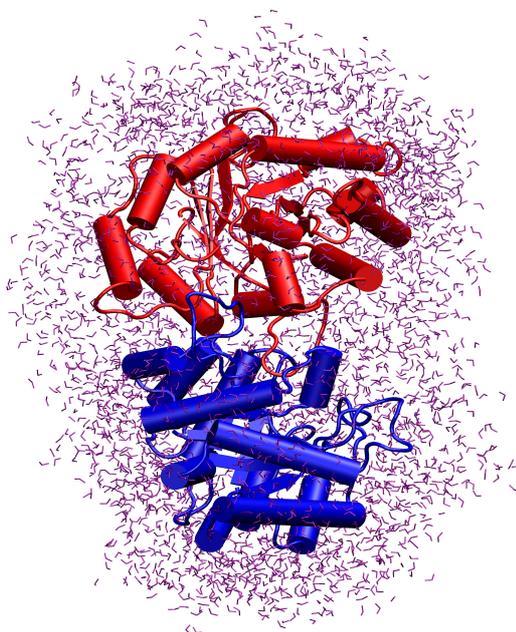
Las interacciones entre átomos no enlazados covalentemente fueron consideradas de la siguiente manera: para distancias menores a 0.9 nm, se calcularon a partir de la ley de Coulomb y fueron evaluadas en cada paso de tiempo. Las fuerzas entre átomos separados de 0.8 a 1.4 nm se calcularon de la misma manera, pero fueron evaluadas cada 5 pasos de tiempo manteniéndose constantes entre ellos; para distancias mayores a 1.4 nm, se consideró un medio continuo con una constante dieléctrica  $\epsilon = 78.0$ , valor con el que fue parametrizado el campo de fuerza utilizado en nuestros cálculos. El paso de tiempo utilizado en todos los casos fue de 2 femtosegundos (fs). Las longitudes de los enlaces y el ángulo H-O-H en el agua se mantuvieron constantes por medio del algoritmo SETTLE [44].

### **3.4.1 Preparación de las simulaciones en agua y en mezclas agua-decano**

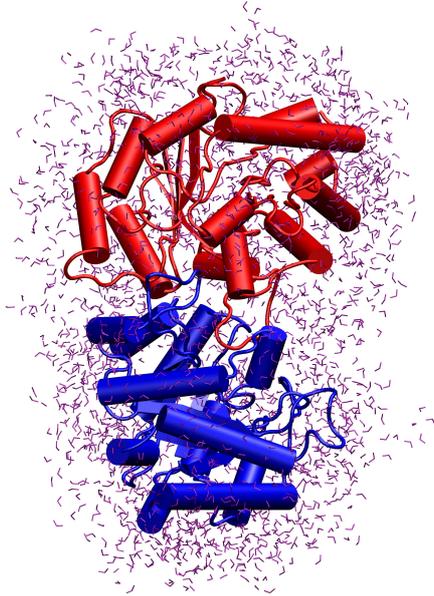
Para simular la TcTIM en agua se tomaron las coordenadas de la estructura cristalina que aparecen en el Protein Data Bank con el código 1TCD. Dicha estructura se colocó en una caja dodecaédrica de tamaño tal que la distancia mínima entre cualquier átomo de la proteína y las paredes de la caja fuera de al menos 2 nm. El espacio libre entre la proteína y la caja fue cubierto con pequeñas cajas de agua previamente equilibradas y el sistema resultante se simuló durante 60 ns. De esta manera se creó la simulación de referencia

(R). Las simulaciones en las que se utilizó una mezcla de agua y decano como disolvente se realizaron a partir de la conformación obtenida de la simulación R después de 20 ns de dinámica molecular (a esta conformación le llamaremos R20).

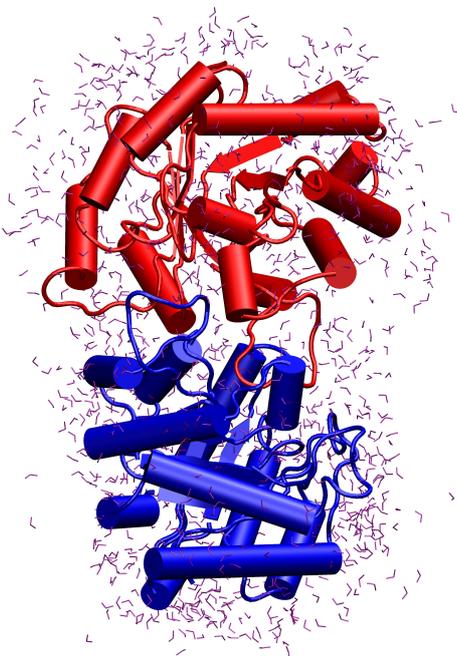
Por medio de un programa desarrollado en código awk, se suprimieron sucesivamente todas las moléculas de agua que tenían algún átomo a una distancia de la proteína mayor a 8, 6 y 4 Å respectivamente, con lo que se obtiene así 3 nuevas cajas de simulación. El tamaño de las cajas resultantes se aumentó ligeramente de manera que ningún átomo de la proteína y del agua estuviese a más de 2 nm de las paredes de la misma y los sistemas resultantes se llenaron de moléculas de decano preequilibradas. A las simulaciones realizadas a partir de estas conformaciones las denominaremos de aquí en adelante D8, D6 y D4, respectivamente. Las fotografías iniciales de las simulaciones D8, D6 y D4, excluyendo las moléculas de decano para mayor claridad, se muestran en las Figuras 3.1, 3.2 y 3.3. Este diseño de las simulaciones se esquematiza en la Figura 3.4.



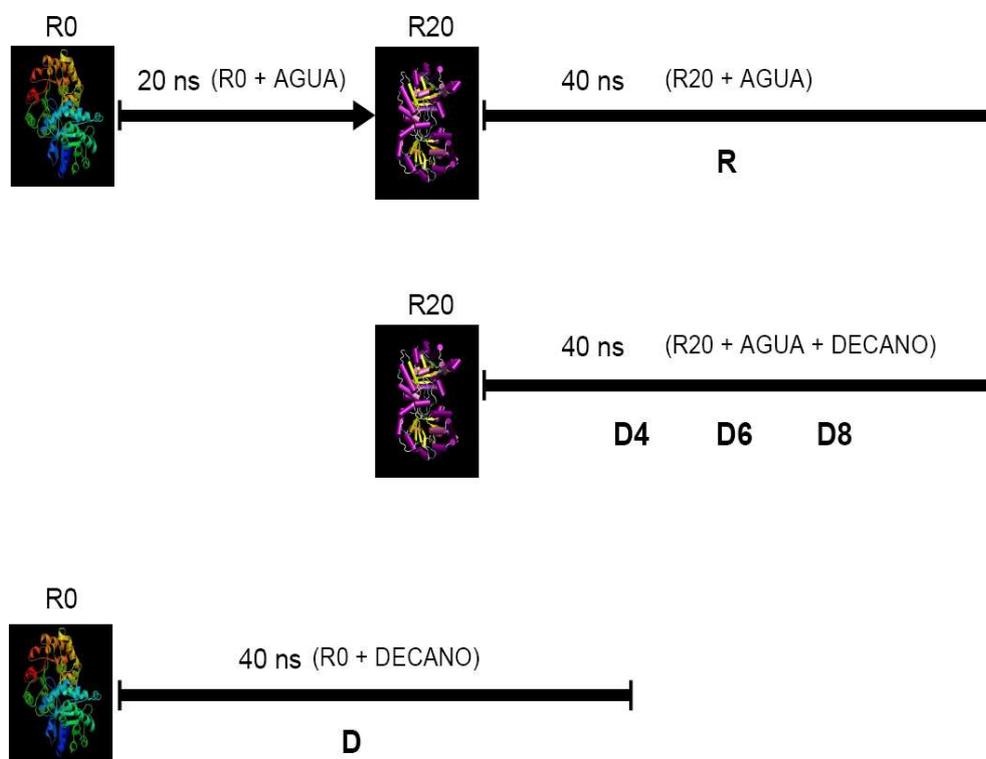
**Figura 3.1:** Conformación inicial de la simulación D8, con una capa de agua inicial de 8 Å (3579 moléculas de agua). En azul se representa el monómero A, en rojo el monómero B y en magenta las moléculas de agua.



**Figura 3.2:** Conformación inicial de la simulación D6, con una capa inicial de agua de 6 Å (2336 moléculas de agua). En azul se representa el monómero A, en rojo el monómero B y en magenta las moléculas de agua.



**Figura 3.3:** Conformación inicial de la simulación D4, con una capa inicial de agua de 4 Å (1175 moléculas de agua). En azul se representa el monómero A, en rojo el monómero B y en magenta las moléculas de agua.



**Figura 3.4:** Esquema representativo del diseño de las simulaciones realizadas.

### 3.4.2 Simulación en decano

Para la realización de la simulación en decano puro se utilizó la estructura de la TcTIM que aparece en el Protein Data Bank [27] con el código 1TCD [13]. La enzima se colocó en una caja con geometría dodecaédrica capaz de llenar todo el espacio cuando se repite de manera periódica en las tres dimensiones del espacio. El tamaño de los sistemas fue seleccionado de tal manera que la distancia mínima entre cualquier átomo de la proteína y las paredes de la caja fuera como mínimo de 2 nm. A esta simulación se le denominó simulación D.

## 3.5 Métodos de análisis

En esta sección se explicarán brevemente los métodos utilizados para el análisis de las distintas trayectorias.

### 3.5.1 Visualización de las trayectorias

La primera parte del análisis consistió en visualizar las distintas trayectorias obtenidas mediante el uso de los programas VMD [45] y Rasmol [46]. Esta inspección visual del comportamiento de las distintas moléculas del sistema fue importante de cara a diseñar las siguientes fases del análisis. Mediante el empleo de las herramientas de visualización fue posible monitorear el comportamiento de la enzima y del disolvente. Un análisis estático de las estructuras permite hacer comparaciones entre los estados inicial y final. Además, gracias a que se conoce la posición y velocidad de todos los átomos de la proteína y los disolventes, es posible observar y analizar la evolución de todo el sistema en función del tiempo.

### 3.5.2 Desviaciones cuadráticas medias (RMSD)

Para todas las trayectorias se calculó la desviación cuadrática media, definida por la ecuación (3.9), en función del tiempo. Este análisis calcula la desviación cuadrática media de algunos átomos en una molécula con respecto a una estructura de referencia mediante el programa *g\_rmsd* del paquete computacional GROMACS [38]. La primer tarea del programa es realizar un ajuste de la estructura por el método de mínimos cuadrados a otra de referencia ( $t_0$ ) y posteriormente calcular la RMSD:

$$RMSD(t, t_0) = \left[ \frac{1}{M} \sum_{i=1}^N m_i \| \mathbf{r}_i(t) - \mathbf{r}(t_0) \|^2 \right]^{\frac{1}{2}} \quad (3.9)$$

donde  $M = \sum_{i=1}^N m_i$  y  $\mathbf{r}_i(t)$  es la posición del átomo  $i$  a un tiempo  $t$ .

Este análisis proporciona una medida cuantitativa no específica de la diferencia entre cada una de las conformaciones en una trayectoria y una estructura de referencia. Así también, la evolución de la curva RMSD vs tiempo en una trayectoria de simulación proporciona un criterio que ayuda a considerar la escala de tiempo en la cual un sistema alcanza el equilibrio.

### 3.5.3 Fluctuaciones cuadráticas medias (RMSF)

El análisis de las fluctuaciones cuadráticas medias fue calculado con el programa *g\_rmsf* del paquete computacional GROMACS [38]. El programa calcula las fluctuaciones cuadráticas medias, es decir, la desviación estándar de la diferencia entre las posiciones atómicas en una conformación del sistema ajustada a otra estructura de referencia. Los resultados se presentan como un promedio para cada residuo.

$$RMSF(i) = \langle (\Delta \mathbf{r}_i^2) \rangle^{\frac{1}{2}} = \langle [\mathbf{r} - \langle \mathbf{r} \rangle]^2 \rangle^{\frac{1}{2}} \quad (3.10)$$

### 3.5.4 Estructura secundaria

Este análisis consistió en calcular la estructura secundaria de cada residuo a partir de la trayectoria del sistema. Para lograr este objetivo se utilizó el programa *do\_dssp* del paquete computacional GROMACS. Este programa utiliza la información del código DSSP para el cálculo de la estructura secundaria en la proteína; DSSP es un acrónimo para "Dictionary of Protein Secondary Structure". El DSSP [47] identifica por medio de un código de una letra a cada tipo de estructura secundaria. Según este código, la estructura se asigna en base a los patrones de enlaces de tipo puentes de hidrógeno entre aminoácidos, propuestos inicialmente por Pauling. Las hélices  $\alpha$  se representan con la letra H y las hebras  $\beta$  con la letra E. Si las hélices  $\alpha$  y hebras  $\beta$  que se forman son muy cortas, éstas son designadas por las letras T para las hélices  $\alpha$  y B para las hebras  $\beta$ . Existen otros patrones de estructura secundaria los cuales no se presentan en este trabajo.

### 3.5.5 Interacción de la TcTIM con el disolvente

Esta parte del análisis consistió en monitorear el número de moléculas de agua que se encontraban a una distancia menor de 4 Å de los átomos que conforman cada residuo. Para esto fue necesario elaborar un programa que calcula la distancia entre los oxígenos de las moléculas de agua con respecto a los átomos de cada residuo de las proteínas y que posteriormente elimina aquéllos que no cumplen con la restricción de encontrarse a una distancia menor de 4 Å. Se diseñó un procedimiento similar para cuantificar el número de átomos de decano pertenecientes al decano que se encontraron a una distancia menor de 4 Å del centro de masas de cada átomo de la TcTIM.

### 3.5.6 Dinámica de las moléculas de agua en el sistema

Durante la simulación se almacenan, a cada paso de tiempo, las posiciones de los átomos del sistema. Para el análisis se utilizó este archivo. La información se analizó a través de un programa en código awk desarrollado específicamente para este análisis. Este programa calcula la distancia entre el centro de masas de cada oxígeno del agua y cada uno de los átomos de la proteína en cada conformación. El análisis de frecuencias consistió en calcular el número de veces que una misma molécula de agua aparece a una distancia menor a 4 Å de un residuo de la TcTIM.

Por ejemplo en el sistema donde fue simulada la TcTIM con 25000 moléculas de agua (simulación R), el programa tomó cada uno de los 4862 átomos de la proteína y calculó la distancia a cada uno de los 25000 oxígenos pertenecientes a las moléculas de agua durante una trayectoria de 20 ns. Después de cada cálculo el programa almacenó las coordenadas e identificadores de aquellas moléculas que aparecieron a una distancia menor a 4 Å ó 2 Å de un residuo de la TcTIM. Este proceso se realizó iterativamente sobre 2000 conformaciones de la trayectoria. El archivo generado contiene la información que permite cuantificar el tiempo de residencia de una misma molécula de agua cercana a un residuo durante la simulación. Si se calcula el número de veces en que una molécula

apareció a distancia menor a 4 Å ó 2 Å de un residuo dado, se obtiene la frecuencia de incidencia de las moléculas de agua en los residuos de la proteína. En un análisis posterior se realizó una suma de las moléculas de agua totales que permanecieron en cada residuo en un entorno definido (4 Å ó 2 Å) durante un tiempo neto de 2-4 ns y 4-20 ns.

El tiempo de cómputo que se requiere para hacer el cálculo en esas 2000 conformaciones es comparable a una simulación molecular. Debido al gran costo computacional se decidió no extender el cálculo hasta las 4000 conformaciones. En las simulaciones D8, D6 y D4 se siguió el mismo procedimiento.

### 3.5.7 Puentes de hidrógeno

Un puente de H es un tipo particular de interacción electrostática atractiva localizada entre un hidrógeno con una carga parcial positiva y un átomo aceptor de protones. Existen diferentes criterios geométricos que se utilizan en la inspección de conformaciones atómicas para contabilizar puentes de H [38]. Para este análisis se utilizó el programa *g\_hbonds* del paquete GROMACS. Los parámetros que utiliza el programa para definir la formación de un enlace de este tipo son una distancia máxima entre el hidrógeno y el grupo aceptor de 3.5 Å ( $r \leq r_{HB} = 3.5 \text{ \AA}$ ), y un ángulo máximo de 30° ( $\alpha \leq \alpha_{HB} = 30^\circ$ ), entre los átomos X-H—A, donde X representa al átomo al cual el H está unido covalentemente. Este cálculo puede realizarse tanto entre átomos de la misma molécula (puentes de H intramoleculares) como entre átomos de moléculas distintas, soluto-soluto, soluto-solvente, etc (puentes de H intermoleculares).

## Capítulo 4

# Resultados y discusión

### 4.1 Análisis visual de las trayectorias

Un análisis visual de las trayectorias permite observar en función del tiempo la evolución de la estructura de la proteína, además del ordenamiento local del disolvente en el entorno de ésta. Debido a la diferencia en polaridad de los disolventes, el comportamiento y la interacción con la enzima resulta de gran interés. La composición del disolvente en los sistemas simulados es diferente en cada caso, en las simulaciones R y D los disolventes fueron agua pura y decano puro respectivamente. En las simulaciones D8, D6 y D4 se utilizó como disolvente una mezcla agua-decano con diferentes concentraciones de ambos compuestos en cada una de ellas. Las Figuras 4.1, 4.2 y 4.3 muestran imágenes de las configuraciones inicial y final, después de 40 ns de simulación de dinámica molecular, obtenidas de las trayectorias D8, D6 y D4. En ellas se muestran en color magenta las moléculas de agua totales empleadas en cada sistema, y en color verde los átomos de moléculas de decano que se localizaron a menos de 4 Å de la enzima. Como se explicó con detalle en el Capítulo 3, las configuraciones iniciales de dichas trayectorias fueron creadas a partir de la conformación de la trayectoria R después de 20 ns de simulación manteniendo una capa de agua alrededor de la proteína de 3579 (D4), 2336 (D6) y 1185 (D4) moléculas, respectivamente.

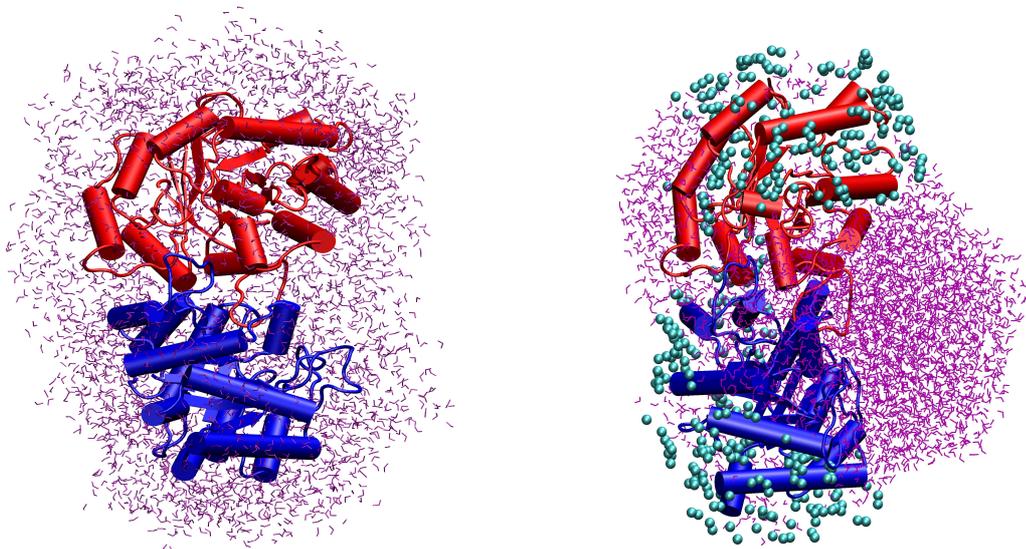
A través de las imágenes que se presentan es posible observar una distribución no homogénea del disolvente alrededor de la proteína al término de cada trayectoria. En los tres casos una gran parte de las moléculas de agua que al inicio formaban una capa de agua de espesor uniforme, se movieron para acumularse en el espacio que se forma entre los dos monómeros de la enzima. La formación de este conglomerado de moléculas de agua, que se observa en todas las figuras, es consecuencia de la insolubilidad entre los dos disolventes. Su formación se relaciona también con la propia geometría de la proteína que proporciona una región favorable para que el agua minimice la superficie interfacial con el decano.

En los tres casos, el espacio que dejaron las moléculas de agua al mudarse a la cavidad de la enzima durante los 40 ns de simulación fue ocupado por moléculas de decano. Además es importante señalar que una cantidad importante de moléculas de agua ocuparon sitios específicos alrededor de la enzima y muy pocas de ellas se fueron al seno del disolvente apolar, como ya era de esperarse debido a la poca afinidad entre ambos solventes.

## **4.2 Análisis cuantitativo de la dinámica de la proteína**

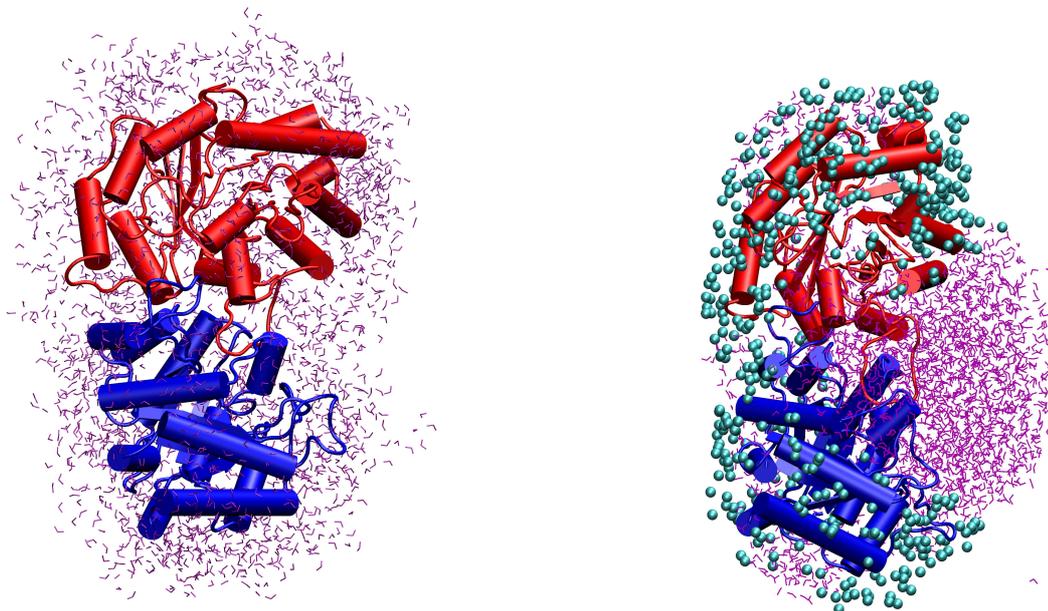
Un primer análisis de la dinámica de la proteína se realizó calculando la desviación cuadrática media (RMSD) de la enzima y las fluctuaciones cuadráticas medias (RMSF) de los diferentes residuos en cada trayectoria.

La RMSD se calculó para todos los átomos de la proteína a lo largo de la trayectoria. De manera independiente, este mismo parámetro se evaluó para los átomos que forman la cadena central y para los átomos que pertenecen a las cadenas laterales de todos los residuos. Análogamente, el valor de RMSF se evaluó considerando dos grupos de átomos diferentes, en un grupo se incluyeron todos los átomos de la proteína y en otro independiente los átomos que sólo pertenecen a las cadenas laterales (esta última selección de átomos excluye a la parte más rígida de la proteína). En las trayectorias D8, D6, D4 y



**Figura 4.1:** *Conformaciones inicial y final de la proteína en la simulación D8, con una capa de agua inicial de 8 Å (3579 moléculas de agua en torno a la proteína). En azul se representa al monómero A, en rojo al monómero B, en magenta a las moléculas de agua y en verde se representan los átomos de carbono pertenecientes a las moléculas de decano encontrados a una distancia menor a 4 Å.*

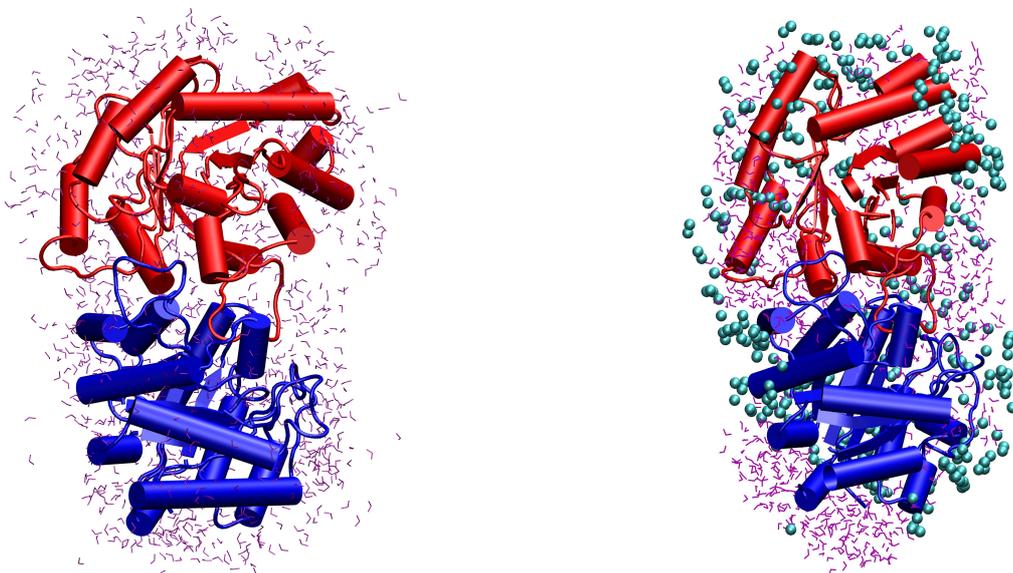
D el análisis de RMSD que se presenta se evaluó a lo largo de los 40 ns de simulación y en la trayectoria R éste fue evaluado sobre los últimos 40 ns de un total de 60 ns de simulación. El cálculo de RMSF en las trayectorias D8, D6, D4 y D fue realizado sobre los últimos 30 ns de simulación ya que los primeros 10 ns se consideraron como tiempo de equilibración del sistema. Para la realización de este análisis en la trayectoria R únicamente se utilizaron los últimos 30 ns de simulación. Los primeros 30 ns se excluyeron con el objetivo de igualar los tiempos sobre los cuales se realizaron todos los cálculos en las diferentes trayectorias.



**Figura 4.2:** Conformaciones inicial y final de la proteína en la simulación D6, con una capa de agua inicial de 6 Å (2336 moléculas de agua entorno a la proteína). En azul se representa al monómero A, en rojo al monómero B, en magenta a las moléculas de agua y en verde se representan los átomos de carbono pertenecientes a las moléculas de decano encontrados a una distancia menor a 4 Å.

### 4.2.1 Desviaciones cuadráticas medias (RMSD)

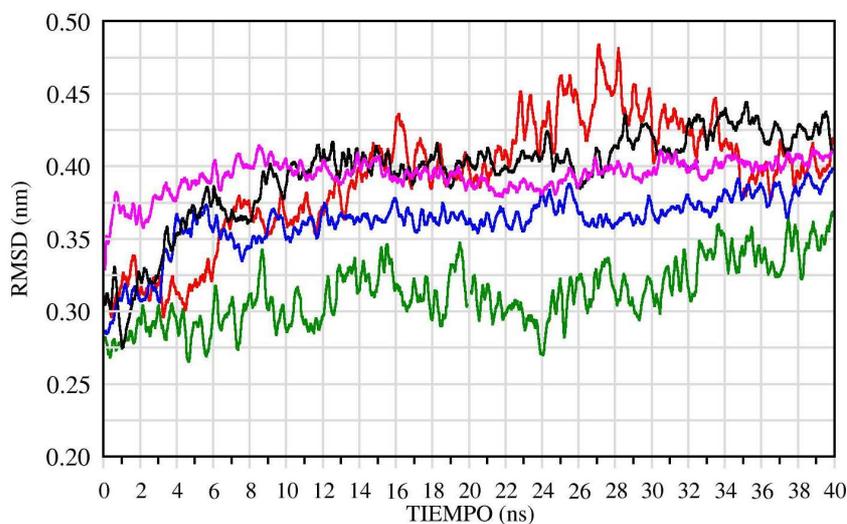
El valor de la RMSD calculada para cada trayectoria nos proporciona una medida neta, no específica, de los cambios conformacionales que tienen lugar en la estructura de la proteína respecto a la estructura que se toma como referencia. En la Figura 4.4 se presentan los valores de la desviación cuadrática media (RMSD) de la proteína en función del tiempo para las trayectorias R, D8, D6, D4 y D. En la Figura 4.5 se presenta únicamente el cálculo de RMSD para los átomos pertenecientes a la cadena central o *backbone*. Como referencia para el cálculo de RMSD en todos los casos que se presentan se utilizaron las



**Figura 4.3:** Conformaciones inicial y final de la proteína en la simulación D4, con una capa de agua inicial de 4 Å (1175 moléculas de agua entorno a la proteína). En azul se representa al monómero A, en rojo al monómero B, en magenta a las moléculas de agua y en verde se representan los átomos de carbono pertenecientes a las moléculas de decano encontrados a una distancia menor a 4 Å.

coordenadas de la estructura cristalizada de la TcTIM.

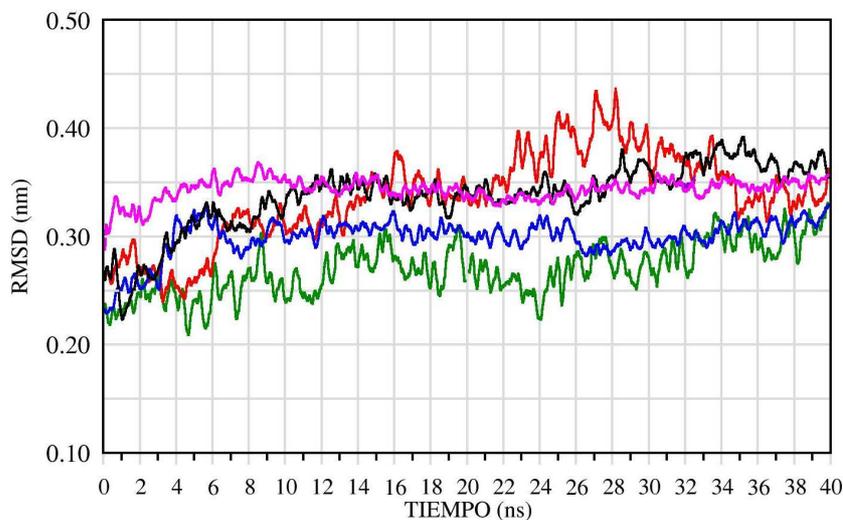
En las Figuras 4.4 y 4.5 se observa que los valores de RMSD obtenidos para cada trayectoria son ligeramente diferentes entre sí. Aproximadamente durante los primeros 10 ns, es posible observar que la tendencia de la RMSD es ligeramente creciente y que a partir de los 10 ns lo que se observa son fluctuaciones con promedios aparentemente constantes en todos los casos. Las fluctuaciones en los valores de RMSD en la región donde el promedio es prácticamente constante podrían atribuirse a transiciones reversibles en la estructura secundaria de la proteína; bajo esta hipótesis en un análisis posterior se presenta un estudio más específico de este nivel de estructura. En el medio acuoso, la proteína adquiere la flexibilidad necesaria para experimentar los cambios conformacionales que



**Figura 4.4:** Desviación cuadrática media (RMSD) de todos los átomos pertenecientes a la TcTIM. Como referencia se utilizó la estructura cristalina 1TCD PDB. Curva verde: Trayectoria R; curva roja: Trayectoria D8; curva negra: Trayectoria D6; curva azul: Trayectoria D4, curva magenta: Trayectoria D. Cada punto en las curvas representa un promedio sobre 200 ps.

se requieren para lograr su función catalítica, por lo cual la enzima no puede encontrarse rígida y debe ser normal observar transiciones locales reversibles en su estructura.

El mismo cálculo de RMSD se realizó sobre las cadenas laterales excluyendo a los átomos de la cadena principal de la proteína y los resultados se presentan en la Figura 4.6. En esta gráfica se encuentra que la tendencia de las curvas para alcanzar un valor estable durante la simulación es similar a la que se observa en la Figura 4.4. En promedio el valor de RMSD es ligeramente mayor cuando se evalúa para los átomos que forman las cadenas laterales de la proteína, ya que éstos están compuestos por los átomos que más cambios conformacionales presentan y que se encuentran más expuestos al disolvente. En el caso de la Figura 4.5, donde fue considerada únicamente la cadena central de la proteína para su evaluación, se obtuvieron los valores más bajos. Lo anterior es consecuencia de que la cadena principal está formada sólo por los átomos que conforman el esqueleto de la



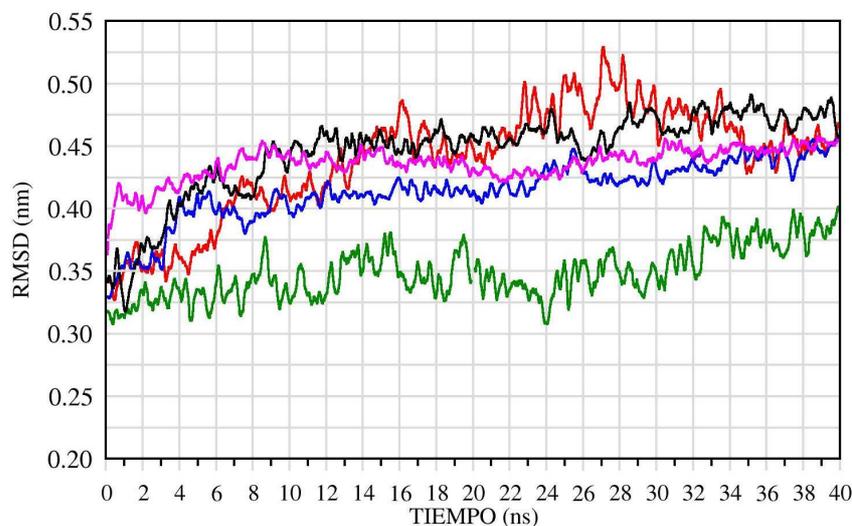
**Figura 4.5:** Desviación cuadrática media (RMSD) de todos los átomos pertenecientes a las cadenas centrales de la TcTIM. Se tomó como referencia la estructura cristalina ITCD PDB. Los colores de las curvas tienen la misma notación de la Figura 4.4.

estructura terciaria y que por tanto manifiestan una mayor rigidez.

Las gráficas de los valores de RMSD en función del tiempo son útiles para obtener un primer criterio de equilibrio en las simulaciones de estudio. El análisis sugiere que es aconsejable descartar los primeros 10 ns de cada trayectoria debido a la evolución que la RMSD muestra en este primer intervalo. Por esta razón, en el análisis de fluctuaciones que a continuación se presenta se utilizaron sólo los últimos 30 ns de cada trayectoria.

#### 4.2.2 Fluctuaciones cuadráticas medias (RMSF)

La movilidad de los residuos de cada monómero se cuantificó en términos de la RMSF. Como se explica en el Capítulo 3 en este análisis se calcula la desviación estándar de la distribución de las posiciones atómicas para cada residuo que tuvieron lugar durante una trayectoria. Para realizar este cálculo es necesario ajustar cada estructura de la trayectoria



**Figura 4.6:** Desviación cuadrática media (RMSD) de los átomos pertenecientes a las cadenas laterales de la TcTIM. Se tomó como referencia la estructura cristalina 1TCD PDB. Los colores de las curvas tienen la misma notación de la Figura 4.4.

a una de referencia, por lo cual en todos los casos se utilizó la última conformación de la proteína en cada sistema.

Este análisis nos permitirá identificar zonas de diferente flexibilidad o rigidez que a simple vista o en una estructura fija dada es imposible de observar. El cálculo de RMSF se realizó de manera independiente para dos grupos de átomos: el primero formado por todos los átomos de la proteína y el segundo únicamente por los átomos que pertenecen a las cadenas laterales de cada residuo.

Los valores de RMSF calculados para todos los átomos de la proteína en cada una de las trayectorias se muestran en la Figura 4.7. Como se mencionó anteriormente, la enzima triosafosfato isomerasa es homodimérica, es decir, que se encuentra formada por la misma cantidad y tipo de residuos en su secuencia, estructura secundaria y terciaria en cada uno de los monómeros. Por esta razón los resultados se presentan para cada subunidad de la proteína por separado haciendo coincidir la secuencia de aminoácidos entre cada

una de ellas. Esta representación de los resultados permite hacer comparaciones entre monómeros. Para facilitar el análisis se colocó en la parte superior de cada figura una representación gráfica que indica los bloques de estructura secundaria hoja- $\beta$  o hélice  $\alpha$  en función de la secuencia en la TcTIM<sup>1</sup>.

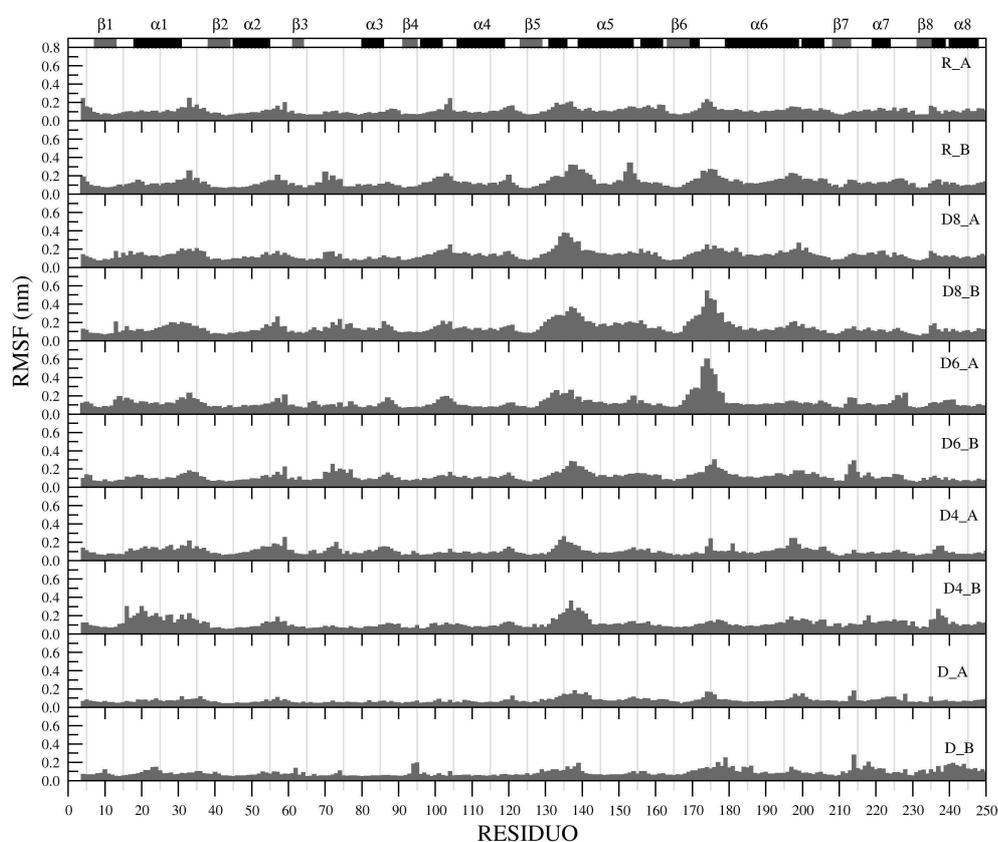
El perfil general que se observa en las gráficas de la Figura 4.7 es muy similar en todas las simulaciones así como para los diferentes monómeros. Las zonas con mayor movilidad en la proteína están representadas por los residuos con los valores más altos de RMSF y éstas coinciden entre sí en la mayoría de los casos. Era de esperarse que las zonas con residuos que forman hebras  $\beta$  presentaran los valores más bajos de RMSF ya que éstas, además de ser una estructura muy estable, se encuentran protegidas en la parte interna del barril, y limita su exposición directa al solvente. Al observar las gráficas, lo anterior resultó ser cierto en la mayor parte de los casos.

De la misma figura se puede concluir que, en términos generales, las regiones de la proteína cuya estructura secundaria corresponde a hebras  $\beta$  son las más rígidas, siguiendo de mayor a menor rigidez las hélices  $\alpha$  y finalmente los lazos entre bloques. Se puede observar que los lazos en general coinciden con los valores más altos de las distribuciones. La diferencia en los valores de RMSF entre las distintas trayectorias para un mismo aminoácido, pueden ser debidas a los diferentes medios en los que la proteína fue simulada o simplemente a que los tiempos de simulación son todavía cortos para explorar con suficiente intensidad todo el espacio de configuraciones. No así, el perfil general en todos los casos es comparable, aunque se observa que el perfil obtenido para la trayectoria D se aleja más del resto.

Como ya se ha mencionado, en todas las trayectorias analizadas los valores más altos de *RMSF* fueron obtenidos en residuos pertenecientes a lazos que unen bloques de hélices  $\alpha$  con hebras  $\beta$  o viceversa. En particular, los lazos más flexibles en cada monómero resultan ser los que unen  $\alpha_1$ - $\beta_2$ (residuos 32-37),  $\alpha_2$ - $\beta_3$  (residuos 56-60),  $\beta_4$ - $\alpha_4$  (residuos

---

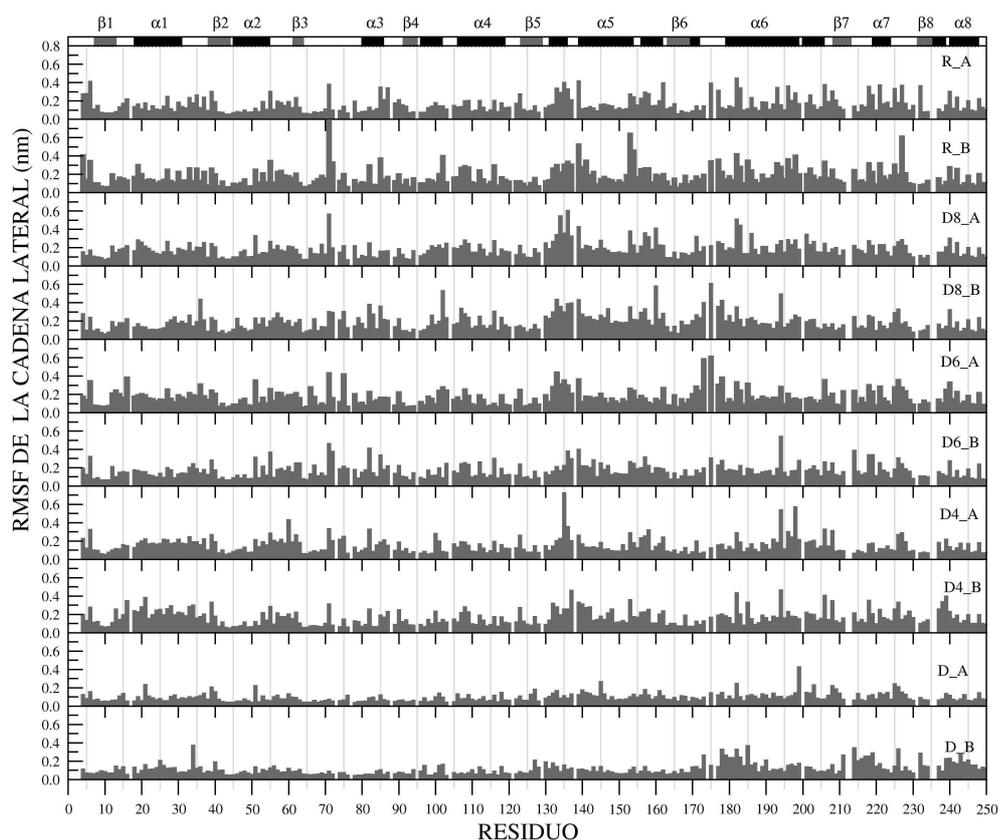
<sup>1</sup>Esta herramienta será llamada a partir de ahora GLES *Representación Gráfica Lineal de Estructura Secundaria*



**Figura 4.7:** Fluctuaciones cuadráticas medias (RMSF) por residuo de la TcTIM calculados sobre los últimos 30 ns de cada simulación. Las letras A y B representan a cada monómero de la proteína. Las notaciones R, D8, D6, D4 y D corresponden a cada una de las trayectorias realizadas. En la parte superior de la gráfica se indica el tipo de estructura secundaria correspondiente a la secuencia de aminoácidos de la TcTIM.

95-106),  $\beta_5$ - $\alpha_5$  (residuos 129-139) y  $\beta_6$ - $\alpha_6$  (residuos 169-179). Los residuos 129-139 que enlazan a  $\beta_5$ - $\alpha_5$  y los residuos 169-179 que enlazan a  $\beta_6$ - $\alpha_6$  son un claro ejemplo de regiones de mayor flexibilidad de la proteína en todas las trayectorias y en cada monómero de la TcTIM. Al comparar estas dos últimas regiones en la secuencia de la TcTIM se observa que en las trayectorias D8 y D6 presentan en todas las trayectorias los valores de RMSF más altos en ambos monómeros. En la trayectoria D8 del monómero B destacan particularmente los residuos Gly-174 y Tyr-175 del resto de residuos con valores de

RMSF de 0.5417 nm y 0.4560 nm, respectivamente. En la trayectoria *D6* los residuos del monómero A con los valores de RMSF más altos fueron los siguientes: Ile-173, Gly-174 y Tyr-175 con los valores 0.5111 nm, 0.5980 nm y 0.4903 nm, respectivamente.



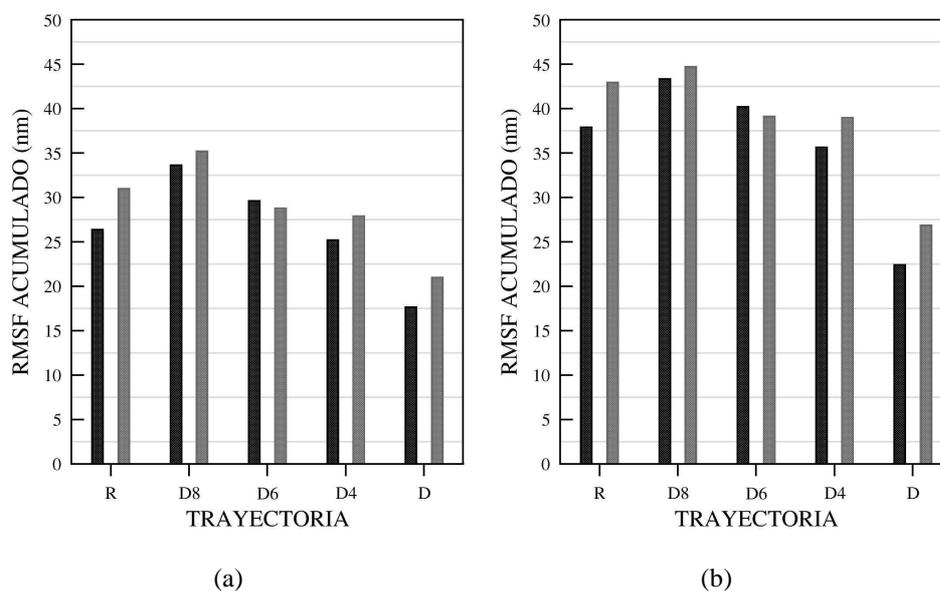
**Figura 4.8:** Fluctuaciones cuadráticas medias (RMSF) de las cadenas laterales de cada residuo por monómero de la TcTIM, calculados sobre los últimos 30 ns de simulación. Las letras A y B representan a cada monómero de la proteína. Las notaciones R, D8, D6, D4 y D corresponden a cada una de las trayectorias realizadas. En la parte superior de la gráfica se indica el tipo de estructura secundaria correspondiente a la secuencia de aminoácidos de la TcTIM.

En la figura 4.8 se muestra una gráfica donde se comparan los valores de *RMSF* de los átomos pertenecientes exclusivamente a las cadenas laterales. Aquellos residuos donde el valor de *RMSF* es cero corresponden a residuos de tipo glicina, cuya cadena lateral está

formada sólo por un átomo de H, no considerado de forma explícita en el campo de fuerzas utilizado. De la misma manera que en el caso anterior, los resultados se presentan para las subunidades A y B de la proteína de forma separada. Para el cálculo, al igual que en la Figura 4.7, se consideraron únicamente los últimos 30 ns de cada trayectoria. Los valores de RMSF que se obtienen después de excluir a los átomos que forman la estructura central son de manera general más altos que los encontrados al considerar todos los átomos que conforman a la enzima. Los resultados de las diferentes trayectorias que se presentan en la Figura 4.8 sugieren, al igual que en la Figura 4.7, un perfil muy similar entre sí, con la trayectoria D como excepción ya que muestra valores de RMSF de menor magnitud.

En la Figura 4.9 se presentan los valores de RMSF acumulados para cada una de las gráficas de la Figura 4.7 y de la Figura 4.8, es decir, la suma de los valores de RMSF sobre los residuos de cada monómero en las diferentes trayectorias. Al comparar el valor acumulado para cada monómero en las simulaciones no es posible afirmar que existe una diferencia significativa entre las trayectorias R, D8, D6 y D4. En cambio, si se compara el valor de RMSF acumulado para la trayectoria D, se encuentra que el resultado sí es significativamente menor que cuando el solvente es acuoso o una mezcla agua-decano. La rigidez que presenta la proteína al estar simulada en un medio no polar es observable con mayor claridad en el valor acumulado de RMSF calculado a partir de los resultados de la Figura 4.8, donde sólo se consideran las fluctuaciones de las cadenas laterales de los aminoácidos. En la figura 4.9(b) se observa que el valor de RMSF acumulado en la trayectoria D es significativamente menor que el calculado en las simulaciones con agua como parte del disolvente.

A partir de estos resultados no es posible afirmar que hay una diferencia significativa entre la movilidad de los dos monómeros de una misma proteína, aunque cabe hacer notar que en cuatro de las cinco trayectorias la RMSF acumulada fue ligeramente mayor para el monómero B. Los resultados de la RMSF de toda la proteína y la RMSF calculada para las cadenas laterales de los residuos presentan la misma tendencia.

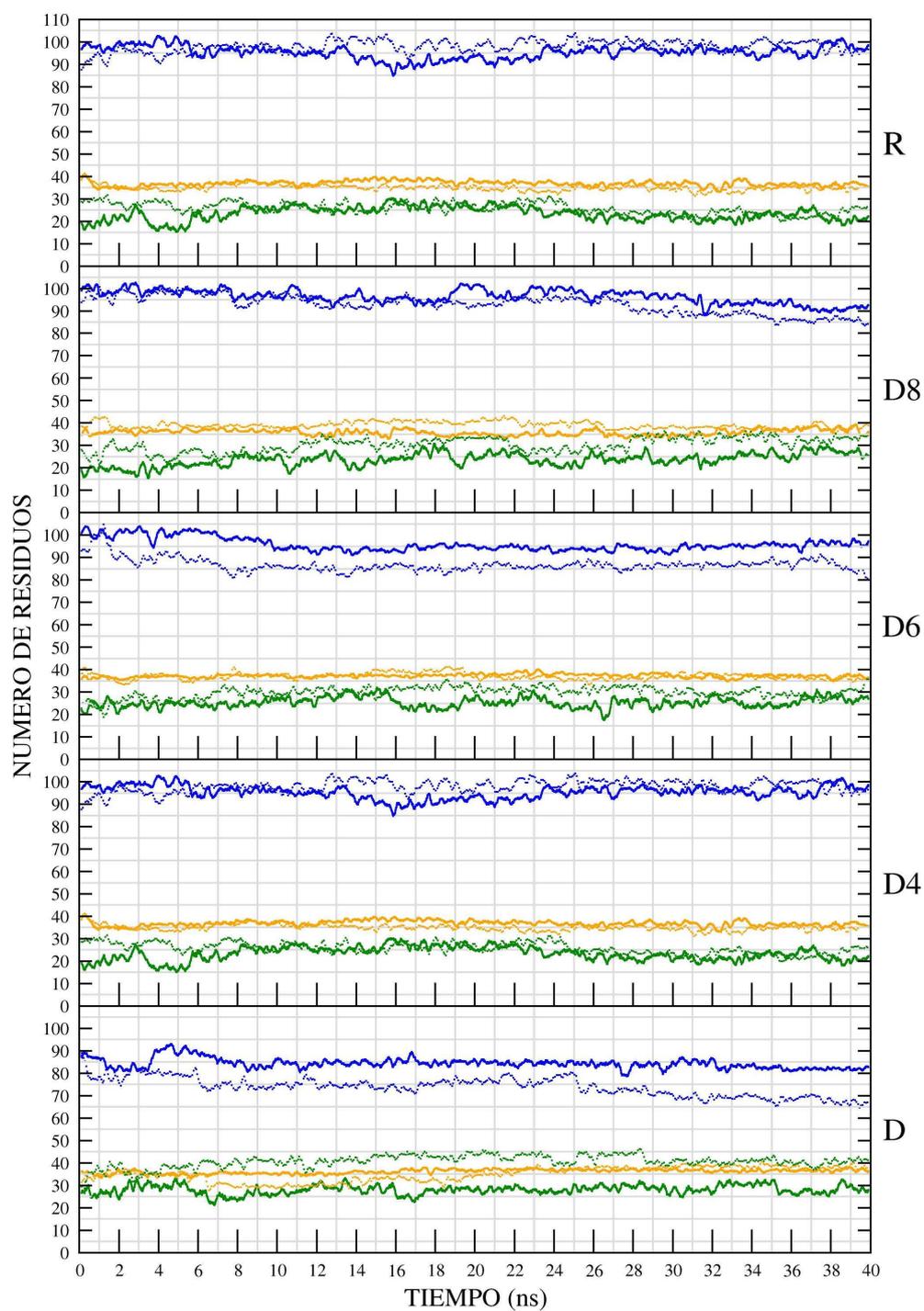


**Figura 4.9:** Fluctuaciones cuadráticas medias acumuladas para cada mómero de la TcTIM en las diferentes trayectorias. (a): RMSF acumulados obtenidos a partir de los valores de la Figura 4.7; (b): RMSF acumulados obtenidos a partir de los valores de la Figura 4.8. Las notaciones R, D8, D6, D4 y D corresponden a cada una de las trayectorias. Las subunidades A y B se representan en color negro y gris, respectivamente.

### 4.2.3 Estructura secundaria

#### Estructura secundaria de la TcTIM

Este análisis consistió en calcular el número de residuos que forman hélices  $\alpha$ , hebras  $\beta$  y giros  $\beta$  en cada conformación de la simulación. En la Figura 4.10 se presentan los resultados obtenidos en cada una de las simulaciones (R, D8, D6, D4 y D). El criterio utilizado para definir el tipo de estructura secundaria es el utilizado en el Diccionario de Estructura Secundaria para proteínas, (DSSP, Dictionary of Protein Secondary Structure) [47].



**Figura 4.10:** Estructura secundaria de la TcTIM según el código DSSP. Curva azul: hélices  $\alpha$ ; curva naranja: hebras  $\beta$ ; curva verde: giros. Las líneas continuas representan al monómero A y las líneas en color más tenue al monómero B.

Al comparar los resultados de cada sistema presentados en la Figura 4.10 se puede ver que R, D8, D6 y D4 presentan comportamientos similares. En las gráficas correspondientes a dichas trayectorias se observa que la estructura secundaria que más contribuye a la estructura de la proteína es la presencia de hélices  $\alpha$ , seguido de los residuos que forman hebras  $\beta$  y en último lugar se encuentran los residuos que forman giros. A diferencia de los otros sistemas en la trayectoria D la cantidad de residuos que forman giros rebasa a los que forman hojas- $\beta$ . Respecto a las hélices  $\alpha$  en esta última trayectoria a pesar de ser la estructura secundaria en mayor proporción, sus residuos nunca alcanzan las 100 unidades como en las trayectorias R, D8, D6 y D4.

En todas las simulaciones, el número de residuos formando hebras  $\beta$  fue de aproximadamente 35 unidades; esta estructura fue la que presentó menos fluctuaciones en las trayectorias, lo cual indica que el barril central formado por hebras  $\beta$  es la estructura más estable.

En cada trayectoria independiente se pueden observar fluctuaciones en los resultados para un determinado tipo de estructura, mismas que pueden interpretarse como transiciones reversibles en la estructura secundaria de la proteína. Por ejemplo, a los 12 ns el número de hélices  $\alpha$  en R parece disminuir en unas 10 unidades que se recuperan volviendo a su valor inicial alrededor de los 20 ns. Estas fluctuaciones concuerdan con el argumento de que la proteína es muy flexible cuando se encuentra en disolución acuosa, dando pie a la formación y destrucción reversible de estructura a nivel local. Los tipos de estructura secundaria con mayores fluctuaciones durante la simulación corresponden a las hélices  $\alpha$  y a los giros. En algunas ocasiones se observan diferencias entre las hélices  $\alpha$  y los giros de cada monómero en una misma trayectoria, por ejemplo, en las simulaciones D6 y D a los 40 ns de simulación el número de residuos que forman hélices  $\alpha$  en el monómero A es aproximadamente 10 unidades mayor que en el monómero B.

En la trayectoria D se observa una diferencia mayor en la estructura secundaria de ambos monómeros. Al término de la simulación el monómero A cuenta aproximadamente con 85 residuos formando hélices  $\alpha$ , unos 15 residuos más que el monómero B. Cuando se

analizan los residuos que forman giros, se observa una diferencia de la misma magnitud a la anterior pero que favorece al monómero B. A partir de esta diferencia, se puede relacionar la pérdida de hélices  $\alpha$  con el aumento en la cantidad de residuos formando giros que perdieron cierto grado de estructura a lo largo de la simulación.

Debido a los cambios conformacionales que se obtienen de este análisis y al diferente comportamiento dinámico que presenta la enzima en las diferentes trayectorias, se elaboró un análisis individual de los principales bloques de hélices  $\alpha$  y hebras  $\beta$  que conforman el barril.

### **Desviación cuadrática por bloque de estructura secundaria.**

Este análisis consiste en calcular la desviación cuadrática media a lo largo de cada trayectoria, de cada una de las hélices  $\alpha$  y hebras  $\beta$  que forman el barril  $\beta/\alpha$  en cada monómero. Los resultados se presentan de manera separada para las hélices  $\alpha$  y para las hebras- $\beta$ , en donde la RMSD fue calculada para todos los átomos que conforman cada bloque de estructura. Este cálculo permite identificar los bloques de estructura que mantienen menos similitudes con la estructura de referencia a lo largo de la simulación. Además permite hacer comparaciones entre monómeros de una misma proteína y, en su caso, encontrar diferencias entre las diferentes trayectorias. Los residuos que conforman cada bloque de estructura se especifican en la Figura 2.4. Como se explicó anteriormente, cada barril está formado por 8 hebras  $\beta$  y 8 hélices  $\alpha$ , además de los lazos que unen estos bloques de estructura.

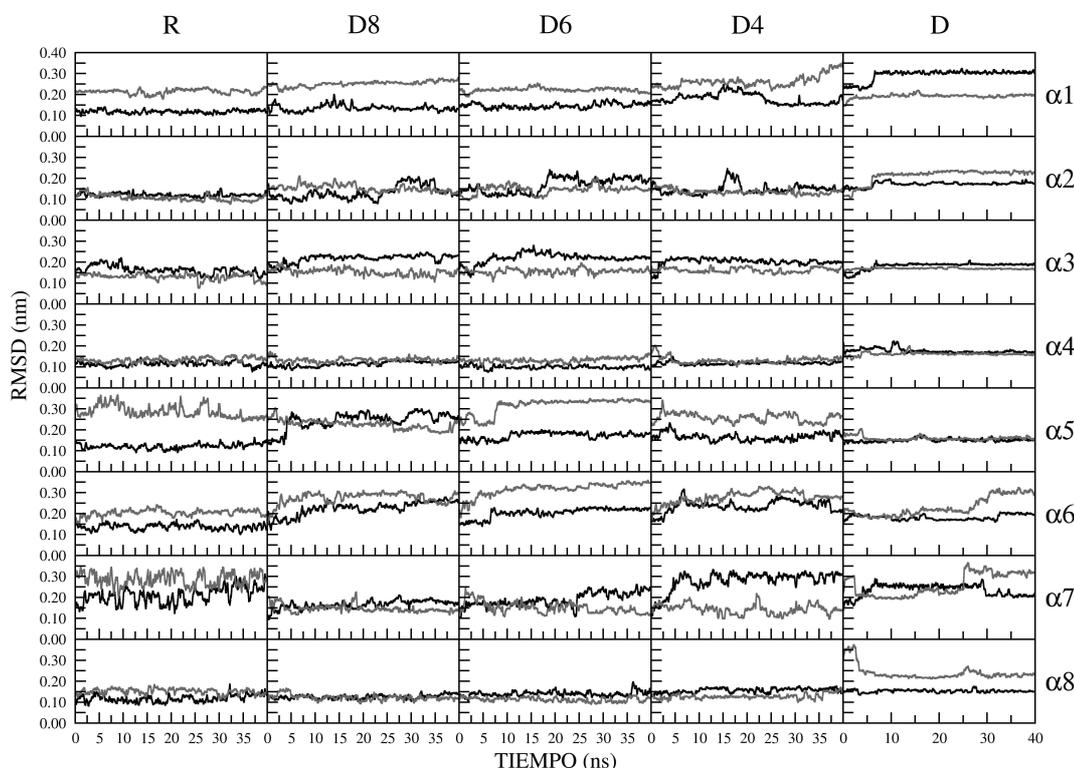
Los valores de RMSD calculados se presentan en las Figuras 4.11 y 4.12. Los resultados se muestran en forma de una matriz cuyo eje horizontal representa a cada una de las trayectorias y el eje vertical a los 8 bloques de residuos que forman la estructura secundaria de interés (hélices  $\alpha$  y hebras  $\beta$ ). Las hebras  $\beta$  y las hélices  $\alpha$  se enumeran con un subíndice del uno al ocho siguiendo la notación y orden al utilizado para describir la estructura cristalizada (ver Figura 2.4). En las figuras, las líneas que se encuentran de color negro, representan al monómero A y en color gris al monómero B. Para obtener la

RMSD, como en los casos anteriores, se utilizó también como referencia a la estructura cristalizada de la TcTIM [27].

En la Figura 4.11 se presentan los valores de RMSD calculados para las ocho hélices  $\alpha$  que pertenecen a un barril en la TcTIM. Al analizar estos resultados se observa que de manera general existe una similitud entre los resultados de las simulaciones D8, D6, D4 y D con la estructura de referencia R. Al comparar los valores obtenidos para un mismo bloque de estructura secundaria entre monómeros de una misma trayectoria se observan diferencias significativas en algunos casos. Por ejemplo, la  $\alpha_1$  del monómero B presentó una desviación aproximadamente de 0.15 nm mayor que el monómero A en la mayoría de las trayectorias. Como cada monómero está compuesto por el mismo número y tipo de residuos, se esperaría un comportamiento similar en el cambio de estructura secundaria en ambos monómeros. Sin embargo, en la Figura 4.11 se observan comportamientos distintos. Las hélices  $\alpha$  que mayores diferencias entre monómeros presentan son las  $\alpha_1$ ,  $\alpha_5$ ,  $\alpha_6$  y  $\alpha_7$ . En cambio, otras hélices, como el bloque de residuos que forma la  $\alpha_4$ , presentaron un RMSD similar en los dos monómeros en todas las trayectorias, sin aumentos considerables.

En la Figura 4.12 se presenta una matriz de los valores de RMSD calculados para las hojas- $\beta$  de la TcTIM. En general, los resultados obtenidos en todas las trayectorias son similares. En algunos casos, las estructuras se alejan inicialmente de la referencia para alcanzar el equilibrio nuevamente sin alguna tendencia que sugiera la pérdida progresiva de estructura secundaria. En las gráficas se pueden observar en algunos casos diferencias entre monómeros de una misma trayectoria. Por ejemplo, la hebra- $\beta_2$  del monómero B presentó un valor mayor de RMSD a lo largo de todas las trayectorias. La hebra- $\beta_4$  y la hebra- $\beta_5$  fueron los bloques de hebras  $\beta$  que presentaron menos diferencias entre monómeros, lo mismo se puede afirmar si se realizan comparaciones entre los diferentes sistemas simulados.

De la gráfica se pueden observar algunas fluctuaciones en la línea que representa la RMSD en forma de elevación o joroba como la que se encuentra en la hebra- $\beta_6$  del mo-

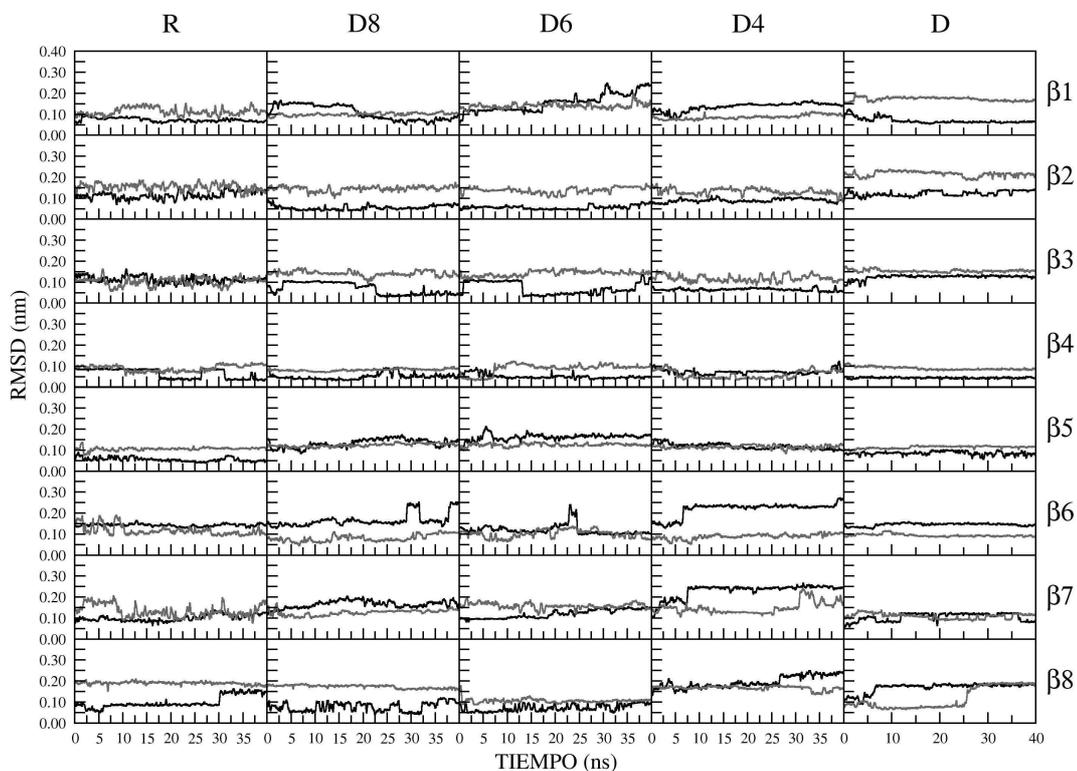


**Figura 4.11:** Matriz comparativa de RMSD para las hélices  $\alpha$  de cada barril  $\beta/\alpha$  en la TcTIM. Como referencia se utilizó la estructura cristalina de la TcTIM. Las notaciones R, D8, D6, D4 y D corresponden a cada una de las trayectorias realizadas. Las líneas que se encuentran de color negro representan al monómero A y en color gris al monómero B. Cada punto en las curvas representa un promedio sobre 200 ps.

número B de la trayectoria D8. Alrededor de los 30 ns la línea que representa la RMSD sugiere un alejamiento de la estructura de referencia que mantiene un equilibrio temporal por unos nanosegundos al cabo de los cuales regresa al valor de equilibrio inicial. Este cambio ocurre de manera repentina en varias de las hebras  $\beta$  que se estudiaron. En algunos casos la RMSD regresa a su valor original, en otros la estructura adquiere una nueva conformación de equilibrio tal como en la hebra- $\beta_6$  del monómero B de la trayectoria D4 y hebra- $\beta_7$  del monómero B de la trayectoria D4.

En las Figuras 4.11 y 4.12 se puede notar que los valores de RMSD que se obtienen en

la trayectoria D son los que menores fluctuaciones presentan. Por el contrario, las líneas que presentaron mayores fluctuaciones fueron obtenidas en la trayectoria R.



**Figura 4.12:** Matriz comparativa de RMSD para las hebras  $\beta$  de cada barril  $\beta/\alpha$  en la TcTIM. Como referencia se utilizó la estructura cristalizada de la TcTIM. Las letras A y B representan a cada subunidad de la TcTIM. Las notaciones R, D8, D6, D4 y D corresponden a cada una de las trayectorias realizadas. Las líneas que se encuentran de color negro representan al monómero A y en color gris al monómero B. Cada punto de las curvas representa un promedio sobre 200 ps.

### Estabilidad de las hélices $\alpha$ de la TcTIM

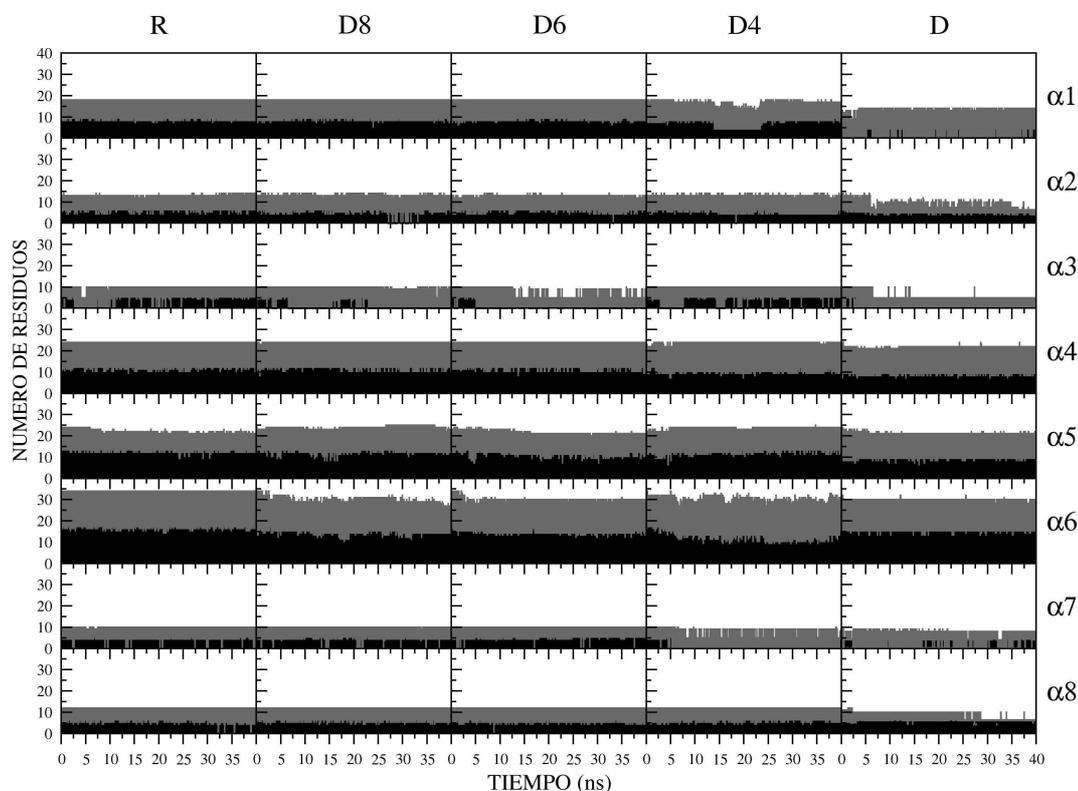
En la La Figura 4.13 se presenta un análisis de la estructura secundaria de los residuos que en la estructura cristalina de la TcTIM componen las ocho hélices  $\alpha$  principales de los barriles  $\beta/\alpha$ . En cada paso de tiempo de la trayectoria, se evaluó el número de residuos

que se encuentran formando parte de estas ocho hélices  $\alpha$  principales en la enzima. Los resultados que se obtienen son congruentes con los resultados de RMSD presentados en la Figura 4.11. En la Figura 4.13 destacan los resultados obtenidos para las hélices  $\alpha_1$ ,  $\alpha_3$  y  $\alpha_7$ , en ella se observa que el número de residuos que conforman a dichas hélices  $\alpha$  en el monómero A disminuye considerablemente durante la simulación, e incluso se hace cero. Estos resultados se aprecian mejor en una representación gráfica en donde se compara la estructura plegada original y las conformaciones obtenidas de las simulaciones. En la Figura 4.14 se muestra una representación de la conformación final de la hélice  $\alpha_1$  del monómero A en cada trayectoria y como referencia se presenta la hélice  $\alpha_1$  del mismo monómero extraída de la estructura cristalina (1TCD.pdb) 4.14(a). En la Figura 4.14(f) está representada la hélice- $\alpha_1$  a los 40 ns de la trayectoria D, en ella se puede apreciar claramente la pérdida de estructura secundaria. Otro comportamiento importante es el que corresponde a la hélice- $\alpha_1$  en la trayectoria D4. En la gráfica correspondiente se observa que, después de los 10 ns de simulación, el número de residuos que forman hélices  $\alpha$  disminuye en ambos monómeros. Este nuevo valor se mantiene constante aproximadamente por 10 ns de la trayectoria y después de este tiempo el número de aminoácidos en la estructura se recupera casi a su valor inicial.

### 4.3 Interacción de la TcTIM con el disolvente

En las Figuras 4.15 y 4.17 se muestra el número de moléculas de agua y el número de átomos de carbono pertenecientes al decano que están a menos de 4 Å, medido desde cualquier átomo de cada residuo para cada uno de los monómeros. El cálculo fue realizado sobre la última conformación de cada trayectoria.

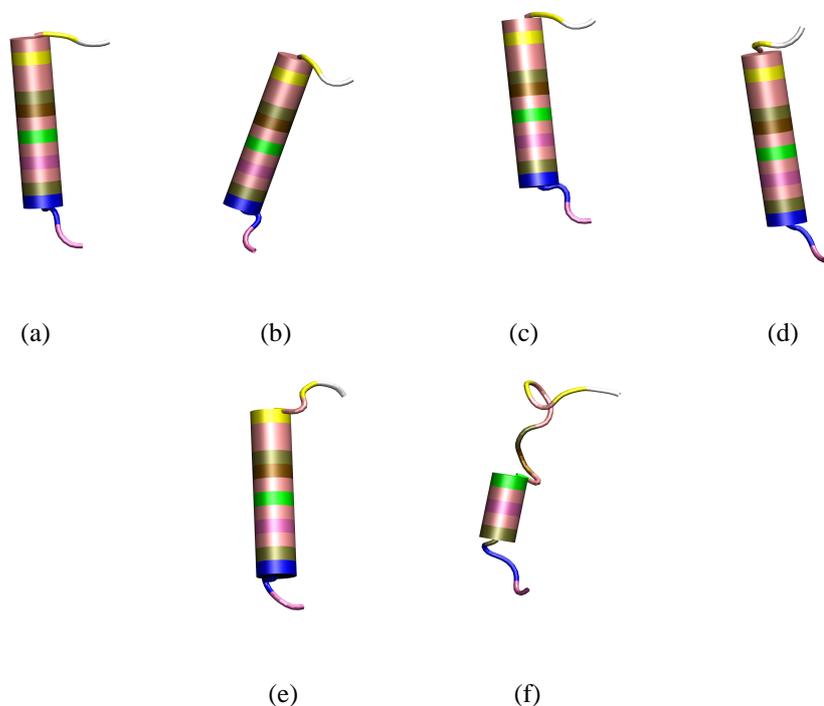
Este análisis es útil para identificar a los residuos de la proteína que interaccionan directamente con el disolvente aun cuando pertenezcan a una región interna de la proteína. Además es posible identificar la afinidad de algunos residuos por el disolvente hidrofóbico o acuoso.



**Figura 4.13:** Número de residuos que se encuentran formando hélices  $\alpha$  en función del tiempo para cada uno de los 8 segmentos helicoidales que forman el barril  $\beta/\alpha$  de la TcTIM. Las notaciones R, D8, D6, D4 y D corresponden a cada una de las trayectorias realizadas. Las subunidades A y B se representan en barras de color negro y gris, respectivamente.

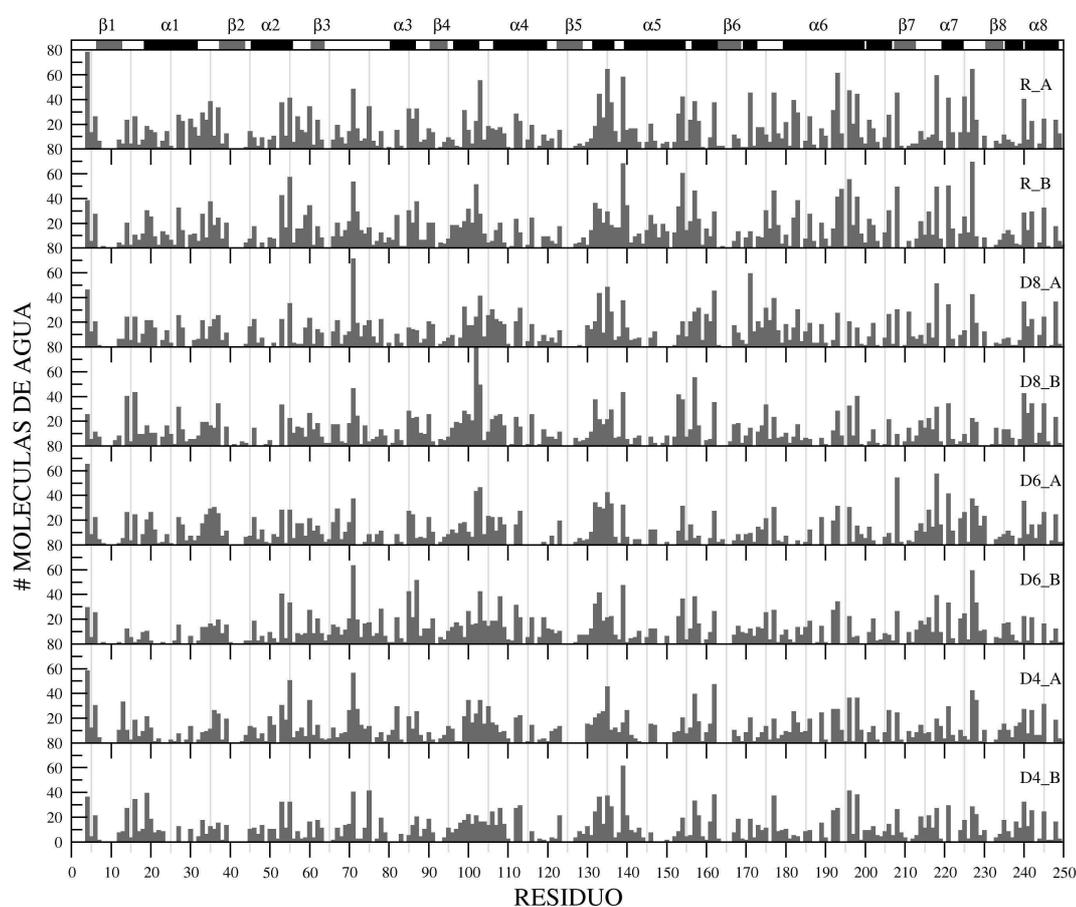
### 4.3.1 Agua

Para calcular el número de moléculas de agua que se presenta en la Figura 4.15, se elaboró un programa de análisis en código awk capaz de sumar el número de moléculas a menos de 4 Å de cada residuo. El programa se ejecutó sobre la última conformación del sistema en cada trayectoria. En la Figura 4.15 se puede observar que el perfil mostrado en las gráficas  $R_A$  y  $R_B$ , donde el único disolvente es el agua, se reproduce en las simulaciones D8, D6 y D4. En todas las gráficas se observa que en la mayoría de los casos los residuos que forman hebras  $\beta$  están en zonas con poco acceso al solvente.



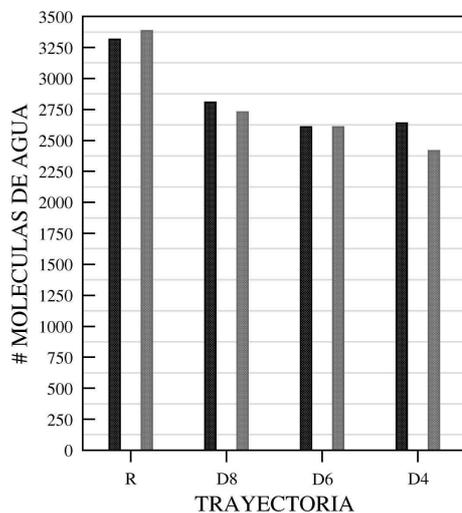
**Figura 4.14:** Representación de la hélice  $\alpha_1$  formada por los residuos 13-29 del monómero A de la TcTIM, en la última conformación de cada una de las trayectorias. (a): segmento  $\alpha_1$  de la estructura cristalina; (b): segmento  $\alpha_1$  de la trayectoria R; (c): segmento  $\alpha_1$  de la trayectoria D8; (d): segmento  $\alpha_1$  de la trayectoria D6; (e): segmento  $\alpha_1$  de la trayectoria D4; (f): segmento  $\alpha_1$  de la trayectoria D.

Los resultados que se presentan en la Figura 4.15 se muestran como una suma acumulada por monómero en cada trayectoria en la Figura 4.16. Al analizar la magnitud de estos resultados acumulados se observa que para algunos monómeros el número de aguas que se contabilizan para cada trayectoria es superior al número de moléculas de agua que fueron colocadas en cada caja de simulación. Esto es debido a que una misma molécula de agua puede estar simultáneamente a menos de 4 Å de más de un residuo de la TcTIM. Las barras más altas en la Figura 4.16 corresponden a la simulación con agua como único disolvente. En esta simulación, se encontraron aproximadamente unas 500 moléculas de



**Figura 4.15:** Número de moléculas de agua en el entorno de la TcTIM calculadas en la conformación final de cada trayectoria, a una distancia menor o igual a  $4 \text{ \AA}$  de cada átomo de la proteína. Las letras A y B representan a cada monómero de la TcTIM. Las notaciones R, D8, D6 y D4 corresponden a cada una de las trayectorias realizadas. En la parte superior de la gráfica se indica el tipo de estructura secundaria correspondiente a la secuencia de aminoácidos de la TcTIM en su estructura cristalina.

agua más que en el resto de las simulaciones. En las trayectorias D8, D6 y D4 el número de moléculas de agua encontrado es comparable en cualquiera de los casos.

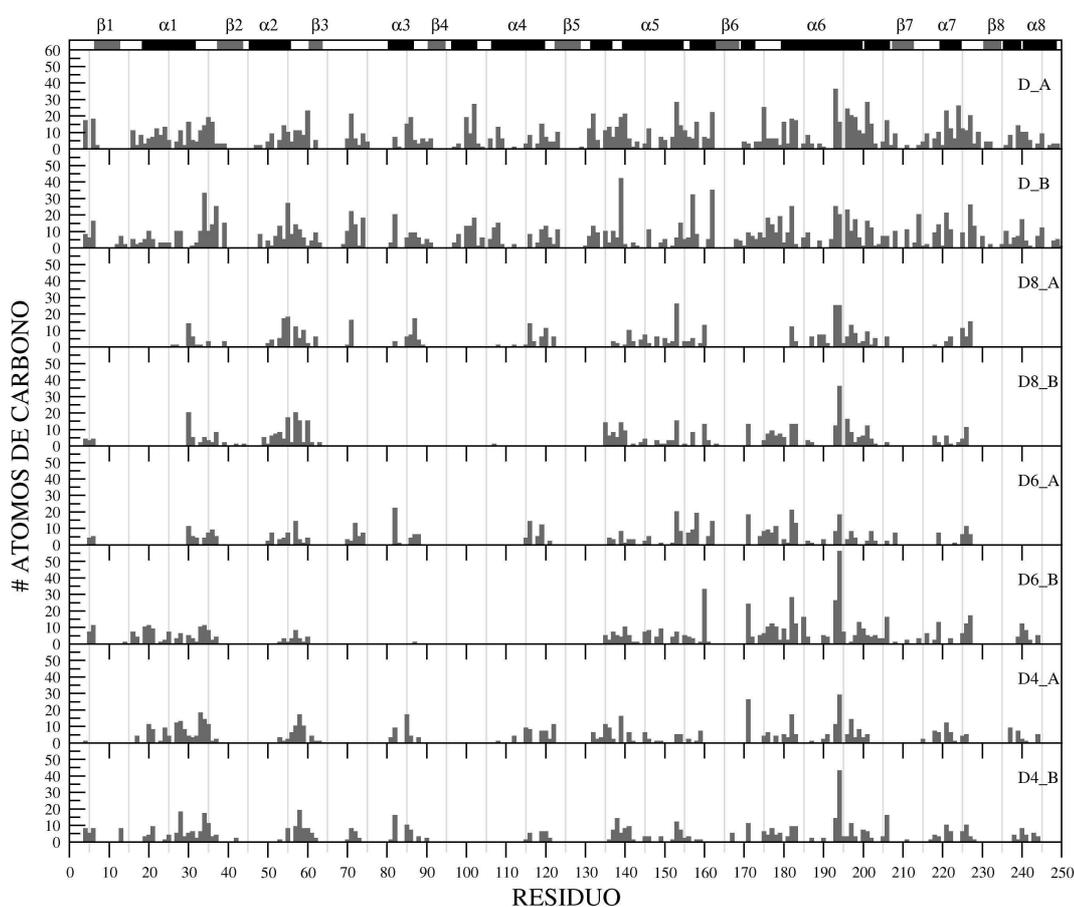


**Figura 4.16:** Número de moléculas de agua totales que se encuentran a una distancia menor a  $4 \text{ \AA}$  de cualquier átomo de la TcTIM, para cada monómero en cada una de las trayectorias. Calculado a partir de los resultados presentados en la Figura 4.15. Las notaciones R, D8, D6 y D4 corresponden a cada una de las trayectorias. La subunidad A y B se representan en color negro y gris, respectivamente.

### 4.3.2 Decano

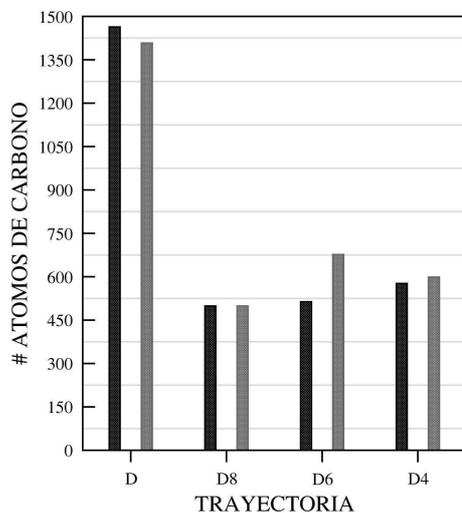
En la Figura 4.17 se presentan las gráficas que contabilizan al número de átomos de carbono a menos de  $4 \text{ \AA}$  de cada residuo en las simulaciones D8, D6, D4 y D. En las gráficas se observa que las moléculas de decano tuvieron mayor acceso a los residuos que pertenecen a la hélice- $\alpha_6$  en los dos monómeros y que el residuo hidrofóbico Trp-194 en todos los casos resulta ser el residuo más expuesto al solvente. Al cuantificar el número de átomos de carbono que estuvieron cerca de los residuos pertenecientes al sitio catalítico: Asn-12, Lys-14, His-96 y Glu-168 se observa que en todos los casos se encontraron en zonas de baja concentración de solvente apolar.

En la Figura 4.18 se presenta el número de átomos de carbono acumulados en cada trayectoria y para cada monómero. Los valores que se obtienen para las simulaciones D8, D6 y D4 son comparables entre sí; sin embargo, el valor acumulado para la trayectoria D es incluso del doble de tamaño que el resto de las trayectorias. En D el único disolvente es decano y por esta razón, su presencia en los alrededores de la proteína es mucho más



**Figura 4.17:** Número de átomos de carbono del disolvente decano en el entorno de la TcTIM calculadas en la última conformación de cada trayectoria, a una distancia menor o igual a 4 Å de cada átomo de la proteína. Las letras A y B representan a cada monómero de la TcTIM. Las notaciones D8, D6, D4 y D corresponden a cada una de las trayectorias realizadas. En la parte superior de la gráfica se indica el tipo de estructura secundaria correspondiente a la secuencia de aminoácidos de la TcTIM en su estructura cristalina.

abundante. En las simulaciones D8, D6 y D4 el decano se ve impedido para acercarse a los residuos de la proteína por la capa de agua que se colocó inicialmente alrededor de ésta y que en muchos de los casos se mantuvo en esa posición.



**Figura 4.18:** Número de átomos de carbono del disolvente totales que se encuentran a una distancia menor a 4 Å de cada residuo de la TcTIM, para cada monómero en cada una de las trayectorias. Calculado a partir de los resultados presentados en la Figura 4.17. Las notaciones D8, D6, D4 y D corresponden a cada una de las trayectorias. La subunidad A y B se representan en color negro y gris, respectivamente.

#### 4.4 Dinámica de las moléculas de agua en el sistema

El comportamiento del disolvente alrededor de la proteína es muy importante porque de ello depende la función de la misma. En el análisis que a continuación se presenta se contabiliza la frecuencia de incidencia de una molécula de agua en determinado residuo de la proteína a lo largo de las trayectorias. Esto permite identificar de manera específica residuos de la proteína que tienen cierta tendencia a estar hidratados con moléculas de agua en mayor proporción que el resto de los residuos. Este análisis complementa la búsqueda de zonas hidrofóbicas que se realizó en el análisis anterior.

Para la realización de este cálculo se elaboró un programa en código awk, descrito en la sección 3.5.6. Como frecuencia se entiende el número de veces que, sobre un total de 2000 conformaciones totales (que corresponden a 20 ns de simulación), una molécula de agua se encontró dentro de un círculo con radio menor a cierta distancia. Las frecuencias con las que se encontraron moléculas de agua en un radio menor a 2 Å de cualquier átomo para cada residuo de la proteína se presentan en las Figuras 4.19 y 4.20. En las Figuras 4.22 y 4.23 se presentan los resultados correspondientes a un radio de corte de 4 Å. Cabe

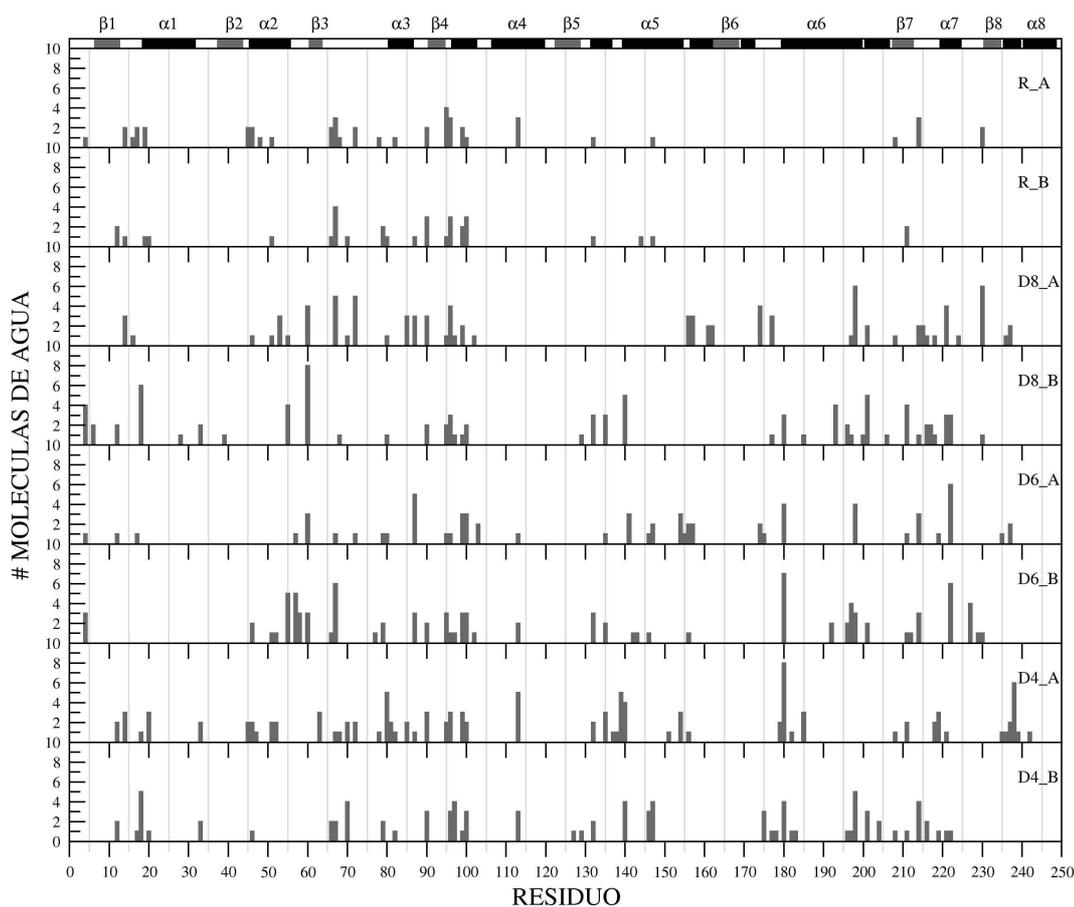
recalcar que el análisis que aquí se presenta se realizó sobre una trayectoria compuesta por 2000 conformaciones diferentes, a diferencia del análisis que se presenta en 4.3 el cual se hizo sobre una conformación fija. Aquí las frecuencias con las que aparece cada molécula de agua en el entorno del mismo residuo se traducen en tiempos. Si una misma molécula se encontró 200 veces a  $2 \text{ \AA}$  del mismo residuo, y considerando que la diferencia de tiempos entre 2 conformaciones consecutivas es de 10 ps, se puede deducir que esa molécula estuvo ligada al residuo durante 2 ns.

Las Figuras 4.19 y 4.20 presentan los resultados con una frecuencia de incidencia entre 2 y 4 ns, los resultados de las Figuras 4.22 y 4.23 corresponden a una frecuencia de incidencia entre 4 y 20 ns.

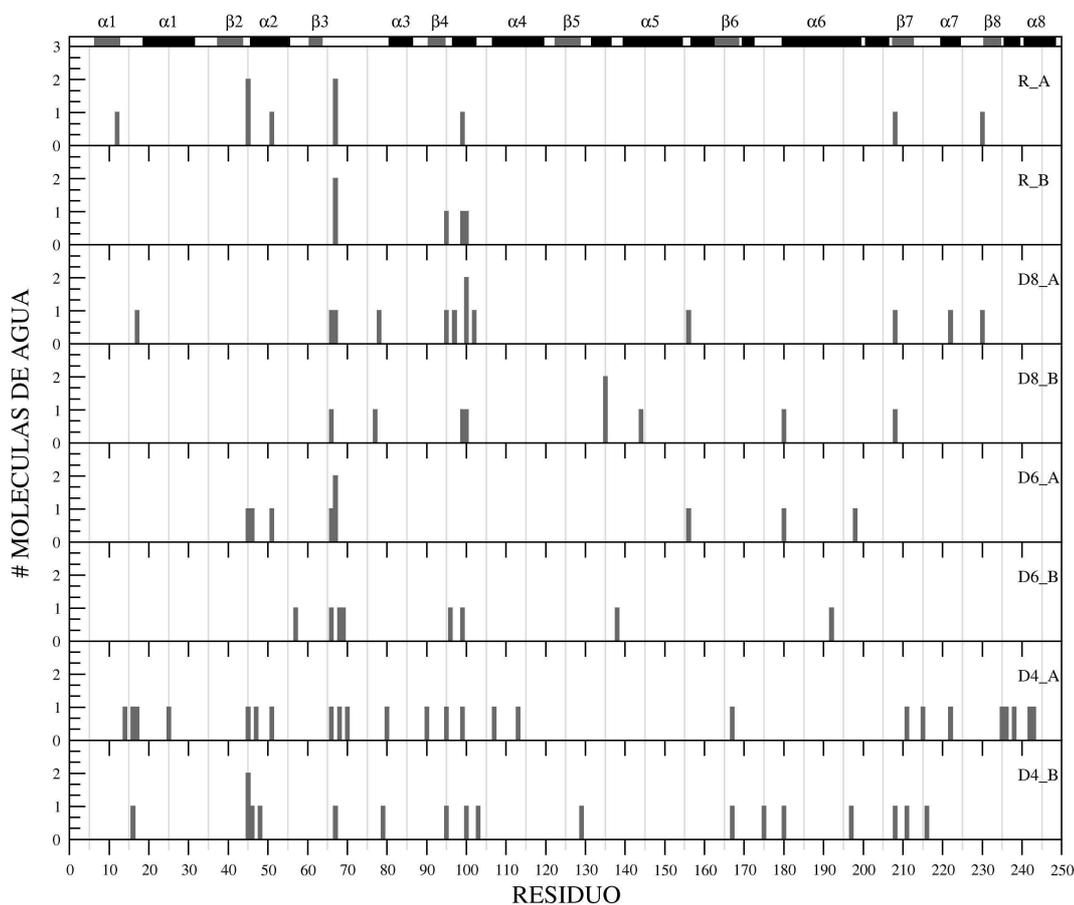
Si se observa la Figura 4.19, se puede notar que las barras obtenidas para  $R_A$  y  $R_B$  son muy pequeñas comparadas con el resto de las trayectorias (nótese que la escala en todas las gráficas es la misma). Este comportamiento se observa en todas las figuras que se presentan para este análisis.

Al comparar el número de moléculas acumuladas en cada gráfica de frecuencias (Figuras 4.21 y 4.24), es fácil notar una gran diferencia entre el número total obtenido en las simulaciones donde el solvente es agua pura y en las simulaciones con mezclas agua-decano. La interpretación de este fenómeno puede ir en el sentido de que las moléculas de decano forman una capa que restringe el movimiento de las moléculas alrededor de la enzima. Dado que el decano no es polar, el agua demuestra una mayor afinidad a la superficie de la proteína que por el medio hidrofóbico que la rodea. En la Figura 4.21 se puede observar lo descrito anteriormente.

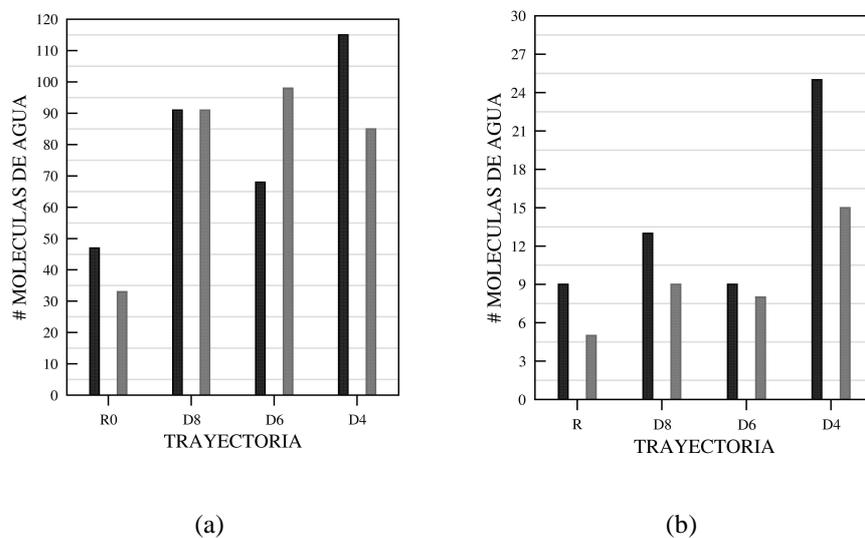
En este análisis, como el que se presenta en la sección 4.3, una misma molécula de agua puede estar cerca de dos residuos diferentes en el mismo paso de tiempo, aunque, debido a las restricciones de distancia y tiempo, la probabilidad de que esto suceda es mucho menor.



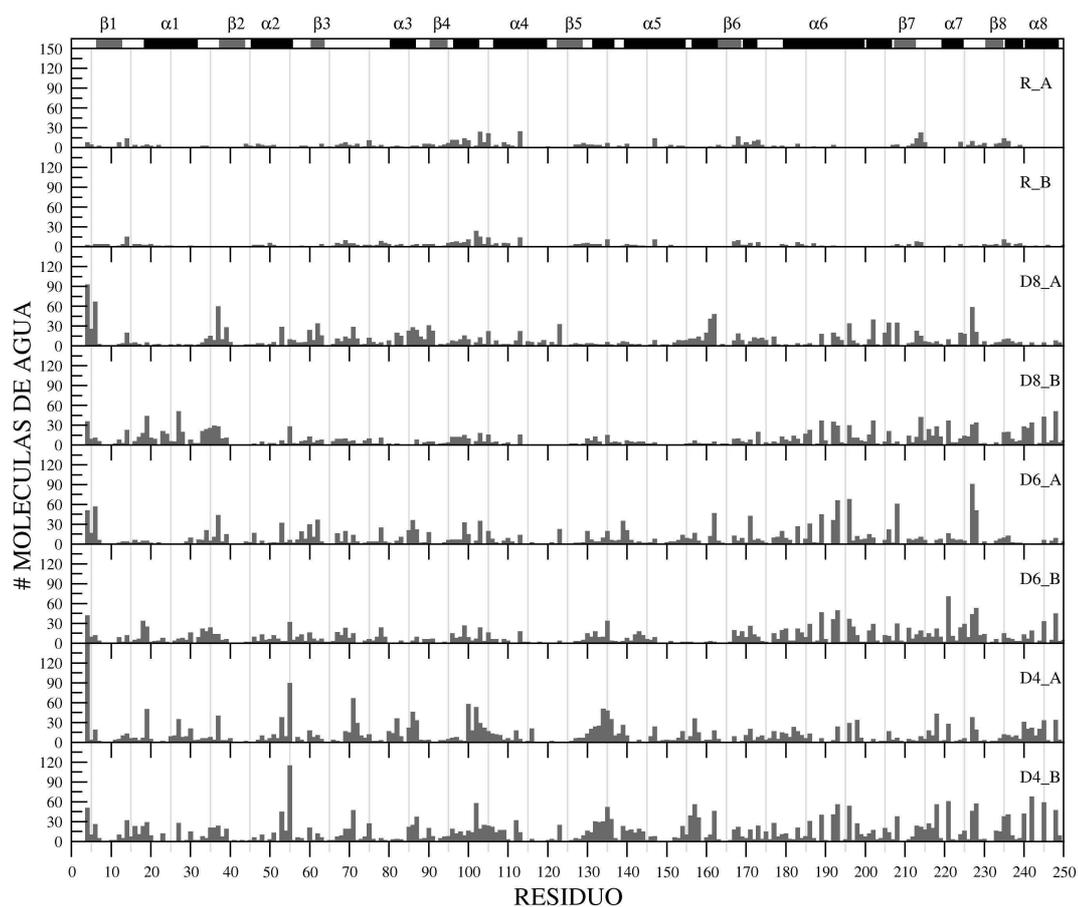
**Figura 4.19:** Incidencia de moléculas de agua a menos de 2 Å cada residuo de la TcTIM, con tiempos de permanencia entre 2 y 4 ns. El análisis se realizó sobre los primeros 20 ns de trayectoria. Las letras A y B representan a cada monómero de la TcTIM, las notaciones R, D8, D6 y D4 corresponden a cada una de las trayectorias realizadas. En la parte superior de la gráfica se indica el tipo de estructura secundaria correspondiente a la secuencia de aminoácidos de la TcTIM.



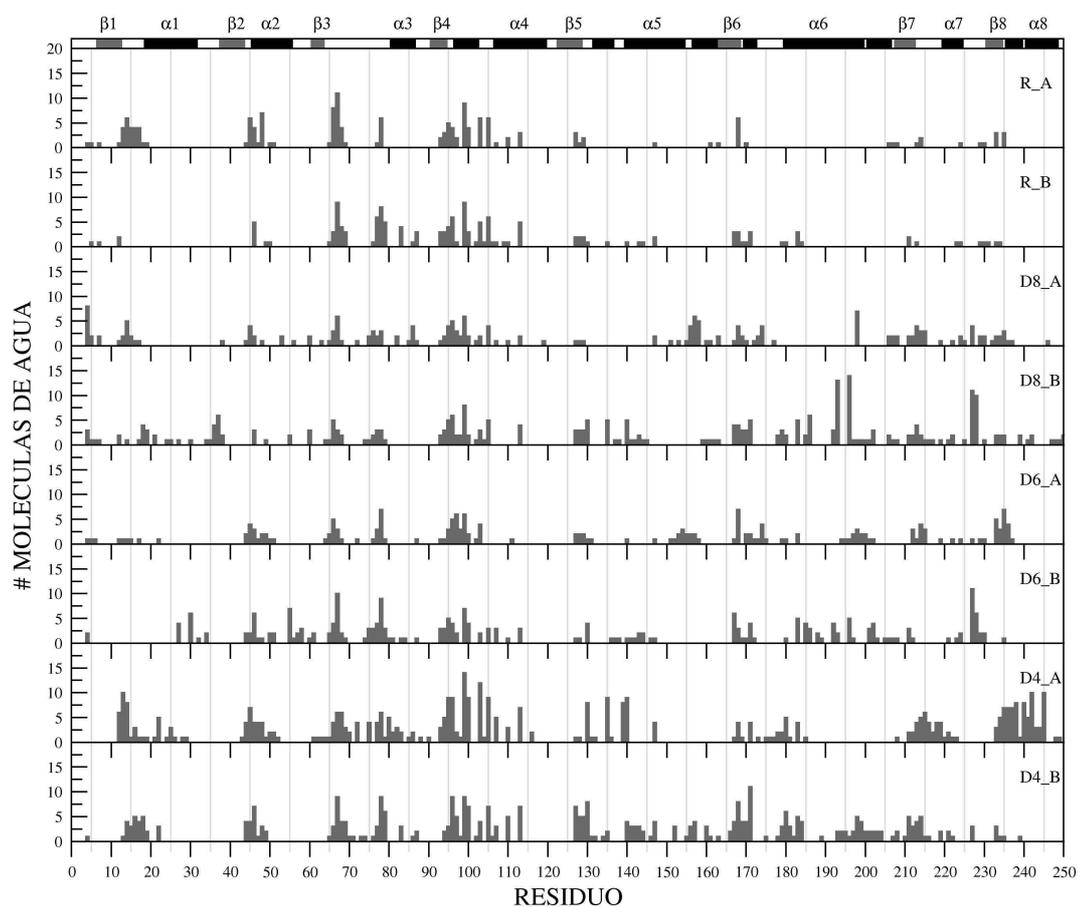
**Figura 4.20:** Incidencia de moléculas de agua a menos de 2 Å de cada residuo de la TcTIM, con tiempos de permanencia de entre 4 y 20 ns. El análisis se realizó sobre los primeros 20 ns de trayectoria. Las letras A y B representan a cada monómero de la TcTIM, las notaciones R, D8, D6 y D4 corresponden a cada una de las trayectorias realizadas. En la parte superior de la gráfica se indica el tipo de estructura secundaria correspondiente a la secuencia de aminoácidos de la TcTIM.



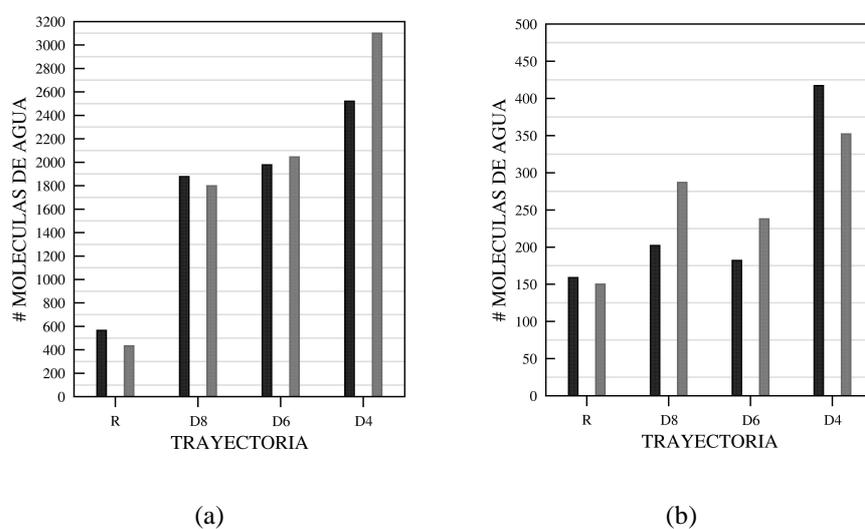
**Figura 4.21:** Frecuencias acumuladas para la incidencia de moléculas a menos de  $2 \text{ \AA}$  de la TcTIM. (a): Frecuencias acumuladas calculadas a partir de los resultados de la Figura 4.19 con una frecuencia entre 2 y 4 ns; (b): Frecuencias acumuladas calculadas a partir de los resultados de la Figura 4.20 con una frecuencia entre 4 y 20 ns. Las notaciones R, D8, D6 y D4 corresponden a cada una de las trayectorias. La subunidad A y B se representan en color negro y gris, respectivamente.



**Figura 4.22:** Incidencia de moléculas de agua a menos de 4 Å de cada residuo de la TcTIM, con tiempos de permanencia de entre 2 y 4 ns. El análisis se realizó sobre los primeros 20 ns de trayectoria. Las letras A y B representan a cada monómero de la TcTIM, las notaciones R, D8, D6 y D4 corresponden a cada una de las trayectorias realizadas. En la parte superior de la gráfica se indica el tipo de estructura secundaria correspondiente a la secuencia de aminoácidos de la TcTIM.



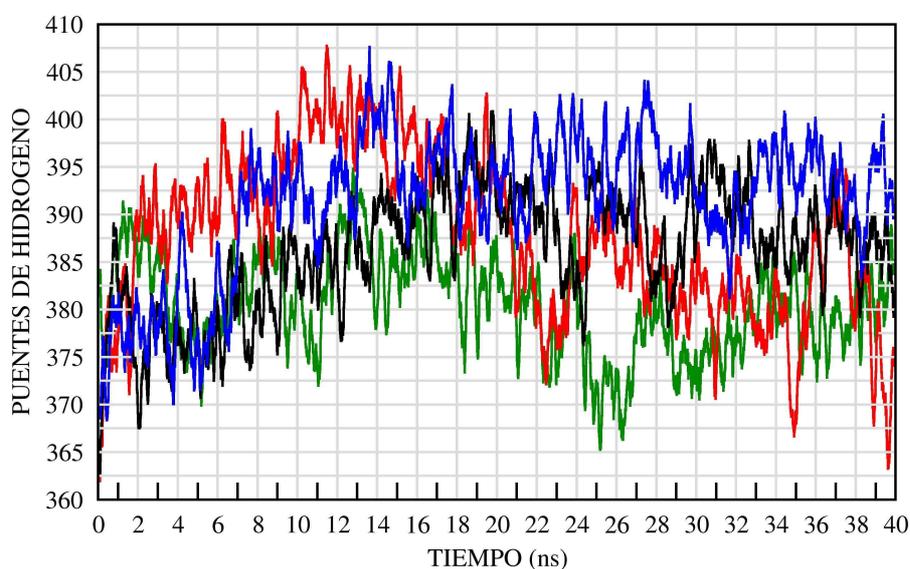
**Figura 4.23:** Incidencia de moléculas de agua a menos de 4 Å de la TcTIM, con tiempos de permanencia entre 4 y 20 ns. El análisis se realizó sobre los primeros 20 ns de trayectoria. Las letras A y B representan a cada monómero de la TcTIM, las notaciones R, D8, D6 y D4 corresponden a cada una de las trayectorias realizadas. En la parte superior de la gráfica se indica el tipo de estructura secundaria correspondiente a la secuencia de aminoácidos de la TcTIM.



**Figura 4.24:** Frecuencias acumuladas para la incidencia de moléculas a menos de  $4 \text{ \AA}$  de la TcTIM. (a): Frecuencias acumuladas calculadas a partir de los resultados de la Figura 4.22 con una frecuencia entre 2 y 4 ns; (b): Frecuencias acumuladas calculadas a partir de los resultados de la Figura 4.23 con una frecuencia entre 4 y 20 ns. Las notaciones R, D8, D6 y D4 corresponden a cada una de las trayectorias. La subunidad A y B se representan en color negro y gris, respectivamente.

## 4.5 Puentes de hidrógeno

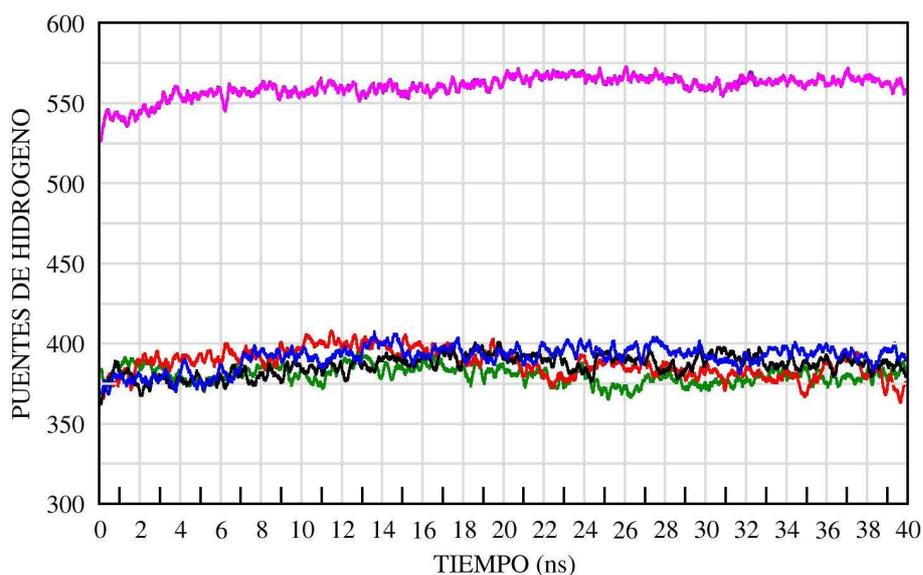
En la Figura 4.25 se presenta el número de puentes de hidrógeno intramoleculares de la proteína para las simulaciones R, D8, D6 y D4 en función del tiempo a lo largo de 40 ns de dinámica molecular. En la literatura se suelen explicar las diferencias en el comportamiento de enzimas en medios de diferente polaridad en función del número de puentes de H intramoleculares y del número de puentes de H que se forman entre la macromolécula y el propio disolvente [3].



**Figura 4.25:** Número de puentes de hidrógeno intramoleculares de la TcTIM a lo largo de las trayectorias R, D8, D6 y D4. Curva verde: Trayectoria R; curva roja: Trayectoria D8; curva negra: Trayectoria D6; curva azul: Trayectoria D4. Cada punto en las curvas representa un promedio sobre 200 ps.

En general, el número de puentes de H intramoleculares debe ser mayor para medios de baja polaridad. Como se puede observar, el número de puentes de H que se forman entre grupos de la propia proteína es similar en las cuatro simulaciones a lo largo de toda la trayectoria.

Cuando se compara el resultado del análisis de la simulación en decano (Figura 4.26), se observa una diferencia aproximadamente de 170 puentes de H. Esta diferencia se explica en función de la polaridad del disolvente, ya que el decano es un solvente apolar y en ese medio la formación de puentes de hidrógeno puede darse sólo con átomos de la propia proteína.



**Figura 4.26:** Número de puentes de hidrógeno intramoleculares de la TcTIM a lo largo de las trayectorias R, D8, D6, D4 y D. Curva magenta: simulación en decano puro; curva verde: Trayectoria R; curva roja: Trayectoria D8; curva negra: Trayectoria D6; curva azul: Trayectoria D4. Cada punto en las curvas representa un promedio sobre 200 ps.

## Capítulo 5

### Conclusiones

En este trabajo se estudió el efecto de la concentración de agua en el comportamiento dinámico y en la estabilidad estructural de la triosafosfato isomerasa de *Trypanosoma cruzi*, TcTIM, por medio de simulaciones de dinámica molecular. Se analizaron además la distribución y la rapidez de intercambio de las moléculas del propio disolvente en el entorno de la enzima en función del tiempo. Para ello se realizaron tres simulaciones de la TcTIM con diferentes concentraciones de decano y agua, una simulación con agua como único disolvente y otra más utilizando exclusivamente decano para solvatar a la enzima. De los resultados obtenidos se concluye que para mantener tanto la estructura como el comportamiento dinámico de la TcTIM es necesaria una mínima cantidad de moléculas de agua distribuidas de manera específica interactuando con determinados grupos de la enzima. Aunque las simulaciones no permiten extraer conclusiones directas acerca de la actividad catalítica de la proteína, a partir de nuestros resultados se puede sugerir, en términos de la relación flexibilidad-actividad-estabilidad explicada en el capítulo 1 de esta memoria, que la movilidad de la enzima en las simulaciones donde existe una mínima cantidad de agua es suficiente para que ésta realice su función biológica. Los resultados aquí obtenidos en términos de flexibilidad, número de puentes de hidrógeno tanto intramoleculares como con el disolvente y de movilidad del agua en el entorno de la proteína podrían extrapolarse para explicar a nivel atómico la capacidad de otras enzimas para realizar su función

en solventes orgánicos. La equivalencia del comportamiento dinámico de la proteína en las simulaciones que contienen pequeñas cantidades de agua respecto a la simulación en agua pura contribuye a validar el estudio sistemático de otras macromoléculas de interés biológico en las mismas condiciones, ahorrando hasta un 70 % de tiempo de cálculo en las interacciones de largo alcance. Por último, este trabajo proporciona información a nivel atómico que permite entender el comportamiento dinámico de la TcTIM en disolución en escalas de tiempo de varias decenas de ns, así como las interacciones con el disolvente, por lo que se espera que estos resultados sean útiles en el diseño de fármacos que inhiban la función de la enzima.

## Apéndice A

### Dinámica molecular

El conocimiento a nivel atómico de aspectos estructurales y dinámicos de sistemas organizados es un tema de gran interés, ya que contribuye a aumentar la comprensión de los sistemas moleculares. Una de las ramas de la Química que permite adentrarse en el comportamiento microscópico de los sistemas moleculares es la química computacional, que por medio de la aplicación de programas informáticos trata de investigar a nivel atómico la estructura, comportamiento dinámico y estabilidad del sistema molecular de interés en relación al medio y las condiciones en que se encuentra.

Actualmente existen distintos paquetes de programas que permiten realizar cálculos computacionales de moléculas a distintos niveles. Éstos pueden ser productos comerciales o bien de libre distribución. De acuerdo a su nivel de teoría, entre los que realizan cálculos *ab initio* destacan GAUSSIAN y GAMESS. A nivel clásico los más populares son CHARM, AMBER y GROMACS.

Para todas las simulaciones realizadas en este trabajo se utilizó la versión 3.2.1 del paquete GROMACS [38, 39, 40, 41] que, con diferencia, es el paquete más optimizado en cuanto a velocidad de cálculo en simulaciones de Dinámica Molecular (DM). Éste es un programa de libre distribución al cual se puede acceder de forma gratuita y, de ser necesario, permite consultar e incluso modificar su código.

En este apéndice se proporciona una descripción detallada de las características y

algoritmos utilizados por GROMACS para la realización de las simulaciones de dinámica molecular que se presentan en este trabajo.

## A.1 Simulaciones de dinámica molecular

Las simulaciones de dinámica molecular se basan en la resolución de las ecuaciones de Newton del movimiento. Un sistema de  $N$  átomos con masas  $m_i$  ( $i = 1, 2, \dots, N$ )  $i$ , definidas por los vectores de posición  $\mathbf{r}_i$  está descrito por:

$$m_i \frac{\partial^2 \mathbf{r}_i}{\partial t^2} = \mathbf{F}_i, \quad i = 1, \dots, N. \quad (\text{A.1})$$

La fuerza ejercida sobre una partícula dentro de un campo de potencial  $V$ , está dada por:

$$\mathbf{F}_i = -\nabla V \quad (\text{A.2})$$

Las ecuaciones (A.1) y (A.2) se resuelven simultáneamente en pequeños pasos de tiempo de forma iterativa para hacer evolucionar el sistema. El sistema se monitorea durante determinado intervalo de tiempo, cuidando que la temperatura y presión permanezcan dentro de los valores requeridos para reproducir las condiciones experimentales. Las coordenadas resultantes del movimiento de los átomos se almacenan en un archivo de salida a intervalos regulares de tiempo. La información contenida en este archivo se puede entender como una *trayectoria*. Si se realizan cálculos sobre los datos almacenados, es posible determinar las propiedades macroscópicas del sistema que proporciona nuestro modelo molecular.

### Nivel de cálculo

Un sistema molecular puede representarse como un sistema de masas puntuales que se mueven bajo un potencial efectivo que, generalmente, es un campo de potencial conservativo (es decir que sólo depende de las coordenadas atómicas instantáneas). La dinámica

de partículas puntuales es, en buen grado de aproximación, gobernada por las leyes de la mecánica clásica. Esta consideración conlleva varias limitaciones. Por ejemplo, el movimiento a bajas temperaturas (0-10 K) no es descrito de forma adecuada y no es posible simular la dinámica de átomos ligeros como el hidrógeno incluso a 278 K. Finalmente, la descripción de reacciones químicas queda fuera de los métodos de simulación basados en la mecánica clásica ya que no consideran intercambio de átomos entre moléculas.

Así, las simulaciones basadas en la mecánica clásica, constituyen una buena aproximación en la mayoría de sistemas en donde el movimiento de los átomos se efectúa a temperaturas normales. Cuando se estudia el movimiento de un sistema de átomos ligeros como el hidrógeno o el helio en determinadas condiciones, el cálculo obligado es esencialmente cuántico. El comportamiento de un oscilador armónico clásico, difiere apreciablemente de un oscilador cuántico cuando la frecuencia de resonancia  $\nu$  es igual o mayor a  $k_B T/h$  (en la Tabla A.1 se muestran los diferentes tipos de grados de libertad intramoleculares que aparecen en las simulaciones de moléculas típicas, así como el periodo de resonancia que los describe). A temperaturas cercanas a 298 K, el número de onda  $\nu = \frac{1}{\lambda} = \frac{\nu}{c}$ , en donde  $h\nu = k_B T$ , es aproximadamente  $200 \text{ cm}^{-1}$ . Por lo tanto, todas las frecuencias de resonancia que superen este valor no podrán ser descritas mediante mecánica molecular clásica.

A continuación se describen dos maneras de abordar este problema:

a) Si se hace una simulación de DM modelando los enlaces mediante osciladores armónicos, se debe corregir la energía total interna  $U = E_{kin} + E_{pot}$  y el calor específico  $C_v$  (también la entropía  $S$  y la energía libre,  $A$  o  $G$ , si son calculadas). Las correcciones a la energía y al calor específico de un oscilador unidimensional con frecuencia  $\nu$  son:

$$U^{QM} = U^{cl} + k_B T \left( \frac{1}{2}x - 1 + \frac{x}{\exp^x - 1} \right) \quad (\text{A.3})$$

$$C_v^{QM} = C_v^{cl} + k_B \left( \frac{x^2 \exp^x}{(\exp^x - 1)^2} - 1 \right) \quad (\text{A.4})$$

donde  $x = h\nu/kT$ . El oscilador clásico absorbe mucha energía  $kT$ , mientras el oscilador

cuántico de alta frecuencia está en su estado basal de energía ( $E = 1/2 h\nu$ ).

b) Se puede restringir el movimiento de los enlaces y ángulos de enlace en las ecuaciones del movimiento. Esto está basado en que un oscilador cuántico en su estado basal se parece más a un enlace rígido que a un oscilador clásico. Esta alternativa, tiene la gran ventaja de que nos permite utilizar pasos de tiempo más largos al eliminar los grados de libertad de mayor frecuencia. GROMACS utiliza esta última aproximación para los enlaces y ángulos de enlace.

**Tabla A.1:** *Periodos de oscilación de movimientos atómicos [48].*

Tipo	Periodo(fs)	
	Experimental	Teórico
Distancia de enlace, H	10	10
Distancia de enlace, átomos pesados	30	20
Agua, vibración	–	20
Agua, rotación	–	1300
Ángulo, H	32	20
Ángulo, átomos pesados	154	45
Ángulo, $-\text{NH}_3^+$ C – N – H	32	22
Ángulo, $-\text{NH}_3^+$ H – N – H	23	13
Ángulos diédros impropios	–	28
Ángulos diédros propios	–	27
Ángulos diédros, enlace peptídico	489	28
Ángulos diédros, $-\text{NH}_3^+$	489	89
Ángulos diédros, OH	53	43

### **Aproximaciones de los campos de fuerza**

En las simulaciones clásicas de DM se utiliza un campo de fuerza conservativo y las contribuciones al potencial se calcularon sólo en función de las posiciones atómicas. Esta aproximación es conocida como Aproximación de Born-Oppenheimer, la cual considera que los movimientos electrónico y nuclear están desacoplados. En ella se supone que los electrones ajustan su dinámica inmediatamente cuando la posición atómica cambia y permanecen en su estado basal. Esta aproximación impide estudiar procesos en los que hay transferencia de electrones o átomos en estados excitados. Cuando se excluyen reacciones químicas en el sistema en condiciones no extremas, es relativamente seguro asumir que el sistema está descrito por las leyes de la mecánica clásica.

Algunas de las características y limitaciones del campo de fuerza utilizado, se enumeran a continuación:

- i. El campo de fuerza es aditivo por pares en todos sus términos excepto en la contribución de las interacciones electrostáticas de largo alcance. Esto quiere decir que todas las fuerzas entre átomos no enlazados resultan de la suma de las interacciones entre pares. Como consecuencia, no es posible evaluar el efecto de la polarizabilidad de manera natural.
- ii. Es necesario utilizar un radio de corte para las interacciones de Lennard-Jones y de Coulomb. Debido a la convención de mínima imagen el radio de corte no puede exceder la mitad del tamaño de la caja de simulación.
- iii. Las condiciones de contorno o de frontera son artificiales. Como el tamaño de los sistemas es pequeño, una parte considerable de partículas en la frontera del sistema tendrá una superficie de interacción con el entorno, que puede ser vacío o algún otro medio más o menos realista. Para evitar este problema y poder simular una fase homogénea, se utilizan condiciones periódicas, es decir, se hacen copias del sistema en tres dimensiones espaciales. Esto resuelve en parte el problema. Además, se debe tener cuidado cuando se selecciona el tamaño de la caja de simulación y la

distancia de corte para evaluar las interacciones electrostáticas sobre todo en sistemas pequeños, ya que pueden presentarse correlaciones artificiales en la estructura o el movimiento. En muchos casos es recomendable realizar pruebas comparativas con cajas de simulación de distintos tamaños antes de elegir las dimensiones finales del sistema.

## A.2 Métodos de búsqueda y minimización de energía

Las simulaciones de un sistema molecular, requieren de la generación de un conjunto estadístico de configuraciones representativo del sistema a la temperatura de estudio para que, a partir de éste, sea posible calcular las propiedades macroscópicas del sistema. Si se parte de una configuración fuera del equilibrio, las fuerzas asociadas al potencial pueden resultar muy grandes y en consecuencia la dinámica molecular del sistema puede fallar. Para eliminar este problema es necesario recurrir a un método de minimización de energía.

La energía potencial de un sistema macromolecular es una función muy compleja que, para ser representada gráficamente, necesitaría recurrir a un extenso y complicado número de dimensiones. La función posee un *mínimo global* que es el punto más profundo de la hipersuperficie a la que daría lugar en dicho espacio multidimensional. Existen además muchos *mínimos locales* que se encuentran representados dentro de esta hipersuperficie. Si se conocieran todos los mínimos y puntos de inflexión sería posible identificar estructuras y conformaciones relevantes. Además, podrían obtenerse sus energías libres, la dinámica del sistema y las diversas transiciones estructurales. Desafortunadamente, el número de dimensiones y de mínimos locales del espacio de configuraciones es tan alto que no es posible barrerlo con la intensidad suficiente como para hacer un estudio completo. No existe ningún método que garantice la determinación del mínimo global en una cantidad práctica de tiempo. Pero en cualquier caso, a partir de una estructura inicial es posible encontrar el *mínimo local* más cercano. Cercano en este contexto implica el mínimo que puede ser alcanzado por un movimiento sistemático bajando paso a paso el

gradiente local. Si lo que se quiere es alcanzar mínimos no tan próximos pero de menor energía, se pueden hacer aumentos y disminuciones sistemáticas de temperatura con la finalidad de saltar barreras de potencial capaces de confinar el sistema en una determinada conformación. Esta técnica se llama *simulated annealing*, que se podría traducir como *recocido simulado* por los sucesivos calentamientos y enfriamientos. Entre los métodos más utilizados para minimizar la energía de un sistema molecular destacan los siguientes:

- Aquéllos que requieren únicamente evaluar la función. Cada paso se realiza en base al resultado de la evaluación previa.
- Los que utilizan las derivadas de la energía potencial (fuerzas) para avanzar hacia el mínimo.
- Aquéllos basados en la segunda derivada de la función potencial. Estos métodos son de gran eficiencia en cuanto a su convergencia cerca del mínimo.

El método *steepest descent*, disponible en GROMACS, es de la segunda clase. Simplemente toma un paso en la dirección del gradiente negativo sin ninguna otra consideración. El método del gradiente conjugado, que utiliza la información obtenida en pasos previos, es muy efectivo para converger hacia el mínimo local pero su funcionamiento es limitado cuando se encuentra lejos del mínimo. Este método es, en general, más lento que el *steepest descent* pero más eficiente para localizar el mínimo energético. En GROMACS también se encuentra disponible el método L-BFGS, cuyo funcionamiento es comparable al método del gradiente conjugado y, en ocasiones, puede llegar a superarlo en rapidez.

### **A.3 Definiciones y unidades**

En cálculo computacional es importante tener cuidado con la definición de las unidades que se van a utilizar. El paquete GROMACS utiliza un grupo de unidades consistentes que producen valores cercanos a la unidad para la mayoría de las magnitudes moleculares

**Tabla A.2:** *Unidades básicas*

Cantidad	Símbolo	Unidades
Longitud	$r$	$nm = 10^{-9}m$
Masa	$m$	$u = 1,6605402(10) \times 10^{-27} Kg$ (1/12 la masa de un átomo $^{12}C$ )
Tiempo	$t$	$ps = 10^{-12} s$
Carga	$q$	$e$ =carga del electrón $= 1,60217733(49) \times 10^{-19} C$
Temperatura	$T$	$K$

**Tabla A.3:** *Unidades derivadas*

Cantidad	Símbolo	Unidad
Energía	$E, V$	$kJ mol^{-1}$
Fuerza	$F$	$kJ mol^{-1} nm^{-1}$
Presión	$p$	$kJ mol nm^3 = 10^{30}/N_{AV} Pa$
Velocidad	$v$	$nm ps^{-1} = 1000 m/s$
Momento dipolar	$\mu$	$e nm$
Potencial eléctrico	$\Phi$	$kJ mol^{-1} e^{-1} = 0,010364272(3) Volt$
Campo eléctrico	$E$	$kJ mol^{-1} nm^{-1} e^{-1} = 1,0364272(3) \times 10^7 V/m$

más relevantes. Las unidades básicas en este sistema son el nanómetro ( $nm$ ), el pico-segundo ( $ps$ ), el Kelvin ( $K$ ), la carga del electrón ( $e$ ) y la unidad de masa atómica ( $u$ ), que se encuentran listadas en el cuadro A.2. Las unidades de las magnitudes derivadas se encuentran listadas en el cuadro A.3.

## A.4 Campo de fuerza

Un campo de fuerza es un grupo de ecuaciones que definen la forma en que varía la energía potencial del sistema. Las ecuaciones describen las interacciones atómicas mediante un conjunto de parámetros que caracterizan, dentro de un contexto específico, a cada elemento del sistema. En la literatura se puede encontrar una gran variedad de campos de fuerza, con diferentes niveles de aproximación en el cálculo. La elección de un campo de fuerza específico dependerá directamente de las propiedades del sistema a estudiar, así como de las propiedades que se desean medir en el sistema. Los campos de fuerza no forman parte de GROMACS, pero este paquete permite utilizar diferentes campos de fuerza como GROMOS96 [49, 50, 51], OPLS [52], AMBER [53], CHARMM [54], entre otros, de los cuales sus parámetros son refinados y optimizados continuamente por grupos de investigación especializados.

La función que expresa la energía potencial del sistema puede desglosarse en tres contribuciones principales:

- i. Términos de no enlace: Potenciales de Lennard-Jones, Buckingham, Coulomb o Coulomb modificado; que definen las interacciones entre átomos no enlazados covalentemente.
- ii. Términos de enlace: distancias entre enlace, ángulos y torsiones. Estos términos se refieren a interacciones entre átomos que están unidos por enlaces covalentes de manera directa o a través de otros átomos (hasta una *distancia* de tres enlaces).
- iii. Especiales: como potenciales relacionados con fuerzas sobreimpuestas al sistema, con objetivos específicos como restringir el movimiento en alguna dirección determinada, separar dos moléculas que están interactuando de manera favorable, etc.



**Figura A.1:** *Fuerzas de Lennard-Jones.*

### A.4.1 Interacciones entre átomos no enlazados

Las interacciones entre átomos no enlazados se consideran aditivas por pares y con simetría esférica.

$$V(\mathbf{r}_1, \dots, \mathbf{r}_N) = \sum_{i < j} V_{ij}(\mathbf{r}_{ij}); \quad (\text{A.5})$$

$$\mathbf{F}_i = - \sum_j \frac{\partial V_{ij}(r_{ij})}{\partial r_{ij}} \frac{\mathbf{r}_{ij}}{r_{ij}} = -\mathbf{F}_j \quad (\text{A.6})$$

Las interacciones entre átomos no enlazados contienen un término de repulsión, un término dispersivo y uno más para las interacciones Coulómbicas. Los términos de repulsión y dispersión se combinan en el potencial de Lennard-Jones o, alternativamente, en el potencial de Buckingham. Por su parte, los átomos parcialmente cargados actúan a través de un potencial de Coulomb.

#### Potencial de Lennard-Jones

El potencial de Lennard-Jones para dos átomos del mismo tipo se expresa por:

$$V_{LJ}(r_{ij}) = \frac{C_{ij}^{12}}{r_{ij}^{12}} - \frac{C_{ij}^6}{r_{ij}^6} \quad (\text{A.7})$$

Los parámetros  $C_{ij}^6$  y  $C_{ij}^{12}$  dependerán del tipo de pares de átomos para los cuales se realice el cálculo. La fuerza derivada de este potencial es:

$$\mathbf{F}_i(r_{ij}) = \left( 12 \frac{C_{ij}^{12}}{r_{ij}^{13}} - 6 \frac{C_{ij}^6}{r_{ij}^7} \right) \frac{\mathbf{r}_{ij}}{r_{ij}} \quad (\text{A.8})$$

El potencial LJ puede escribirse también de la siguiente forma:

$$V_{LJ}(r_{ij}) = 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (\text{A.9})$$

Cuando se construye la matriz de parámetros necesaria para utilizar este potencial, se utilizan en GROMACS dos tipos de reglas de mezcla:

$$C_{ij}^{(6)} = (C_{ii}^6 C_{jj}^6)^{\frac{1}{2}} \quad (\text{A.10})$$

$$C_{ij}^{(12)} = (C_{ii}^{12} C_{jj}^{12})^{\frac{1}{2}} \quad (\text{A.11})$$

o, alternativamente,

$$\sigma_{ij} = \frac{1}{2}(\sigma_{ii} + \sigma_{jj}) \quad (\text{A.12})$$

$$\epsilon_{ij} = (\epsilon_{ii}\epsilon_{jj})^{\frac{1}{2}} \quad (\text{A.13})$$

### Potencial de Buckingham

El potencial de Buckingham posee un término más realista y flexible que el potencial de Lennard-Jones, para representar la contribución debida a la repulsión entre pares de átomos. Esta ventaja tiene el inconveniente de incrementar el costo en el tiempo de cómputo.

La forma analítica del potencial es:

$$V_{bh}(r_{ij}) = A_{ij} \exp(-B_{ij}r_{ij}) - \frac{C_{ij}}{r_{ij}^6} \quad (\text{A.14})$$

la fuerza derivada de este potencial es:

$$\mathbf{F}_i(r_{ij}) = \left[ -A_{ij}B_{ij} \exp(-B_{ij}r_{ij}) - 6 \frac{C_{ij}}{r_{ij}^7} \right] \frac{\mathbf{r}_{ij}}{r_{ij}} \quad (\text{A.15})$$

### Potencial de Coulomb

El potencial de Coulomb, que caracteriza la interacción entre dos partículas parcialmente cargadas, viene dado por:



**Figura A.2:** *Potencial de Coulomb.*

$$V_c(r_{ij}) = f \frac{q_i q_j}{\varepsilon_r r_{ij}} \quad (\text{A.16})$$

donde  $f = (4\pi\varepsilon_0)^{-1} = 138.935485$ . La fuerza que deriva de este potencial es representada por:

$$\mathbf{F}_i(\mathbf{r}_{ij}) = f \frac{q_i q_j}{\varepsilon_r r_{ij}^2} \frac{\mathbf{r}_{ij}}{r_{ij}} \quad (\text{A.17})$$

donde  $\varepsilon_r$  es la constante dieléctrica.

### Potencial de Coulomb con Campo de Reacción

Las interacciones coulómbicas pueden modificarse, asumiendo un entorno dieléctrico más allá de un radio de corte  $r_c$ . El potencial para esta interacción quedaría descrito por:

$$V_{crf} = f \frac{q_i q_j}{r_{ij}} \left[ 1 + \frac{\varepsilon_{rf} - 1}{2\varepsilon_{rf} + 1} \frac{r_{ij}^3}{r_c^3} \right] - f \frac{q_i q_j}{r_c} \frac{3\varepsilon_{rf}}{2\varepsilon_{rf} + 1} \quad (\text{A.18})$$

En la cual, el término constante de la derecha anula el valor del potencial en el radio de corte  $r_c$ . Así se puede reescribir el potencial como:

$$V_{crf} = f q_i q_j \left[ \frac{1}{r_{ij}} + k_{rf} r_{ij}^2 - C_{rf} \right] \quad (\text{A.19})$$

con

$$K_{rf} = \frac{1}{r_c^3} \frac{\epsilon_r f - 1}{(2\epsilon_r f + 1)} \quad (\text{A.20})$$

$$C_{rf} = \frac{1}{r_c} + K_{rf} r_c^2 = \frac{1}{r_c} \frac{3\epsilon_r f}{(2\epsilon_r f + 1)} \quad (\text{A.21})$$

Para  $\epsilon_{rf}$  muy grandes la  $k_{rf}$  tiende a  $0.5 r_c^{-3}$  mientras que para  $\epsilon_{rf} = 1$  la corrección desaparece. Esto hace posible el uso de la misma expresión con o sin campo de reacción.

La fuerza que deriva de esta función potencial viene dada por la siguiente expresión:

$$\mathbf{F}_i(\mathbf{r}_{ij}) = f q_i q_j \left[ \frac{1}{r_{ij}^2} - 2k_{rf} r_{ij} \right] \frac{\mathbf{r}_{ij}}{r_{ij}} \quad (\text{A.22})$$

#### A.4.2 Interacciones entre átomos enlazados

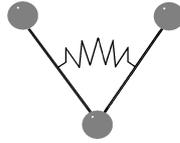
La evaluación de las interacciones entre átomos enlazados se basan en una lista fija de átomos. Éstas no son exclusivamente interacciones entre pares de átomos, como lo son las interacciones de longitud de enlace, si no que además se incluyen interacciones simultáneas entre tres átomos, como en un ángulo de enlace, e interacciones entre cuatro átomos que forman ángulos diédros. A continuación, se describe la manera en la que las interacciones entre átomos enlazados es evaluada en terminos de su contribución al potencial.

##### Enlaces covalentes

La longitud de enlace entre dos átomos enlazados covalentemente  $j$  e  $i$  (Figura A.3), queda definida por un potencial armónico:

$$V_b(r_{ij}) = \frac{1}{2} k_{ij}^b (r_{ij} - b_{ij})^2 \quad (\text{A.23})$$

con la fuerza,


**Figura A.3:** *Enlaces.*

**Figura A.4:** *Ángulos simples.*

$$\mathbf{F}_i(\mathbf{r}_{ij}) = k_{ij}^b (r_{ij} - b_{ij}) \frac{\mathbf{r}_{ij}}{r_{ij}} \quad (\text{A.24})$$

### Ángulos entre grupos de tres átomos

Las oscilaciones del ángulo formado por tres diferentes átomos  $i - j - k$  enlazados covalentemente entre sí (Figura A.4), son representadas de igual manera por un potencial armónico sobre el ángulo  $\theta_{ijk}$  formado por dos enlaces consecutivos

$$V_\alpha(\theta_{ijk}) = \frac{1}{2} k_{ijk}^\theta (\theta_{ijk} - \theta_{ijk}^0)^2 \quad (\text{A.25})$$

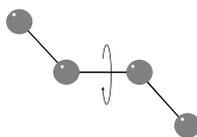
las fuerzas correspondientes, quedan expresadas de la siguiente manera:

$$\begin{aligned} \mathbf{F}_i &= -\frac{dV_\alpha(\theta_{ijk})}{d\mathbf{r}_i} \\ \mathbf{F}_k &= -\frac{dV_\alpha(\theta_{ijk})}{d\mathbf{r}_k} \quad \text{donde} \quad \theta_{ijk} = \arccos \frac{(\mathbf{r}_{ij} \cdot \mathbf{r}_{kj})}{r_{ij} r_{kj}} \\ \mathbf{F}_j &= -\mathbf{F}_i - \mathbf{F}_k \end{aligned} \quad (\text{A.26})$$

la numeración  $i, j, k$  representa a los átomos enlazados,  $j$  es el átomo intermedio.

### Ángulos diédros propios

Los ángulos diédros propios describen las torsiones que se dan entre ángulos formados por cuatro átomos  $i, j, k, l$ . Por convención  $\phi$  es el ángulo entre los planos, cada uno de ellos generado por un grupo de tres átomos consecutivos ( $i, j, k$ ) por un lado y ( $j, k, l$ ) por otro (Figura A.5). Un valor de cero para  $\phi$  corresponde a la configuración *cis* ( $i$  y  $l$  orientados hacia el mismo lado). Este ángulo se utiliza para describir las torsiones que tienen lugar en cadenas hidrocarbonadas por medio de un potencial.



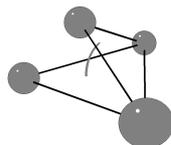
**Figura A.5:** *Ángulos diédros propios.*

$$V_{rb}(\phi_{ijkl}) = \sum_{n=0}^5 C_n (\cos \Psi)^n \quad (\text{A.27})$$

donde  $\Psi = \phi - 180^\circ$

### Ángulos diédros impropios

Éstos describen las interacciones entre grupos de cuatro átomos,  $i, j, k$  y  $l$ . Los ángulos diédros impropios se definen como los ángulos que se forman entre dos planos, cada uno de ellos generado por un grupo de tres átomos ( $i, j, k$ ) por un lado y ( $j, k, l$ ) por otro (Figura A.6).



**Figura A.6:** *Ángulos diédros impropios.*

$$V_{id}(\xi_{ijkl}) = \frac{1}{2}k\xi(\xi_{ijkl} - \xi_o) \quad (\text{A.28})$$

## A.5 Algoritmos

### A.5.1 Condiciones periódicas

Al simular un sistema de tamaño finito, es necesario establecer las condiciones de frontera, sobre las cuales será definido el tamaño del sistema. En un sistema pequeño compuesto por entre 100 y 1000 átomos, las partículas más externas formarán una superficie de contacto con sus alrededores. Si lo que se desea simular es el bulto de un sistema, esta frontera deberá eliminarse. Por esta razón en mecánica clásica se trata a los límites del sistema aplicando condiciones periódicas de frontera. Para ello los átomos del sistema se colocan dentro de una caja de simulación capaz de llenar todo el espacio cuando se rodea de imágenes periódicas de sí misma en las tres dimensiones. La caja con geometría definida es rodeada por 26 ( $= 3^3 - 1^3$ ) imágenes idénticas de sí misma trasladadas una distancia  $\pm R$  en las direcciones  $x, y$  y  $z$ . La cantidad de imágenes de la siguiente capa que rodea la caja de simulación, será formada por  $5^3 - 3^3 - 1^3$  y así consecutivamente.

Las cajas de simulación son típicamente cúbicas, aunque se utilizan diferentes geometrías bajo determinadas circunstancias, como por ejemplo, cajas rectangulares, triclinicas, octaédricas, dodecaédricas, etc. Para sistemas moleculares en solución, las cajas de forma dodecaédrica suelen ser más convenientes, porque permiten utilizar un 29 % menos de volumen de disolvente, y por tanto, ahorrar tiempo de cómputo para llevar a cabo la simulación. Además, las cajas de forma dodecaédrica incrementan la isotropía del espacio al aproximarse más a la geometría de una esfera.

Cuando la distribución del sistema es incompatible con la representación periódica, la solución es tratar una parte del sistema como una pared en donde el movimiento de los átomos queda restringido parcialmente. En algunos casos no es conveniente considerar las interacciones de un átomo en la caja central con todas sus imágenes periódicas, por

lo que las únicas interacciones que se toman en cuenta serán aquellas formadas con las imágenes vecinas más cercanas. Para evitar los problemas de anisotropía debido a la forma de la imagen más cercana es común aplicar un radio de corte esférico para la evaluación de determinado tipo de interacción. Las condiciones periódicas de frontera no sólo afectan al cálculo de las fuerzas, sino también las posiciones de los átomos. En la práctica es conveniente mantener a los átomos agrupados dentro de la caja central de simulación; cuando ocurre que un átomo sale de la la caja de simulación por un extremo, este mismo entra con la misma velocidad con la que dejó la caja, pero por el lado opuesto.

En las simulaciones que se presentan en este trabajo se utilizan condiciones periódicas más la convención de *imagen mínima*: esto es, que sólo la imagen más cercana de cada partícula, es considerada para las interacciones de largo alcance entre átomos no enlazados.

### **Radio de corte**

La convención de *imagen mínima* implica que el radio de corte utilizado para definir el límite de las interacciones de no-enlace, no deberá exceder la mitad del lado más pequeño de la caja. Lo cual se expresa:

$$R_c < \frac{1}{2} \min(\| a \|, \| b \|, \| c \|) \quad (\text{A.29})$$

Cuando se estudia una macromolécula como en el caso de una proteína en solución, la restricción mas conveniente para definir el radio de corte es:

$$R_c < \frac{1}{2} \min(a_x, b_y, c_z) \quad (\text{A.30})$$

### **A.5.2 El concepto de grupo**

Tanto en las simulaciones de dinámica molecular realizadas con GROMACS como en las rutinas de análisis de este paquete de programas, se utiliza el concepto de grupo para llevar a cabo determinadas acciones. Un grupo es un conjunto de átomos sobre los que

queremos imponer una determinada restricción o que nos interesa analizar de manera independiente. Típicamente un grupo puede estar formado por los átomos de una molécula, por los átomos del disolvente, o por varios átomos o moléculas distribuidas bajo algún criterio específico. Por ejemplo, utilizando el concepto de grupo la temperatura puede ser controlada de manera independiente para diferentes moléculas o grupos de moléculas, también es posible *congelar* una parte del sistema simulando sólo la dinámica del resto de los átomos o imponer una fuerza controlada sobre un conjunto de átomos, entre otras posibilidades.

### A.5.3 Algoritmos para la presión y temperatura

Para realizar simulaciones de dinámica molecular en sistemas fuera del equilibrio, como en el estudio de procesos irreversibles, efectos de catalizadores o propiedades de transporte, es necesario de implementar límites y restricciones al sistema para el control de temperatura y presión. Del mismo modo, si se desea simular de sistemas en equilibrio, el control automático de la temperatura y presión como variables independientes es muy conveniente. Para controlar estas propiedades, se utilizan algoritmos que introducen correcciones a cada paso de tiempo de la simulación.

#### Termostato de Berendsen

El algoritmo de Berendsen imita el acoplamiento, mediante una ecuación cinética de primer orden de la temperatura instantánea del sistema, a un baño externo con una temperatura de referencia  $T_0$  [?]. La corrección entre temperatura del sistema y la del baño ocurre lentamente de acuerdo a la siguiente expresión:

$$\frac{\partial T}{\partial t} = \frac{T_0 - T}{\tau} \quad (\text{A.31})$$

Lo cual indica que la desviación de la temperatura decae exponencialmente con una constante de tiempo  $\tau$ . Esta constante de acoplamiento puede modificarse en función de las

características de la simulación. El flujo de calor que entra o sale del sistema se corrige escalando las velocidades de cada partícula en cada paso con un factor dependiente del tiempo dado por:

$$\lambda = \sqrt{1 + \frac{\Delta t}{\tau_T} \left[ \frac{T_0}{T(t - \frac{\Delta t}{2})} - 1 \right]} \quad (\text{A.32})$$

con:

$$\tau_T = \tau N_{df} k / 2Cv \quad (\text{A.33})$$

Los valores de  $\tau$  y  $\tau_T$  son diferentes porque el cambio de energía cinética debido al escalamiento de las velocidades se distribuye entre la energía cinética y la energía potencial. Por tanto el cambio de temperatura es menor que el escalamiento de energía. En la práctica el cociente  $\tau/\tau_T$  varía entre 1 y 3 en función del sistema.

Hay otros algoritmos que se utilizan para controlar la temperatura, como el de Nosé-Hoover, pero en este trabajo hemos utilizado el de Berendsen por razones de eficiencia.

### Barostato de Berendsen

El algoritmo de Berendsen para controlar presión consiste en un reescalamiento de las coordenadas y los vectores de la caja de simulación en cada paso de tiempo. Como resultado, la presión del sistema se corrige lentamente de acuerdo a la siguiente expresión:

$$\frac{dP}{dt} = \frac{P_0 - P}{\tau_p} \quad (\text{A.34})$$

La matriz de escalamiento  $\mu$  viene dada por:

$$\mu_{ij} = \delta_{ij} - \frac{\Delta t}{3\tau_p} \beta_{ij} (P_{0ij} - P_{ij}(t)) \quad (\text{A.35})$$

donde  $\beta$  es el factor de compresibilidad isotérmica del sistema. Para agua a 1 atm y 300 K,  $\beta = 4.6 \times 10^{-10} Pa^{-1} = 4.6 \times 10^{-5} bar^{-1}$ . Los valores para otros líquidos típicamente utilizados como disolventes son del mismo orden.

Hay otros algoritmos que se utilizan para controlar la presión, como el de Parrinello-Rahman, pero en este trabajo hemos utilizado el de Berendsen por razones de eficiencia.

## Bibliografía

- [1] Pal S.K. y Zewail A.H. Dynamics of water in biological recognition. *Chem. Rev.*, 104:2099–2123, **2004**.
- [2] Pal S.K., Peon J. y Zewail A.H. Biological water at the protein surface: dynamical solvation probed directly with femtosecond resolution. *Proc. Natl. Acad. Sci. USA*, 99:1763–1768, **2002**.
- [3] Zaks, A. y Klibanov, A.M. The effect of water on enzyme action in organic media. *J. Biol. Chem.*, 263:8017–8021, **1998**.
- [4] Zaks, A. y Klibanov, A.M. Enzymatic catalysis in nonaqueous solvents. *J. Biol. Chem.*, 263:3194–3201, **1998**.
- [5] Zaks, A. y Klibanov, A.M. Substrate specificity of enzymes in organic solvents vs water is reversed. *J. Am. Chem. Soc.*, 108:2767–2768, **1986**.
- [6] Zaks, A. y Klibanov, A.M. Enzyme-catalyzed processes in organic solvents. *Proc. Natl. Acad. Sci. USA*, 82:3192–3196, **1985**.
- [7] Nandi N. y Bagchi B. Dielectric relaxation of biological water. *J. Phys. Chem. B*, 101:10954–10961, **1997**.
- [8] Gómez-Puyou, M.T. y Gómez-Puyou, A. Enzymes in low water systems. *Crit. Rev. Biochem. Mol. Biol.*, 33:53–89, **1998**.

- 
- [9] Gonnelli, M. y Strambini G.B. Protein dynamical structure by tryptophan phosphorescence and enzymatic activity in reverse micelles: 2. alkaline phosphatase. *J. Phys. Chem.*, 92:2854–2857, **1988**.
- [10] Soares C.M., Teixeira, V.H., y Baptista, A.M. Protein structure and dynamics in nonaqueous solvents: insights from molecular dynamics simulations studies. *Biophys. J.*, 84:1628–1641, **2003**.
- [11] Yang L., Dordick J.S. y Garde S. Hydration of enzyme in nonaqueous media is consistent with solvent dependence of its activity. *Biophys. J.*, 87:812–821, **2004**.
- [12] World Health Organization (<http://www.who.org>).
- [13] Maldonado, E., Soriano-García, M., Moreno A., Cabrera, N., Garza-Ramos, G., Gómez-Puyou, M.T., Gómez-Puyou, A. y Perez-Montfort, R. Differences in the intersubunit contacts in triosephosphate isomerase from two closely related pathogenic trypanosomes. *J. Mol. Biol.*, 283:193–203, **1998**.
- [14] Téllez-Valencia, A., Ávila-Ríos, S., Pérez-Montfort, R., Rodríguez-Romero, A., Gómez-Puyou, M.T., López-Calahorra, F. y Gómez-Puyou, A. Highly specific inactivation of triosephosphate isomerase from *Trypanosoma cruzi*. *Biochem. Biophys. Res. Commun.*, 295:958–963, **2002**.
- [15] Téllez-Valencia, A., Olivares-Illana, V., Hernández-Santoyo, A., Pérez-Montfort, R., Costas, M., Rodríguez-Romero, A., López-Calahorra, F., Gómez-Puyou, M. T. y Gómez-Puyou, A. Inactivation of triosephosphate isomerase from *Trypanosoma cruzi* by an agent that perturbs its dimer interface. *J. Mol. Biol.*, 341:1355–1365, **2004**.
- [16] Zomosa-Signoret, V., Hernández-Alcántara, G., Reyes-Vivas, H. Martínez-Martínez, E. Garza-Ramos, G., Pérez-Montfort, R., Gómez-Puyou, M.T. y Gómez-Puyou, A.

- Control of the reactivation kinetics of homodimeric triosephosphate isomerase from unfolded monomers. *Biochemistry*, 42:3311–3318, **2003**.
- [17] Gómez-Puyou, A., Saavedra-Lira, E., Becker, I., Zubillaga, R.A., Rojo-Domínguez, A. y Pérez-Montfort, R. Using evolutionary changes to achieve species-specific inhibition of enzyme action-studies with triosephosphate isomerase. *Chem. Biol.*, 2:847–855, **1995**.
- [18] Olivares-Illana, V., Pérez-Montfort, R., López-Calahorra, F., Costas, M., Rodríguez-Romero, A., Gómez-Puyou, M.T. y Gómez-Puyou, A. Structural differences in triosephosphate isomerase from different species and discovery of a multitritypanosomatid inhibitor. *Biochemistry*, 45:2556–2560, **2006**.
- [19] Cox, F.E.G. History of human parasitology. *Clinical Microbiology Reviews*, 15:595, **2002**.
- [20] Chagas, C. Nova Tripanozomiase humana. estudos sobre a morfologia e o ciclo evolutivo do *Shizotrypanum cruzi* n. gen., n. sp., agente etiologico de nova entidade morbida do homem. *Mem. Inst. Oswaldo Cruz, Rio de Janeiro.*, 1:1–80, **1909**.
- [21] Ramsey, J.M. y Schofield, C.F. Control of chagas disease vectors. *Salud Pública de México*, 45, **2003**.
- [22] Rassi, A. Jr., Rassi, A. y Little, W. Chagas' heart disease. *Clin. Cardiol.*, 23:883–889, **2000**.
- [23] Dias, J.C.P., Silveira, A.C. y Schofield, C.J. The impact of chagas disease control in latin america-A review. *Mem. Inst. Oswaldo Cruz, Rio de Janeiro.*, 97:603–612, 2002.
- [24] Urbina, J.A. Chemoteraphy of chagas disease. *Curr. Pharm. Design.*, 8:287–295, **2002**.

- [25] Davenport, R.C., Bash, P.A., Seaton, B.A., Bash, P.A., Seaton, B.A., Karplus, M., Petsko, G.A. y Ringe, D. Structure of the triosephosphate isomerase-phosphoglycolohydroxamate complex: An analogue of the intermediate on the reaction pathway. *Biochemistry*, 30:5821–5826, **1991**.
- [26] Reardon, D. y Farber, G.K. The structure and evolution of barrel proteins. *FASEB*, 9:497–503, **1995**.
- [27] <http://www.rcsb.org/pdb/>.
- [28] Nagano, N., Hutchison, E.G. y Thornton, J.M. Barrel structures in proteins: automatic identification and classification including a sequence analysis of TIM barrels. *Protein Sci.*, 8:2072–2084, **1999**.
- [29] Maes, D., Zeelen, J.Ph., Thanki, N., Beaucamp, N., Alvarez, M., Hoa Dao Thi, M., Backmann, J., Martial, J.A., Wyns, L., Jaenicke, R. y Wierenga, R.K. The crystal structure of triosephosphate isomerase (TIM) from *Thermotoga maritima*: a comparative thermostability structural analysis of ten different TIM structures. *Proteins*, 37:441–453, **1999**.
- [30] Urfer, R. y Kirschner, K. The importance of surface loops for stabilizing an eightfold  $\beta\alpha$  barrel protein. *Protein Sci.*, 1:31–45, **1992**.
- [31] Pompliano, D.L., Peyman, A. y Knowles, J.R. Stabilization of a reaction intermediate as a catalytic device: definition of the functional role of the flexible loop in triosephosphate isomerase. *Biochemistry*, 29:3186–3194, **1990**.
- [32] Garza-Ramos, G., Cabrera, N., Saavedra-Lima, E., Gómez-Puyou M.T., Ostoa-Saloma, P., Pérez-Montfort, R. y Gómez-Puyou, A. Sulfhydryl reagent susceptibility in proteins with high sequence similarity, triosephosphate isomerase from *Trypanosoma brucei*, *Trypanosoma cruzi* and *Leishmania mexicana*. *Eur. J. Biochemistry*, 253:684–691, **1998**.

- 
- [33] Mainfrod, V., Terpstra, P., Beauregard, M., Frere, J.-M., Mande, S.C., Hol, W.G.J., Martial, J.A. y Goraj, K. Three hTIM mutants that provide new insights on why TIM is a dimer. *J. Mol. Biol.*, 257:441–456, **1996**.
- [34] Gao, X., Maldonado, E., Pérez Montfort, R., Garza-Ramos, G., Gómez-Puyou, M.T., Gómez-Puyou, A. y Rodríguez-Romero, A. Crystal structure of triosephosphate isomerase from *Tripanosoma cruzi* in hexane. *Proc. Natl. Acad. Sci. USA.*, 96:10062–10067, **1999**.
- [35] Espinoza-Fonseca, L.M. y Trujillo-Ferrara, J.G. Exploring the possible binding sites at the interface of triosephosphate isomerase dimer as a potential target for anti-trypanosomal drug design. *Bioorganic Medicinal Chemistry Letters*, 14:3151–3154, **2004**.
- [36] Espinoza-Fonseca, L.M. y Trujillo-Ferrara, J.G. Structural considerations for the rational design of selective anti-trypanosomal agents: The role of the aromatic clusters at the interface of triosephosphate isomerase dimer. *Biochem. Biophys. Res. Commun.*, 328:922–928, **2005**.
- [37] Van Gunsteren, W.F., y Berendsen, H.J.C. Computer simulation of molecular dynamics: Methodology, applications, and perspectives in chemistry. *Chem. Int. Ed. Engl.*, 29:992–1023, **1990**.
- [38] [www.gromacs.org](http://www.gromacs.org).
- [39] Berendsen, H.J.C., van der Spoel, D. y van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Phys. Comm.*, 95:43–56, **1995**.
- [40] Lindahl, E., Hess, B. y van der Spoel, D. GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Model.*, 7, **2001**.

- 
- [41] Van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E. y Berendsen, H.J.C. GROMACS: Fast, flexible and free. *J. Comp. Chem.*, 25:1701–1718, **2005**.
- [42] Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F. y Hermans, J. Interactions models for water in relation to protein hydration, in intermolecular forces. *Eds. Pullman, B. Dordrecht, The Netherlands*, pages 331–342, **1981**.
- [43] Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., y Haak, J.R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684–3690, **1984**.
- [44] Miyamoto S. y Kollman, P.A. Settle: an analytical version of the shake and rattle algorithms for rigid water models. *J. Comp. Chem.*, 13:952–962, **1992**.
- [45] Humphrey, W., Dalke, A. y Schulten, K. VMD: VISUAL MOLECULAR DYNAMICS. *J. Mol. Graph.*, 14:33–38, **1996**.
- [46] Sayle, R.A. y Milner-White, E.J. Rasmol: Biomolecular graphics for all. *Trends in Biochem. Sci.*, 20:374–376, **1995**.
- [47] Wolfgang, K. y Christian, S. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22:2576–2637, 1983.
- [48] Feenstra, K.A., Hess, B. y Berendsen, H.J.C. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J. Comp. Chem.*, 20:786–798, **1999**.
- [49] Van Gunsteren, W.F., Billeter, S.R., Eising, A.A., Hünenberger, P.A., Krüger, P., Mark, A.E., Scott, W.R. y Tironi, I.G. Biomolecular simulation, the GROMOS96 manual and user guide. *Hochschulverlag AG der ETH, Zürich*, **1996**.

- [50] Schuler, L.D., Daura, X. y van Gunsteren, W.F. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *J. Comp. Chem.*, 22:1205–1218, **2001**.
- [51] Oostenbrink, C., Villa, A., Mark, A.E. y van Gunsteren, W.F. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53a5 and 53a6. *J. Comp. Chem.*, 25:1656–1676, **2004**.
- [52] Jorgensen, W.L. y Tirado-Rives, J. The OPLS force field for proteins. energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.*, 110:1657–1666, **1988**.
- [53] Case, D.A., Cheatman, T.E. III, Darden, T., Gohlke, H., Luo, R., Merz, K.M. Jr., Onufriev, A., Simmerling, S. Wang, B. y Woods, R.J. The amber biomolecular simulation programs. *J. Comp. Chem.*, 26:1668–1688, **2005**.
- [54] Brooks, B.R., Cruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. y Karplus, M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculation. *J. Comp. Chem.*, 4:187–217, **1983**.