



UNIVERSIDAD NACIONAL AUTÓNOMA DE
MÉXICO

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

**DETERMINACIÓN DEL EFECTO DEL DISOLVENTE Y DE LA ESTABILIDAD ESTRUCTURAL
DE TRIOSAFOFATO ISOMERASA DE DISTINTAS ESPECIES POR MEDIO DE
SIMULACIONES DE DINÁMICA MOLECULAR**

TESIS
PARA OPTAR POR EL GRADO DE

DOCTOR EN CIENCIAS

PRESENTA

M. EN C. NORMA DELIA DÍAZ VERGARA



TUTOR: DR. ÁNGEL PIÑEIRO GUILÉN AÑO: 2009



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

AGRADECIMIENTOS

- A la Universidad Nacional Autónoma de México
- Al Doctor Ángel Piñeiro Guillén por su apoyo incondicional, confianza y amistad.
- Al Doctor Xavier Daura Ribera por las facilidades y el apoyo proporcionado durante la estancia de investigación en su grupo de trabajo en el Instituto de Biotecnología y Biomedicina (IBB), UAB, Barcelona, España.
- A los Doctores Miguel Costas Basín, Silvia Pérez Casas y Luis Miguel Trejo por su apoyo, alojamiento y facilidades brindadas en el laboratorio 102 durante el desarrollo del trabajo.
- Al Biotecnólogo Ricard Boned por el apoyo técnico brindado para la elaboración de herramientas de cálculo usados en este trabajo.
- Al Biólogo Oscar Conchillo por las facilidades y ayuda proporcionadas como administrador de los sistemas cómputo en el IBB.
- A la QFB. Josefina Tenopala por sus consejos y junto con el personal del Posgrado en Ciencias Químicas por su amable atención y ayuda siempre que se requirió.
- A los M. en C. Leny Álvarez y Rodolfo Reynoso por su ayuda durante el periodo fuera de México.
- A los compañeros de trabajo y del posgrado en la Facultad de Química, y a los compañeros del IBB por la retroalimentación y apoyo durante el desarrollo del proyecto.
- Al Posgrado en Ciencias Químicas.

POR EL FINANCIAMIENTO OTORGADO:

- Al Consejo Nacional para la Ciencia y la Tecnología (CONACyT) de México por las becas otorgadas (No. Registro CONACyT 202514).
- Al Consejo Nacional para la Ciencia y la Tecnología (CONACyT) de México (Proyectos número J49811-Q).
- A la Facultad de Química.
- Al Posgrado en Ciencias Químicas.

POR LOS RECURSOS DE CÓMPUTO PROPORCIONADOS:

- A la Dirección General de Servicios de Cómputo Académico (DGSCA), UNAM, por el tiempo de cómputo utilizado en las simulaciones cuyos resultados se muestran en este trabajo.
- Al clúster de Computadoras Valhalla en el IBB, UAB. Barcelona.

LUGARES DONDE SE DESARROLLÓ EL TEMA:

- Laboratorio 102 Termofísica, Departamento de Fisicoquímica, Facultad de Química, Universidad Nacional Autónoma De México, México, D. F.
- Laboratorio de Bioinformática y Biología Computacional. Instituto de biotecnología y biomedicina. Universidad Autónoma de Barcelona, Barcelona, España.

LOS RESULTADOS DE ESTE TRABAJO DE INVESTIGACIÓN FUERON PRESENTADOS EN LOS SIGUIENTES CONGRESOS:

- Estimating differences of Dimeric Stabilities between Several triosephosphate isomerases by Molecular Dynamics Simulations Norma Díaz-Vergara, Ángel Piñero and Xavier Daura; Grand Challenges in Computational Biology, Barcelona, Spain, June 2008.
- Principal Motions of the Triosephosphate isomerase from Trypanosome cruzi at the Time Scale of 140 ns. Norma Díaz-Vergara and Ángel Piñero; 2nd Latin American Protein Society Meeting, Acapulco, Guerrero, México, November 2007.
- Understanding relevant internal motions of the triosephosphate isomerase from Trypanosoma cruzi by molecular dynamics simulations. Norma Díaz-Vergara and Ángel Piñero; 21st Annual Symposium of the Protein Society, Boston, Massachusetts, EUA, July 2007.
- A molecular dynamics study of triosephosphate isomerase from Trypanosoma cruzi in water/decane mixtures. Norma Díaz-Vergara and Ángel Piñero Guillén; 5th International workshop on Current Problems in Complex Fluids. Physical and Chemical Aspects of Molecular Biology, Puebla, México. January 4-6, 2007.

PUBLICACIÓN DERIVADA:

- Molecular Dynamics Study of Triosephosphate Isomerase from Trypanosoma cruzi in Water/Decane Mixtures. Norma Díaz-Vergara y Ángel Piñero. Journal of Physical Chemistry B, 2008, 112, 3529-3539.

Índice general

Capítulo 1. Motivación	1
1.1. Objetivo general	2
1.2. Objetivos específicos	2
Capítulo 2. Simulaciones de dinámica molecular y métodos de análisis	5
2.1. Simulaciones de dinámica molecular	5
2.2. Métodos de análisis	7
2.3. Energía de Gibbs y energía de Helmholtz	11
Parte 1. Estudio de la trifosfato isomerasa de <i>Trypanosoma cruzi</i> en mezclas agua/decano. Un estudio de dinámica molecular.	19
Capítulo 3. Introducción	21
Capítulo 4. Metodología	25
4.1. Construcción de los sistemas	25
4.2. Parámetros de simulación	26
4.3. Métodos de análisis	27
Capítulo 5. Resultados y discusión	29
5.1. Análisis visual de las trayectorias	29
5.2. Análisis de la dinámica y estructura de la TcTIM	31
5.3. Componentes principales del movimiento	34
5.4. Análisis estructural del lazo 6	36
5.5. Movilidad de los residuos catalíticos	36
5.6. Análisis del disolvente en el entorno de la proteína	38
5.7. Puentes de hidrógeno intramoleculares	44
Capítulo 6. Conclusiones	47
Parte 2. Estudio de la estabilidad dimérica de trifosfato isomerasa de <i>Trypanosoma cruzi</i>, <i>Trypanosoma brucei</i> y humana.	49

Capítulo 7. Introducción	51
7.1. Aproximación de Jarzynski	54
Capítulo 8. Metodología	59
8.1. Parámetros de simulación	59
8.2. Construcción de los sistemas	59
8.3. Métodos de análisis	60
Capítulo 9. Resultados y discusión	63
9.1. Análisis estructural	63
9.2. Análisis de las trayectorias de referencia	64
9.3. Análisis estructural	65
9.4. Cálculo de la Energía libre	68
9.5. Análisis visual de las trayectorias	70
Capítulo 10. Conclusiones	75
Capítulo 11. Conclusiones Generales	77
Bibliografía	79
Parte 3. Artículo publicado	87
Molecular Dynamics Study of Triosephosphate Isomerase from <i>Trypanosoma cruzi</i> in Water/Decane Mixtures. Norma Díaz-Vergara y Ángel Piñeiro. <i>Journal of Physical Chemistry B</i> , 2008, 112, 3529-3539.	89

CAPÍTULO 1

Motivación

No es fácil obtener información experimental detallada del comportamiento dinámico de una proteína a nivel atómico, así como de las moléculas del disolvente con quien ésta interactúa en escalas de tiempo en las que tienen lugar algunos de los movimientos que determinan su función (entre femtosegundos y microsegundos). Existen técnicas como la espectrometría de femtosegundos que permiten observar cambios conformacionales en regiones específicas de la proteína en estas escalas de tiempo. Sin embargo, estas técnicas presentan diferentes limitaciones dependiendo de la complejidad del sistema de estudio. En este sentido, las simulaciones de dinámica molecular, gracias a la implementación de nuevos algoritmos y métodos teóricos, y por supuesto a los procesadores y arquitecturas computacionales disponibles en la actualidad, permiten modelar de manera realista el comportamiento dinámico de proteínas en escalas de tiempo que van desde los femtosegundos hasta los microsegundos, e incluso milisegundos cuando los cálculos se realizan bajo ciertas aproximaciones.

Este trabajo se basa en un estudio de la enzima triosafosfato isomerasa (TIM) del parásito *Trypanosoma cruzi* mediante simulaciones computacionales de dinámica molecular. Específicamente, se analizó el comportamiento dinámico de la proteína en presencia de un modelo de solvente explícito. Además, se realizó un estudio comparativo de la estabilidad dimérica de TIM de distintas especies (*Trypanosoma cruzi*, *Trypanosoma brucei* y *humana*). La elección de la enzima se basó en la relación que ésta tiene con el parásito que produce la enfermedad denominada mal de Chagas (trypanosomiasis americana). En la actualidad, esta enfermedad parasitaria se considera como una de las más serias y con más repercusiones socioeconómicas en América Latina, en donde se estima que hay más de 20 millones de personas infectadas (www.who.org, WHO). Dada su importancia, numerosos grupos de investigación dedican grandes esfuerzos para tratar de combatir y erradicar esta enfermedad. Algunas de las líneas de investigación en vigor apuntan sus esfuerzos a la búsqueda de un fármaco que sea capaz de inhibir la actividad catalítica de la TIM del parásito sin afectar al anfitrión. Esta enzima es fundamental en la realización del proceso metabólico de la glucólisis, por medio del cual el parásito adquiere la energía necesaria

para su supervivencia. Sin embargo, inhibir esta enzima de manera específica no es sencillo, entre otras razones porque también se encuentra presente en el ser humano con una estructura ligeramente diferente. Considerando que el diseño del inhibidor debe ser específico para la enzima del parásito, cualquier información acerca de su comportamiento estructural y dinámico obtenida desde una aproximación a nivel atómico resulta de gran utilidad para el diseño de fármacos. Además, con el uso de esta técnica no sólo es posible extraer información acerca de la propia enzima, sino también de sus interacciones con el disolvente en el que ésta se encuentra.

En base a lo anterior con el desarrollo de este trabajo se plantean los siguientes objetivos:

1.1. Objetivo general

Describir el comportamiento dinámico que muestran las proteínas en solventes no polares con bajas concentraciones de agua en comparación con el que presenta en agua pura, por medio de simulaciones de dinámica molecular. Analizar los factores que brindan la estabilidad de los agregados diméricos en los que tiene lugar la estructura funcional de la triosafosfato isomerasa.

1.2. Objetivos específicos

- Describir a nivel atómico el comportamiento dinámico que muestran las proteínas en solventes no polares con bajas concentraciones de agua.
- Proporcionar información detallada del comportamiento dinámico de la triosafosfato isomerasa de *Trypanosoma cruzi*, *Trypanosoma brucei* y humana en agua.
- Identificar los movimientos dinámicos internos de triosafosfato isomerasa de *Trypanosoma cruzi*, *Trypanosoma brucei* y humana que puedan ser relacionados con su funcionalidad.
- Calcular las diferencias de energía libre del proceso de monomerización de triosafosfato isomerasa de *Trypanosoma cruzi*, *Trypanosoma brucei* y humana en agua.
- Estudiar las interacciones que participan directamente en la estabilidad de las subunidades en el dímero de cada estructura de TIM .
- Realizar una comparación entre la estabilidad dimérica de la triosafosfato isomerasa de *Trypanosoma cruzi*, *Trypanosoma brucei* y humana.

El trabajo de investigación que se presenta en esta memoria se divide en dos partes dedicadas a estudios de la TIM por medios de simulaciones computacionales.

La primera parte (1), está basada en el análisis del comportamiento dinámico de la triosafosfato isomerasa de *Trypanosoma cruzi* tanto en agua como en mezclas agua/decano como modelo de disolvente polar/apolar-orgánico. Aquí nos planteamos analizar el comportamiento dinámico de la propia enzima de manera que la información aportada contribuya a un mejor entendimiento y sea útil en el diseño de ligandos específicos. La información obtenida de este estudio será útil además para explicar a nivel atómico la actividad que muchas enzimas presentan en medios orgánicos con bajas concentraciones de agua. Complementamos nuestro estudio previo de los mismos sistemas (Díaz-Vergara, 2004) con un análisis más detallado en el que se incluye una descomposición de la trayectoria de la proteína en sus componentes principales, con especial atención en las regiones involucradas en su actividad catalítica. En la segunda parte de esta memoria se evalúa la estabilidad estructural en forma dimérica de la TIM de tres distintas especies (*Trypanosoma cruzi*, *Trypanosoma brucei* y *humana*) además de realizar un análisis comparativo del comportamiento dinámico de las tres enzimas. El método utilizado para evaluar la estabilidad dimérica es utilizado por primera vez en este trabajo para dicho fin e involucra múltiples simulaciones en las que se pasa de dímeros a monómeros de manera irreversible a través de la imposición de un potencial artificial repulsivo entre las subunidades de las que consta la proteína. Con esta información, y utilizando la expresión conocida como ecuación de Jarzynski, es posible estimar el cambio de energía de libre a lo largo de este proceso y por tanto determinar la estabilidad de cada uno de los tres homodímeros.

CAPÍTULO 2

Simulaciones de dinámica molecular y métodos de análisis

Conocer a nivel atómico el comportamiento dinámico de sistemas organizados y sus características estructurales contribuye en gran manera al entendimiento de los procesos biológicos que ocurren en los seres vivos. Ésto es posible por medio de técnicas computacionales en química, que pueden ir desde simulaciones a nivel de mecánica cuántica hasta la mecánica clásica por medio de simulaciones de dinámica molecular de agregados moleculares de tamaño relativamente grande. Por medio de simulaciones de dinámica molecular a nivel clásico es posible extraer información de la estructura, estabilidad y del comportamiento dinámico a partir de simulaciones clásicas del sistema.

2.1. Simulaciones de dinámica molecular

Las simulaciones de dinámica molecular presentadas en este trabajo están basadas en la resolución de las leyes de Newton de la mecánica clásica. Las ecuaciones de movimiento se emplean para describir las interacciones entre los átomos del sistema de estudio. Para un sistema de N átomos de masas $m_i (i = 1, 2, \dots, N)$ definido por los vectores de posición \mathbf{r}_i se tiene que (2.1.1):

$$(2.1.1) \quad m_i \frac{\partial^2 \mathbf{r}_i}{\partial t^2} = \mathbf{F}_i, \quad i = 1, \dots, N.$$

Las fuerzas se obtienen a partir de la función de potencial $\mathbf{V}_i (\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_3)$ (2.1.2):

$$(2.1.2) \quad \mathbf{F}_i = -\frac{\partial V}{\partial \mathbf{r}_i}$$

Por lo tanto, a lo largo de las simulaciones de dinámica molecular las ecuaciones 2.1.1 y 2.1.2 se resuelven simultáneamente para todas las partículas del sistema en pequeños pasos de tiempo. Este proceso se repite iterativamente a la vez que las coordenadas, datos de velocidad y energía son almacenadas generando así la trayectoria del sistema, el proceso se detiene hasta el tiempo deseado de simulación.

2.1.1. El campo de fuerza. La forma en la cual interactúan los átomos del sistema queda definido por el campo de fuerza. Éste consiste en un grupo de ecuaciones que determinan como varía la energía potencial del sistema en función de la posición de los átomos que lo componen así como de la distribución de los átomos del entorno. Además, implica la definición de una serie de tipos de átomos caracterizados dentro de un contexto específico, y de un conjunto de parámetros que relacionan las ecuaciones con los tipos de átomos involucrados.

En este trabajo se utilizó un campo de fuerza *conservativo*, además se consideró que las contribuciones al potencial son debidas únicamente a las posiciones de los átomos. La suposición anterior está basada en la aproximación de Born-Oppenheimer, la cual considera que los movimientos electrónicos y nucleares están desacoplados. Puesto que no se incluyen reacciones químicas en el sistema las consideraciones tomadas se consideran apropiadas.

Un campo de fuerza común para un sistema de N átomos de masas m_i ($i = 1, 2, \dots, N$) con posiciones \mathbf{r}_i tiene la forma siguiente:

$$(2.1.3) \quad V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = \sum_{\text{Enlaces}} \frac{1}{2} K_b [b - b_o]^2 +$$

$$(2.1.4) \quad \sum_{\text{ngulos}} \frac{1}{2} K_\theta [\theta - \theta_o]^2 +$$

$$(2.1.5) \quad \sum_{\substack{\text{diédros} \\ \text{impropios}}} \frac{1}{2} K_\xi [\xi - \xi_o]^2 +$$

$$(2.1.6) \quad \sum_{\substack{\text{diédros} \\ \text{propios}}} K_\varphi [1 + \cos n\varphi - \delta]^2 +$$

$$(2.1.7) \quad \sum_{\text{pares}(i,j)} [C_{12}(i,j)/r_{ij}^{12} - C_6(i,j)/r_{ij}^6] +$$

$$(2.1.8) \quad \sum_{\text{pares}(i,j)} \left[\frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} \right]$$

El primer término 2.1.3 describe la energía asociada a las interacciones de dos átomos unidos entre sí por un enlace covalente, y queda expresado como un potencial armónico. La constante b_o representa la longitud del enlace y K_b la constante de fuerza, los valores de dichos parámetros quedan determinados de acuerdo al tipo de átomo. La ecuación 2.1.4 describe la contribución al potencial de la energía

asociada a la variación del ángulo entre dos enlaces interatómicos contiguos. Esta interacción se modela mediante un oscilador armónico como en el caso de dos átomos enlazados, aunque en este caso la coordenada será el propio ángulo formado. Las energías asociadas a las oscilaciones entre enlaces covalentes formados por 4 átomos enlazados (i, j, k, l) se representan por los términos 2.1.5 y 2.1.6. La primera de estas contribuciones se describe nuevamente por un oscilador armónico parametrizado por el ángulo diedro que se genera entre los planos formados por los grupos $i - j - k$ y $j - k - l$. La ecuación 2.1.6 se refiere a los ángulos diedros φ definidos de la misma manera que el ángulo ξ de la ecuación 2.1.5, parametrizados esta vez mediante una función sinusoidal que permite giros completos de hasta 360° . Las torsiones de los átomos de carbono en cadenas alifáticas quedan descritas por la anterior contribución al potencial. Con los términos 2.1.7 y 2.1.8 se describe la interacción entre todos los pares de átomos que no interactúan a través de enlaces covalentes. Dichas contribuciones se expresan 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 a una distancia \mathbf{r}_{ij} .

2.2. Métodos de análisis

2.2.1. Desviaciones cuadráticas medias (RMSD). Representa una medida de la diferencia entre los valores observados de un parámetro con respecto a un valor de referencia. Aplicado al campo de las simulaciones de dinámica molecular este parámetro proporciona una medida de la desviación promedio entre una estructura de referencia y las estructuras generadas lo largo de la de simulación. La expresión que define este cálculo se expresa en la ecuación siguiente:

$$(2.2.1) \quad RMSD(t, t_0) = \left[\frac{1}{M} \sum_{i=1}^N m_i \|\mathbf{r}_i(t) - \mathbf{r}_i(t_0)\|^2 \right]^{\frac{1}{2}}$$

donde $M = \sum_{i=1}^N m_i$ y $\mathbf{r}_i(t)$ es la posición del átomo i a un tiempo t y $\mathbf{r}_i(t_0)$ se refiere a la estructura de referencia. El monitoreo del RMSD se puede utilizar como una medida de la convergencia de las simulaciones y de la cercanía de la estructura simulada con la obtenida experimentalmente. Cuando no se tiene una estructura experimental para ser utilizada como referencia en este cálculo, el RMSD se calcula usualmente comparando con la estructura inicial o una estructura promedio obtenida de una simulación.

2.2.2. Fluctuaciones cuadráticas medias (RMSF). Las fluctuaciones cuadráticas medias representan una medida de la desviación estándar de una partícula i con respecto a su posición de referencia. A diferencia del RMSD que proporciona una cantidad promediada sobre todas las partículas, el RMSF es una medida de la desviación de la posición de cada una de las partículas promediada sobre el número de configuraciones. La expresión que describe el cálculo es la siguiente:

$$(2.2.2) \quad RMSF(i) = \langle (\Delta \mathbf{r}_i^2) \rangle^{\frac{1}{2}} = \langle [\mathbf{r} - \langle \mathbf{r} \rangle]^{\frac{1}{2}} \rangle$$

2.2.3. Análisis de componentes principales (PCA). Este análisis permite desacoplar la dinámica de macromoléculas en movimientos concertados más simples, es decir, en movimientos que se llevan a cabo de manera simultánea por un grupo específico de átomos y que ocurren en una determinada dirección. El análisis de componentes principales, conocido como PCA por sus siglas en inglés de Principal Component Analysis (Amadei *et al.*, 1993), es un método analítico que permite identificar y prescindir de la información redundante filtrando un conjunto de variables independientes a partir de las cuales se puede describir el comportamiento dinámico del sistema casi en su totalidad. Analizar mediante este método trayectorias de dinámica molecular permite dividir el espacio configuracional en dos subespacios. Uno de ellos, al que se llamará espacio esencial, comprenderá los movimientos en la proteína que contienen la mayoría de las fluctuaciones atómicas relevantes en la escala de tiempo considerada. El segundo subespacio quedará conformado por los movimientos atómicos de baja frecuencia.

El método consiste en calcular la matriz de covarianzas \mathbf{C} de las posiciones atómicas, esta es calculada para toda una trayectoria dada después de eliminar los movimientos rotacionales y traslacionales globales de la molécula. Este ajuste usualmente se logra tras superponer todas las estructuras de la trayectoria sobre una de referencia. Para M configuraciones del sistema extraídas de la trayectoria y N número de átomos, \mathbf{C} se convierte en una matriz $3N \times 3N$ con elementos:

$$(2.2.3) \quad C_{ij} = \left\langle M_{ii}^{\frac{1}{2}} (x_i - \langle x_i \rangle) M_{jj}^{\frac{1}{2}} (x_j - \langle x_j \rangle) \right\rangle$$

donde M es la matriz diagonal que contiene el análisis de masa ponderada (mass weighted analysis) o la matriz unitaria en caso de que no se aplique la ponderación de la masa (non mass weighted analysis). C es una matriz simétrica $3N \times 3N$, que se diagonalizará con una matriz de transformación R :

$$(2.2.4) \quad R^T C R = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{3N}) \text{ donde } \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{3N}$$

Las columnas de la matriz R están compuestas por los autovectores o modos esenciales del movimiento que describen los movimientos dinámicos en la estructura, y con los autovalores asociados a cada autovector, que pueden interpretarse como la amplitud con la que cada modo normal contribuye a la dinámica global. Si se proyecta la trayectoria sobre los modos principales se puede obtener una película para cada modo esencial, la proyección se expresa como:

$$(2.2.5) \quad \mathbf{p}(t) = R^T M^{\frac{1}{2}} (x(t) - \langle x \rangle)$$

A menudo, es posible describir todos los movimientos que contribuyen a la dinámica del sistema únicamente con los primeros modos normales. Si la trayectoria se filtra a lo largo de un movimiento entonces se puede obtener una trayectoria en la dirección del componente principal del movimiento, es decir:

$$(2.2.6) \quad x^f(t) = \langle x \rangle + M^{-\frac{1}{2}} R_{\star i} p_i(t)$$

2.2.4. Clusters de estructuras. Esta parte del análisis consistió en generar *clusters* de estructuras, es decir agrupaciones de estructuras similares entre sí. La formación de cada *cluster* se realiza a través de la selección de conformaciones de la molécula estudiada que se generan a lo largo de cada una de las simulaciones de dinámica molecular. Usualmente este análisis está basado en medidas de distancias entre estructuras. Específicamente, el método empleado en este trabajo utiliza un algoritmo propuesto por Daura y cols. (Daura *et al.*, 1999) basado en el cálculo de RMSD de posiciones atómicas entre pares de estructuras. Para cada estructura se calcula el RMSD con respecto al resto de conformaciones y posteriormente se contabiliza el número de conformaciones en las que el valor de RMSD se encuentra dentro de un radio de corte determinado, dando así como resultado al número total de vecinos de dicha conformación. Aquella estructura con el mayor número de vecinos será tomada como la estructura central del *cluster*, y junto con sus vecinos formarán el primer *cluster*. Todas las estructuras del primer cluster serán excluidas del conjunto total de estructuras para continuar aplicando el algoritmo de manera iterativa. Los elementos de cada cluster generado compartirán la característica de que su valor de RMSD es menor a un determinado radio de corte de la estructura central. El valor

del radio de corte es arbitrario, y se especifica usualmente en función de los valores de RMSD obtenidos para cada estructura así como del número integrantes de cada grupo que se deseen obtener.

2.2.5. Puentes de Hidrógeno. Un puente de Hidrógeno puede describirse como un tipo particular de interacción electrostática atractiva de atracción localizada entre un hidrógeno con una carga parcial positiva y un átomo aceptor de protones. Existen diferentes criterios energéticos y geométricos que se utilizan en la inspección de conformaciones atómicas para contabilizar puentes de Hidrógeno en simulaciones de dinámica molecular. Nosotros utilizamos 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 Hidrógeno está unido covalentemente y A es el grupo aceptor (Oxígeno o Nitrógeno). Este cálculo puede realizarse tanto entre átomos de la misma molécula (puentes de Hidrógeno intramoleculares) como entre átomos de moléculas distintas, soluto-soluto, soluto-solvente, etc. (puentes de Hidrógeno intermoleculares).

2.2.6. Radio de Giro (RoG). El radio de giro proporciona una medida del compactamiento de una estructura. El radio de giro en los ejes x , y y z como función del tiempo para un grupo de átomos se calcula mediante la aplicación de la siguiente fórmula.

$$(2.2.7) \quad RoG = \left(\frac{\sum_i \|\mathbf{r}_i\|^2 m_i}{\sum_i m_i} \right)^{1/2}$$

2.2.7. Área de la superficie accesible al solvente (SAS). El área accesible al solvente (SAS) de una molécula es el área total de la superficie de van der Waals de la estructura de dicha molécula, que se encuentra accesible al solvente. El área puede ser clasificada en función del tipo de átomos que la conforman en hidrofóbica o hidrofílica. Este cálculo aplicado a las estructuras obtenidas por medio de las simulaciones de dinámica molecular permite monitorear el SAS a lo largo de la trayectoria. Para realizar este análisis se utilizó la metodología propuesta por Connolly (Connolly, 1983). Este método consiste en calcular la accesibilidad al solvente haciendo rodar una molécula de agua con geometría esférica sobre la superficie de la proteína generando el contorno de una superficie, posteriormente ésta es transformada por medios analíticos en el área accesible al solvente.

2.3. Energía de Gibbs y energía de Helmholtz

La energía de Gibbs (G) es una función termodinámica que determina la espontaneidad de un proceso termodinámico a temperatura y presión constantes. Está directamente relacionada con la probabilidad de que un proceso se lleve a cabo y, por tanto, con la constante de equilibrio asociada. Cuando se trabaja a temperatura y volumen constantes la propiedad equivalente es la energía de Helmholtz (F).

Utilizando la mecánica estadística, la energía de Helmholtz puede expresarse en función de la función de partición Z .

$$(2.3.1) \quad F = -k_B T \ln Z$$

en donde Z se define como

$$(2.3.2) \quad Z = \frac{1}{h^{3N} N!} \iint \exp^{-H(\mathbf{p}, \mathbf{r})/k_B T} d\mathbf{p} d\mathbf{r}$$

para un sistema de N partículas indistinguibles. Los vectores de $3N$ dimensiones \mathbf{r} y \mathbf{p} corresponden, respectivamente, a las coordenadas y los momentos conjugados de las partículas del sistema. Cada par de vectores (\mathbf{r}, \mathbf{p}) representa un punto en el espacio de fases del sistema definido por todos los posibles valores de \mathbf{r} y \mathbf{p} . De la definición de la función de partición Z se puede observar claramente que tanto la energía de Gibbs o de Helmholtz no pueden obtenerse directamente, dado que requeriría el conocimiento del espacio de fases en su totalidad. Sin embargo, es más sencillo determinar la diferencia entre F o G de dos estados de un sistema. Por lo tanto, la diferencia de energía F entre dos estados A y B está formalmente descrita de la manera siguiente

$$(2.3.3) \quad \Delta F_{BA} = -\Delta F_{AB} = F_B - F_A = -\beta \ln \left(\frac{Z_B}{Z_A} \right)$$

donde ΔF representa la diferencia de energía de Helmholtz, β es $1/k_B T$ y Z es la función de partición canónica.

Todos los métodos computacionales aplicados en el cálculo de la función de partición se encuentran con el mismo problema ya que para su obtención se requiere de un extenso muestreo del espacio de fases en los dos estados A y B . Ese exhaustivo análisis es inalcanzable a excepción de sistemas muy pequeños y tras asumir diversas suposiciones, por lo cual los métodos para calcular diferencias de energía F o G han tomado mayor interés.

La diferencia de energía de Helmholtz o Gibbs entre dos estados A y B de un sistema pueden determinarse a partir del trabajo realizado sobre el sistema para forzar la transición desde un estado hasta el otro. Los métodos estándar para calcular este cambio de energía son la Integración termodinámica (Thermodynamic Integration, TI) (Frenkel & Smit, 2002) y el método de perturbación de la energía libre (Free Energy Perturbation, FEP)(Frenkel & Smit, 2002). Ambas metodologías hacen uso de un parámetro de acoplamiento λ y definen al Hamiltoniano del sistema como una función de λ que conecta a los estados inicial y final del sistema. De tal manera que $\mathcal{H}(\lambda_A) = \mathcal{H}_A$ corresponde al Hamiltoniano en el estado A y $\mathcal{H}(\lambda_B) = \mathcal{H}_B$ al del estado B. Si el Hamiltoniano es una función de λ entonces la energía libre es también una función de λ . En este caso, la diferencia de energía libre entre dos estados A y B se expresa como:

$$(2.3.4) \quad \Delta F_{BA} = F(\lambda_B) - F(\lambda_A) = \int_{\lambda_A}^{\lambda_B} \frac{\partial F(\lambda)}{\partial \lambda} d\lambda$$

$$(2.3.5) \quad = \int_{\lambda_A}^{\lambda_B} \left\langle \frac{\partial \mathcal{H}(\lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda$$

donde $\langle \dots \rangle_{\lambda}$ representa un promedio sobre el colectivo del sistema en el correspondiente valor de λ .

A la ecuación 2.3 se le conoce como la fórmula de integración termodinámica. Los cálculos de TI pueden realizarse bajo diferentes esquemas. La integración puede realizarse continuamente mientras lentamente se cambia el parámetro λ desde λ_A a λ_B a lo largo de la simulación, este método se le conoce comúnmente por su nombre en inglés, *slow growth method*. Este esquema conlleva varias dificultades puesto que la equilibración pocas veces se alcanza con eficacia. El otro esquema se utiliza para realizar simulaciones cuando λ toma cierto valor y se evalúa la integral numéricamente. En este método la convergencia de las simulaciones puede ser verificado independientemente a cada valor de λ y de ser necesario simulaciones adicionales con distintos valores de λ pueden ser añadidos. El método de perturbación de la energía libre es muy demandante computacionalmente, dado que en cada estado intermedio deberán realizarse simulaciones en equilibrio durante una considerable escala de tiempo para asegurar un muestreo representativo del colectivo termodinámico. Cabe señalar que con esta metodología no se requiere del muestreo completo del espacio de fases, aunque la calidad de los resultados está directamente relacionada con el tamaño del muestreo realizado.

Combinando la ecuación 2.3.3 con la aproximación del parámetro de acoplamiento obtenemos que:

$$(2.3.6) \quad \Delta F_{BA} = F(\lambda_B) - F(\lambda_A) = -k_B T \ln \frac{Z(\lambda_B)}{Z(\lambda_A)}$$

$$(2.3.7) \quad \Delta F_{BA} = -k_B T \ln \left\langle e^{-[\mathcal{H}(\lambda_B) - \mathcal{H}(\lambda_A)]/k_B T} \right\rangle_{\lambda_A}$$

la cual tiene la forma del promedio del colectivo sobre el estado A. A la ecuación 2.3.7 se le conoce como la fórmula de la perturbación de la energía de Helmholtz (Zwanzig, 1954).

2.3.1. Aproximación de Jarzynski. Recientemente una nueva metodología para calcular diferencias de energía libre ha sido propuesta por Cristhopher Jarzynski, quien derivó una interesante relación entre el promedio del trabajo termodinámico de un proceso fuera del equilibrio y la energía libre. Si se considera un sistema clásico de dimensiones finitas que depende de un parámetro externo λ y que puede intercambiar energía en forma de trabajo con los alrededores. Una vez que se haya dejado que el sistema alcance el equilibrio con un reservorio térmico cambiamos el parámetro de manera infinitamente lenta desde $\lambda = 0$ a $\lambda = 1$. Entonces, el sistema permanecerá en equilibrio cuasiestático con el reservorio a lo largo del proceso y por lo tanto, el trabajo total realizado sobre el sistema será igual a la diferencia de energía libre de Helmholtz entre las configuraciones inicial y final del sistema. Es decir $W_\infty = \Delta F \equiv F^B - F^A$; en donde F_λ corresponde a la energía libre de Helmholtz de equilibrio del sistema a una temperatura T para un determinado valor de λ . El subíndice de W nos dice que esta igualdad es válida cuando la transformación de λ se lleva a cabo infinitamente lenta. Sin embargo, si el sistema es llevado por el mismo camino λ a una velocidad de transformación finita y definida, el trabajo W dependerá en este caso de las condiciones microscópicas del sistema y del reservorio, y éste excederá ΔF (Jarzynski, 1997a):

$$(2.3.8) \quad \bar{W} \geq \Delta F$$

El promedio del trabajo está realizado sobre un colectivo de valores de W irreversible, donde cada medición se realiza una vez que el sistema y el reservorio se encuentran en equilibrio. La desigualdad de la ecuación (2.3.8) se complementa con el trabajo disipado en este proceso $W_{dis} = \bar{W} - \Delta F$. La aproximación de Jarzynski

(Jarzynski, 1997b,a) propone que es posible extraer una propiedad termodinámica estrictamente de equilibrio de un colectivo obtenido a partir de cambios dinámicos de un sistema. En términos de la ecuación de Jarzynski tenemos (2.3.9):

$$(2.3.9) \quad \exp(-\beta\Delta F) = \langle \exp(-\beta W_i) \rangle$$

El lado izquierdo de la ecuación es un exponencial de la diferencia de energía de Helmholtz, mientras que del lado derecho de la igualdad W_i es el trabajo de la i ésima transformación para llevar el sistema del estado A a B. El promedio del exponencial se lleva a cabo sobre un colectivo termodinámico. Esta fórmula realiza una clara conexión entre la distribución de probabilidad de valores del trabajo, con una propiedad de equilibrio como la diferencia de energía Helmholtz entre dos estados.

$$(2.3.10) \quad \Delta F_{A \rightarrow B} = -\frac{1}{\beta} \ln \sum_{i=1}^N \frac{1}{N} \exp(-\beta W_{i,A \rightarrow B})$$

donde $\beta \equiv 1/k_B T$.

Por métodos computacionales se puede simular la transformación de un sistema desde A hasta B a lo largo de una coordenada de reacción a una velocidad finita. A este tipo de simulaciones se les conoce como simulaciones de *slow growth* o crecimiento lento. Por medio de simulaciones de dinámica molecular es posible generar un conjunto de simulaciones de *slow growth*, en donde el sistema se lleva desde el estado inicial A al estado final B, y para cada una de ellas calcular el trabajo W requerido. De tal manera que realizando un número N repeticiones independientes del proceso se obtenga una distribución de valores del trabajo y la ecuación anterior sea aplicable.

La ecuación de Jarzynski representa una novedosa forma de calcular el cambio de energía de Gibbs o de Helmholtz entre dos estados de un sistema. No obstante, se ha de tener cuidado con la precisión con que esta ecuación evalúa la diferencia en energía. Para una distribución extraída de N valores del trabajo, la diferencia de energía obtenida a partir de la ecuación de Jarzynski estará influenciada por el número de repeticiones y por la velocidad de cambio de la coordenada de reacción. Esto produce un error estadístico y otro sistemático al que llamaremos *bias*. En algunos casos esta puede ser la principal fuente del error en la estimación de la diferencia de energía libre y en general la magnitud de este error es desconocida. Gore y cols. 2003 propusieron una manera de evaluar tanto el error sistemático

o bias, como del error estadístico que conlleva la aplicación de esta aproximación (Gore *et al.*, 2003).

2.3.2. Cálculo del error asociado. Dependiendo de las características de la perturbación, las siguientes expresiones pueden utilizarse para estimar la diferencia de energía libre de un conjunto de simulaciones de *slow growth* por medio de simulaciones de dinámica molecular. Basadas en el promedio del trabajo (Ecuación 2.3.11), en el teorema de fluctuación-disipación (Ecuación 2.3.12) o la ecuación de Jarzynski (Ecuación 2.3.13):

$$(2.3.11) \quad \Delta F_{AW} = \langle W \rangle = \frac{1}{N} \sum_{i=1}^N W_i$$

$$(2.3.12) \quad \Delta F_{FD} = \langle W \rangle - \frac{1}{2} \beta \sigma_W^2 = \frac{1}{N} \sum W_i - \frac{\beta}{2} \frac{1}{N-1} \sum_{i=1}^N (W_i - \langle W \rangle_N)^2$$

$$(2.3.13) \quad \Delta F_J = -\frac{1}{\beta} \ln \langle \exp^{-\beta W_i} \rangle_N = -\frac{1}{\beta} \ln \left[\frac{1}{N} \sum_i \exp^{-\beta W_i} \right]$$

donde $\beta = 1/k_B T$, con k_B como la constante de Boltzmann y T la temperatura.

La Ecuación 2.3.13 es válida para cualquier perturbación. ΔF_{AW} y ΔF_{FD} corresponden a los dos primeros términos de una expansión en serie de ΔF_J , y su validez se mantiene únicamente cuando la perturbación está hecha bajo condiciones de reversibilidad (ΔF_W) o cercanas al equilibrio (ΔF_{FD}). Bajo condiciones de no equilibrio la diferencia entre el trabajo promedio y la diferencia de energía de Helmholtz es igual al trabajo disipado (W_{dis}).

Cuando la distribución de trabajo que se genera adquiere una forma gaussiana $P(W)$, la distribución del trabajo disipado es una gaussiana también, con promedio:

$$(2.3.14) \quad \langle W_{dis} \rangle = \frac{1}{2} \beta \sigma_W^2$$

y varianza ,

$$(2.3.15) \quad \sigma_{W_{dis}}^2 = \sigma_W^2 = \frac{2}{\beta} \langle W_{dis} \rangle$$

Cuando una transformación se realiza cerca del equilibrio generará entonces una distribución gaussiana. Sin embargo, es necesario remarcar que una distribución gaussiana no siempre representa a una perturbación realizada cercana al equilibrio, a pesar de que lo contrario es cierto.

La aproximación de Jarzynski puede escribirse también en términos de W_{dis} como:

$$(2.3.16) \quad \langle \exp^{-\beta W_{dis}} = 1 \rangle$$

De la Ecuación 2.3.16 se puede deducir que las simulaciones con valores negativos de W_{dis} han de ser muestreadas con mayor intensidad para que la igualdad se cumpla. Se ha demostrado también que la probabilidad de observar una trayectoria con $W_{dis} < 0$ decrece con el incremento de $\langle W_{dis} \rangle$. Por lo tanto, a mayor $\langle W_{dis} \rangle$, mayor será también el número N de trayectorias que deberán obtenerse para que el valor de ΔF_J se aproxime más a ΔF . Gore y cols. propusieron tres propiedades para caracterizar ΔF , que son: (i) el *bias* o error sistemático que mide la diferencia entre el valor esperado y el valor real de ΔF ; (ii) la varianza: que es un estimador del error estadístico; y finalmente (iii) el error cuadrático medio (MSE) que se define por la suma del *bias* y la varianza.

Como el *bias* es causado por el número finito de repeticiones realizadas para generar la distribución, dependerá de las N repeticiones realizadas. La frontera entre un valor de N pequeño y/o grande se encuentra de la siguiente manera (Gore *et al.*, 2003):

$$(2.3.17) \quad N_C = C_b (\exp^{2\beta W_{dis}-1})$$

donde $C_b = 15$ es un parámetro fijo empleado para determinar el *bias* entre límite entre N pequeño y N grande: $B_J(N_C) = \frac{1}{2\beta C_b} = 0,08 \text{ kJmol}^{-1}$.

Adicionalmente, los mismos autores derivaron expresiones para calcular el *bias*, la varianza y el error cuadrático medio, que resultan de la utilización de la ecuación de Jarzynski ΔF_J . En el caso concreto cuando N es pequeño y las simulaciones se realizan cerca del equilibrio, se tiene que:

$$(2.3.18) \quad B_J = \frac{\langle W_{dis} \rangle}{N^\alpha}$$

donde α es función de $\langle W_{dis} \rangle$ y está dada por:

$$(2.3.19) \quad \alpha = \frac{\ln [2\beta \langle W_{dis} \rangle C]}{\ln (N_C)}$$

la varianza :

$$(2.3.20) \quad \sigma_J^2 \leq \frac{2B_J}{\beta}$$

y el error cuadrático medio:

$$(2.3.21) \quad MSE = B_J^2 + \sigma_J^2 \approx B_J^2$$

para obtener finalmente la corrección a la ecuación de Jarzynski:

$$(2.3.22) \quad \Delta F_{JC} = \Delta F_J - B_J$$

Además, propusieron la siguientes relaciones que constituyen otra forma de calcular el error:

$$(2.3.23) \quad \langle W_{dis,1} \rangle = \langle W \rangle - \Delta F_J$$

$$(2.3.24) \quad \langle W_{dis,2} \rangle = \langle W_{dis,1} \rangle + B_{J,1}$$

obteniendose así:

$$(2.3.25) \quad \Delta F_{Jc,1} = \Delta F_J - B_{J,1}$$

$$(2.3.26) \quad \Delta F_{Jc,2} = \Delta F_J - B_{J,2}$$

en donde $B_{J,1}$ y $B_{J,2}$ se refieren al *bias* calculado utilizando $\langle W_{dis,1} \rangle$ y $\langle W_{dis,2} \rangle$, respectivamente.

Parte 1

Estudio de la triosfosfato isomerasa de
Trypanosoma cruzi en mezclas agua/decano.

Un estudio de dinámica molecular.

CAPÍTULO 3

Introducción

La existencia de una correlación entre rigidez, estabilidad térmica y actividad catalítica de enzimas ha sido propuesta por varios autores (Tuena de Gomez-Puyou & Gomez-Puyou, 1998). Esta correlación implica que cuanto menor es la temperatura a la cual se despliega la estructura mayor es la flexibilidad y la actividad catalítica. En este sentido, parece que a lo largo de la evolución las proteínas han sorteado un balance entre flexibilidad y estabilidad térmica que las provee de su habilidad catalítica, pero que al mismo tiempo les permite preservar una estructura definida y estable. En este contexto el agua juega un papel muy importante (Finney & Poole, 1984; Levy & Onuchic, 2004), en particular se sabe que el comportamiento dinámico de aquellas moléculas que se encuentran en contacto directo con el soluto presentan diferencias tanto en su organización como en sus propiedades termodinámicas y dieléctricas si se comparan con las que se encuentran en el seno de la disolución (Pal & Zewail, 2004; Rupley & Careri, 1991; Rupley *et al.*, 1983). Además, el comportamiento de las moléculas de agua se puede ver afectado por la presencia de co-solventes, y por tanto termina afectando el comportamiento dinámico de las macromoléculas.

A mediados de la década de los 80's Zaks y Klibanov observaron que la actividad catalítica que presentaban algunas enzimas cuando estaban suspendidas en solventes hidrofóbicos a bajas concentraciones de agua era comparable e incluso mayor que en soluciones acuosas (Zaks & Klibanov, 1984, 1985, 1986, 1988a,b). En base a sus resultados, llegaron a la conclusión de que bajo estas condiciones cuanto más hidrofóbico era el solvente, menos agua necesitaba la proteína para mantener su actividad (Zaks & Klibanov, 1988a). Además, los mismos autores sugirieron que el agua actúa como un lubricante formando puentes de Hidrógeno con la propia enzima y facilitándole la flexibilidad que requiere para mantener su actividad catalítica. Dichas explicaciones son generalmente aceptadas, a pesar de que no se cuenta con evidencia directa del fenómeno a nivel experimental. En la práctica se ha sacado provecho de este fenómeno, principalmente en el área de biotecnología en donde se han propuesto diversas aplicaciones basadas en la utilización de proteínas en bajas concentraciones de agua o en micelas invertidas (Gonnelli & Strambini, 1988; Strambini & Gonnelli,

1988; Dordick, 1988; Bell *et al.*, 1995; Klibanov, 1997, 1989). A la fecha, ninguna técnica experimental ha podido proporcionar en detalle una descripción estructural y dinámica del proceso. En este sentido se han realizado algunos trabajos de simulación molecular para el estudio de proteínas en mezclas de agua y solventes orgánicos (Soares *et al.*, 2003; Micaelo *et al.*, 2005; Colombo *et al.*, 1999; Norin *et al.*, 1994; Peters *et al.*, 1996; Toba *et al.*, 1996; Zheng & Ornstein, 1996a; Ornstein & Ya-Jun, 1996; Yang *et al.*, 2004; Micaelo & Soares, 2007; Zheng & Ornstein, 1996b). Sin embargo, las escalas de simulación reportadas (7 *ns* la trayectoria más larga) no son suficiente para observar claras diferencias dinámicas o estructurales del sistema bajo esas condiciones. En 2003, Soares y cols. (Soares *et al.*, 2003) reportaron el primer estudio sistemático de dos proteínas (ubiquitina y cutinasa) en hexano a diversas condiciones de hidratación. A partir de simulaciones de entre 2 y 4 *ns* estos autores reportaron correlaciones entre RMSD, RMSF y el porcentaje de agua en la proteína. Lamentablemente, es prácticamente imposible observar cambios estructurales importantes a estas escalas de tiempo de simulación ya que los movimientos ocurren principalmente en las cadenas laterales de las proteínas. Yang y cols. (Yang *et al.*, 2004) realizaron simulaciones con la proteína subtilisina BPN' solubilizada en presencia de tensoactivos disueltos en octano, tetrahidrofurano y acetonitrilo, a dos distintos niveles de hidratación y en una escala de tiempo de 4 *ns*. En este trabajo no reportan grandes cambios estructurales o dinámicos de la proteína como efecto del disolvente. Estos autores basaron su trabajo en la partición de las moléculas de agua entre la superficie de la proteína y la fase orgánica, además de la hidratación del sitio activo. Recientemente, Micaelo y Soares (Micaelo & Soares, 2007) publicaron un estudio del mecanismo de hidratación de la cutinasa en hexano, di-isopropil eter, 3 pentanona, etanol y acetonitrilo con distintos niveles de hidratación en una escala de tiempo de 7 *ns*. En este trabajo, se observaron algunos cambios estructurales en la proteína, sin embargo no se profundizó en el análisis de su comportamiento dinámico. Principalmente, estuvo basado en el comportamiento del agua. Estos estudios demuestran la falta de simulaciones de dinámica molecular a mayores escalas de tiempo, las cuales pueden proporcionar mayor información de la estructura y dinámica de las enzimas en bajas concentraciones de agua, así como del papel del solvente en la función de la proteína.

En este trabajo se presenta un estudio de simulaciones de dinámica molecular de la trifosfato isomerasa de *Trypanosoma cruzi* (TcTIM) en varias mezclas agua/decano. El tiempo de simulación para cada sistema fue de 40 *ns*. Como referencia se realizaron dos simulaciones de la enzima en decano puro y agua pura, en

el caso de la simulación en agua pura, la trayectoria se llevó hasta los 140 *ns* de simulación.

En este trabajo se intenta reducir el complejo espacio configuracional para extraer los componentes esenciales del movimiento que describen los cambios estructurales en la proteína en medio acuoso y a bajas concentraciones de agua en mezclas agua/decano. En general, se busca identificar aquellos movimientos internos que puedan estar relacionados con su función y su actividad catalítica.

La triosafosfato isomerasa es una enzima homodimérica con 251 residuos por cada subunidad. Cada subunidad se pliega en un barril α/β con ocho hebras formando el barril central y rodeado por ocho hélices alfa (Maldonado *et al.*, 1998). Esta enzima se eligió porque tiempo atrás se probó experimentalmente que TIMs de otras especies mostraron actividad catalítica a bajas concentraciones de agua. Además, esta enzima es una proteína blanco en el diseño de fármacos efectivos contra el mal de Chagas (Maldonado *et al.*, 1998; Garza-Ramos *et al.*, 1992; Gao *et al.*, 1999; Garza-Ramos *et al.*, 1998; Perez-Montfort *et al.*, 1999; Reyes-Vivas *et al.*, 2001; Tellez-Valencia *et al.*, 2002, 2004).

En las páginas siguientes se presenta un estudio detallado de la estructura y dinámica de la TcTIM, así como del comportamiento de las moléculas del disolvente realizado en distintas condiciones de hidratación de la proteína. El principal objetivo de este trabajo consiste en estudiar diferencias en el comportamiento dinámico de la proteína y el solvente en función de la concentración que puedan explicar a nivel atómico la razón por la cual algunas enzimas son catalíticamente activas en solventes orgánicos. Adicionalmente, se busca estudiar a fondo el comportamiento dinámico de la TcTIM en agua pura.

CAPÍTULO 4

Metodología

4.1. Construcción de los sistemas

Se construyeron seis cajas de simulación con la TcTIM en 3 distintas mezclas formadas por moléculas de agua y decano, además de una caja conteniendo la proteína en agua pura y otra diferente en decano puro. La selección de decano como solvente secundario se debió a su baja hidrofobicidad, dado que en cada sistema se buscaba mantener capas de agua de distinto espesor en torno a la enzima. Puesto que la partición de las moléculas de agua en el decano es muy baja, con este solvente se asegura entonces que de esta manera el agua permanezca rodeando la proteína. A cada una de las trayectorias generadas se les denominó RW_a, RW_b, W8D, W6D, W4D y RD. El sistema RW_a se construyó colocando a la TcTIM en una caja de simulación dodecaédrica y posteriormente se solvató con 24,825 moléculas de agua previamente equilibradas. La estructura cristalina de la TcTIM con la cual se construyó el sistema fue tomada del Protein Data Bank (PDB ID 1TCD). El tamaño de la caja de simulación fue elegido tal que la distancia entre la proteína y las paredes de la caja de simulación fuese mayor o igual a 2 nm. El sistema RW_a fue simulado durante 20 ns. A partir de la última configuración de esta simulación que se llamó RW_a(20) se prepararon los sistemas RW_b, W8D, W6D, W4D. En primer lugar, la simulación RW_a(20) se extendió durante 40 ns de simulación más, dando lugar a la simulación RW_b. Para la construcción de los sistemas W8D, W6D, W4D se utilizó la estructura RW_a(20) como estructura inicial. Para obtener W8D se eliminaron todas las moléculas de agua que estuviesen a una distancia mayor a 8 Å de cualquier átomo de la proteína, la estructura resultante se colocó dentro de una caja de simulación dodecaédrica y solvatada con moléculas de decano. El mismo procedimiento se repitió para construir los sistemas W6D y W4D, aunque en estos casos el radio de corte o espesor de la capa de agua fue de 6 Å y 4 Å, respectivamente. Este procedimiento asegura que tanto la proteína como el agua en su entorno se encuentren en equilibrio al inicio de la simulación y que la estructura inicial sea la misma en los 3 sistemas. La metodología para construir el sistema RD es similar al RW solo que se utilizó decano puro como único disolvente. El número final de moléculas de cada componente del

sistema se muestra a detalle en la Tabla 5.6.2. Los sistemas W8D, W6D, W4D y RD fueron simulados durante 40 *ns*.

La trayectoria RW_b fue extendida hasta llegar a los 140 *ns* de simulación. Así mismo se generaron dos réplicas de estas trayectorias con las mismas condiciones pero modificando la estructura inicial.

4.2. Parámetros de simulación

Todas las simulaciones se realizaron utilizando el paquete computacional GROMACS versión 3.2.1 (Berendsen *et al.*, 1995; www.gromacs.org, GROMACS; Lindahl *et al.*, 2001; Van Der Spoel *et al.*, 2005) con el campo de fuerza GROMOS96 (43a2) (Schuler & Gunsteren, 2000; Daura *et al.*, 1998) tanto para simular las interacciones en la proteína como en el decano. Para describir las interacciones entre las moléculas de agua se utilizó el modelo de agua SPC (por sus siglas en inglés de Simple Point Charge) que define un modelo de cargas puntuales (Berendsen *et al.*, 1981). En todos los casos se utilizaron condiciones periódicas a la frontera con cajas de simulación de geometría dodecaédrica. En estas simulaciones se hizo uso del colectivo termodinámico NPT. A lo largo de la simulación la presión se mantuvo a 1 bar y la temperatura a 298 K. Para mantener la presión y la temperatura dentro del sistema se utilizó un barostato y un termostato de Berendsen (Berendsen *et al.*, 1984) con unas constantes de acoplamiento de 0.5 ps y 0.1 ps, respectivamente. Las interacciones de largo alcance se calcularon utilizando un radio de corte dual de 0.9 *nm* y 1.4 *nm*. Las interacciones dentro del radio de corte interno se evaluaron a cada paso de tiempo y las interacciones comprendidas entre las distancias del radio de corte a cada 5 pasos de tiempo. Más allá del radio de corte se consideró un medio continuo. La constante dieléctrica empleada fue de 78 ϵ en todas las simulaciones, incluyendo aquellas en las que el disolvente principal es el decano. Puesto que las interacciones mas allá del radio de corte no nos interesan consideramos seguro utilizar la misma constante dieléctrica en todos los casos. Para evitar interacciones desfavorables antes de cada simulación se realizó una minimización de energía mediante el método *SteepestDescent*. Las velocidades iniciales se asignaron aleatoriamente en base a una distribución de Maxwell-Boltzmann a 298 K. Las ecuaciones de movimiento se integran utilizando el método *leap-frog* con un paso de tiempo de 2 fs. Tanto los ángulos como las distancias de enlace entre moléculas de agua se restringieron por el algoritmo SETTLE (Miyamoto & Kollman, 1992). Por otro lado, para mantener las distancias de los enlaces dentro de la proteína y el decano se utilizó el algoritmo

LINCS (Hess *et al.*, 1997). A lo largo de cada simulación coordenadas, velocidades y energías se almacenaron cada 10 *ps* para su posterior análisis.

4.3. Métodos de análisis

En primer lugar se calculó el RMSD de la proteína en cada trayectoria utilizando como estructura de referencia en cada caso la estructura cristalina de la proteína. La movilidad de cada residuo se calculó en términos del RMSF. Para este cálculo se utilizaron los últimos 20 *ns* de simulación de las trayectorias RW_b, W8D, W6D, W4D y RD. La estructura de referencia para estos casos fue también la estructura cristalina. Como a cada paso de tiempo se generó una configuración del sistema, se realizó un tamizaje con dichas configuraciones, al cual se le llamó análisis de clusters. El método empleado se describe en el Capítulo 2 de la presente memoria. El número de configuraciones utilizadas para este análisis fue de 2,000, extraídas de los últimos 20 *ns* de simulación de cada trayectoria. El parámetro de corte para el RMSD fue de 0.15 para las trayectorias RW_b, W8D, W6D y W4D y de 0.10 para RD. A lo largo de las simulaciones se calculó el radio de Giro, el número de puentes de Hidrógeno intramoleculares, la distancias entre moléculas y el área hidrofóbica e hidrofílica accesible al disolvente.

Parte del análisis realizado requirió del uso de herramientas bioinformáticas tales como los visualizadores RASMOL 2.7, VMD 1.8.2 (Humphrey *et al.*, 1996) y PyMOL 0.99 y los graficadores Xmgrace y GNUPLOT. Sin embargo, también fue necesario la elaboración de material propio de análisis utilizando lenguajes de programación como awk, perl, bash, entre otros. En el cálculo de RMSD, RMSF, *clusters*, puentes de Hidrógeno, radio de giro, SAS y distancias entre moléculas se utilizaron herramientas de GROMACS.

Análisis de componentes principales . La dinámica interna de la TcTIM se analizó extrayendo sus componentes principales de las distintas trayectorias. Para cada una de ellas, las distintas configuraciones se ajustaron a una estructura de referencia para eliminar movimientos de traslación y rotación de la proteína. Posteriormente, esta trayectoria fue utilizada para obtener la matriz de covarianza de posiciones y después diagonalizarla para extraer los componentes principales del movimiento. En el cálculo únicamente se tomaron en cuenta los átomos correspondientes a la cadena central de la proteína. De esta manera se pretende encontrar sólo los movimientos de mayor amplitud que tienen lugar en la cadena central eliminando así

los de alta frecuencia en la cadena lateral. Nuevamente, el cálculo se realizó con las herramientas del paquete computacional GROMACS.

CAPÍTULO 5

Resultados y discusión

5.1. Análisis visual de las trayectorias

Después de analizar visualmente todas las trayectorias se observó que la estructura dimérica de la TcTIM permaneció unida en todos los casos manteniendo su estructura terciaria y su estructura secundaria. Sin embargo, es evidente que la estructura de la TcTIM en la simulación RD muestra un tamaño más pequeño que las otras conformaciones finales. Además, según se observa en la Figura 5.1.1 hubo una mayor pérdida de estructura secundaria en algunas zonas de la proteína.

En las simulaciones W8D, W6D y W4D se puede observar un desplazamiento de las moléculas de agua entre las configuraciones inicial y final, cambiando la distribución de las moléculas de agua que se encontraba al inicio de la simulación y formándose una acumulación en la zona hidrofóbica cercana a la intercara del dímero. Lo anterior ocurrió en las tres simulaciones en las que se utilizó una mezcla agua/decano como disolvente, lo cual parece indicar que este hecho no es accidental, sino que sugiere la existencia de un fenómeno subyacente que ocasione esta acumulación. Dado que el disolvente decano es altamente hidrofóbico y que la geometría de la enzima permite un punto de nucleación, es posible que ambos factores hayan colaborado a la formación de una burbuja de agua en este sitio específico. Además, para que esto ocurriera es posible que los propios movimientos de la enzima hayan favorecido esta nucleación moviendo moléculas de agua hacia dicha región. El área que comprende la interfase y el área entre los dos monómeros de la enzima es concebida como una zona de alta hidrofobicidad, pero de ninguna manera el ambiente en esta zona es más hidrofóbico que el solvente decano. Por esta razón es que el agua que se acumula en esta cavidad hidrofóbica permite la minimización del área de las moléculas de agua expuestas al decano. La presencia de regiones donde las moléculas de agua forman un núcleo y se acumulan se observa también en otras áreas de la superficie de la proteína aunque en menor medida.

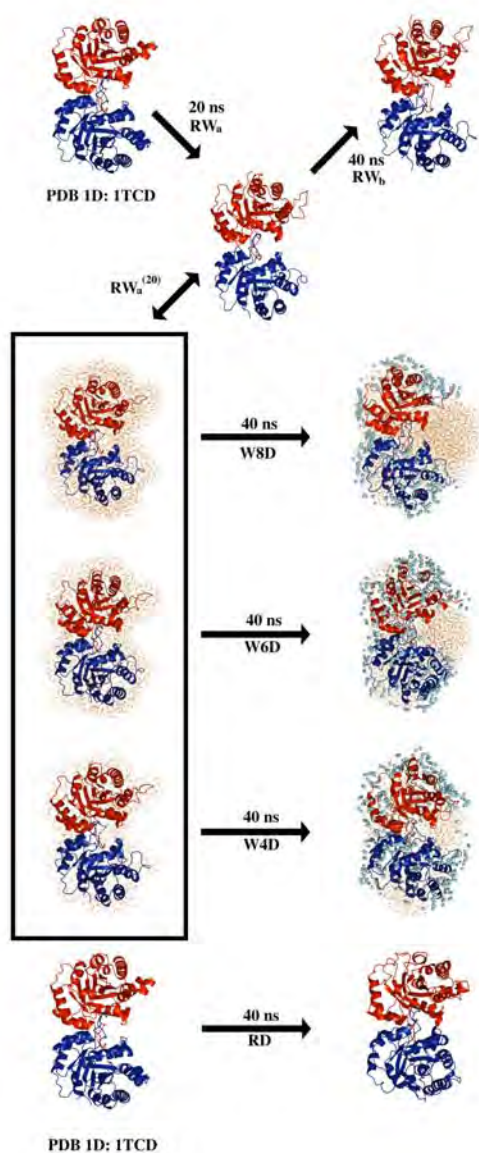


FIGURA 5.1.1. Configuraciones inicial y final de la TcTIM en cada una de la trayectoria. Las escalas de tiempo y los nombres de cada sistema se encuentran indicados mediante etiquetas en las flechas y en las estructuras. Para W8D, W6D y W4D se muestran todas las moléculas de agua (naranja) y los átomos de Carbono (esferas verdes) que se encontraran a una distancia menor a 4 \AA . Las subunidades A y B de la proteína están coloreadas en azul y rojo, respectivamente. La flecha verde en la conformación RD se encuentra apuntando a la región de la enzima que más pérdida de estructura secundaria experimentó.

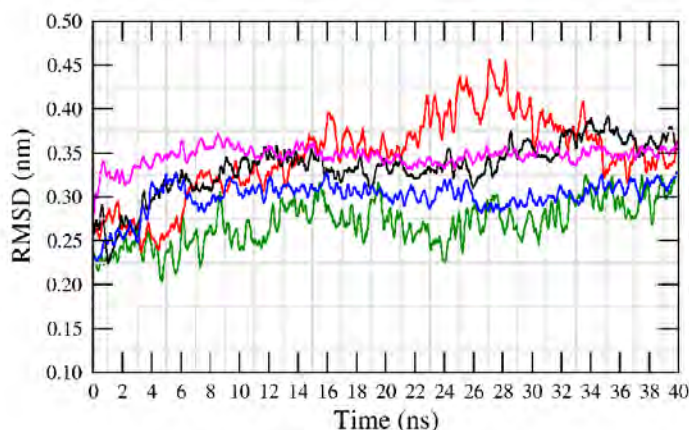


FIGURA 5.2.1. Desviaciones cuadráticas medias (RMSD) de la TcTIM tomando en cuenta únicamente los átomos de la cadena central. RW_b (verde), W8D (rojo), W6D (negro), W4D (azul), y RD (magenta). Como referencia en todos los casos se utilizó la estructura cristalina (PDB ID 1TCD).

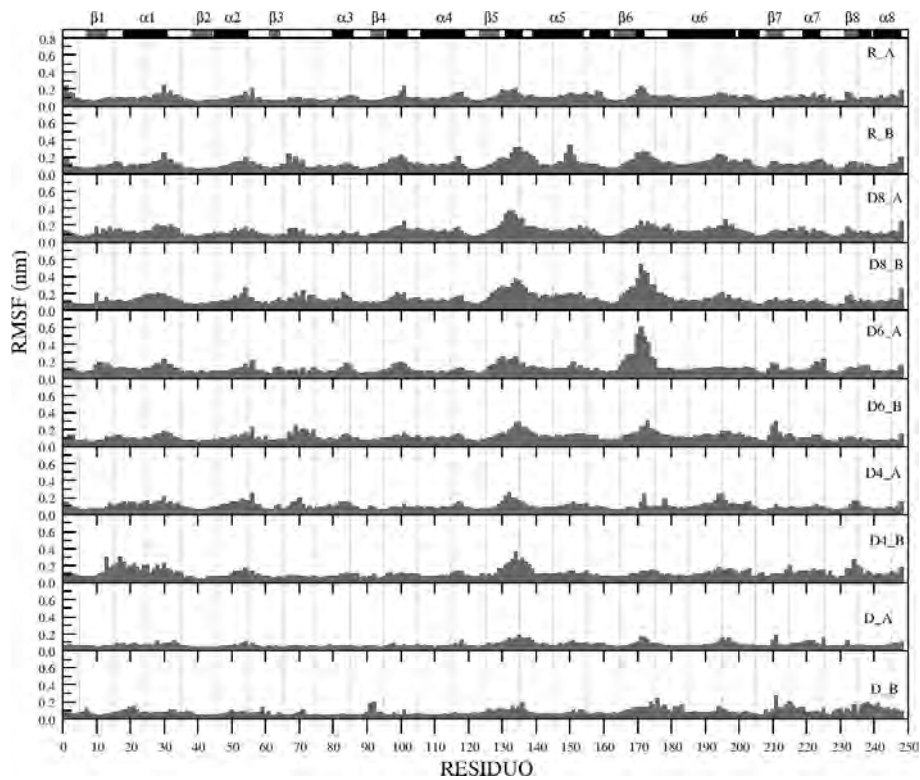
5.2. Análisis de la dinámica y estructura de la TcTIM

5.2.1. Dinámica global de la proteína en las trayectorias. En la Figura 5.2.1 se muestra una gráfica con los valores de RMSD de la proteína a lo largo de cada trayectoria. En general, a partir de los primeros nanosegundos el RMSD de cada una de las trayectorias mostró ligeras fluctuaciones en torno a un valor promedio constante, es decir, no se observan algunas tendencias claras que sugieran inestabilidad en el sistema o en la proteína. Adicionalmente, durante los primeros nanosegundos en las simulaciones W8D, W6D y W4D se observa un pequeño incremento en el valor de RMSD, el cual puede ser ocasionado por el reagrupamiento del disolvente alrededor de la proteína. En el caso de la simulación W8D el RMSD presenta un incremento significativo alrededor de los 22 *ns* seguido más adelante de una recuperación del valor de RMSD alrededor de los 34 *ns*. Debido a la extensión de las simulaciones en comparación con trabajos previos se esperan fluctuaciones importantes en la estructura, debido a que en estas escalas la proteína puede experimentar más cambios conformacionales incluyéndolos aquellos dentro de la cadena central y no sólo movimientos de baja frecuencia. La línea de color magenta que corresponde a la simulación en decano presenta el mismo comportamiento que las otras trayectorias, sin embargo, es evidente que las fluctuaciones que sufre la proteína parecen ser de menor intensidad. Dado que el RMSD es una magnitud que no proporciona información específica del comportamiento dinámico de la proteína a nivel local, se procedió a estudiar en

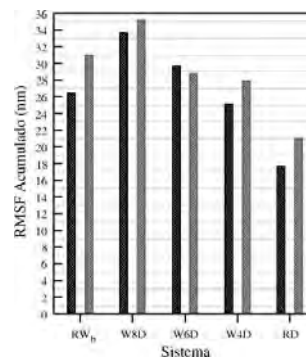
detalle las partes de la proteína que contribuyeron a los cambios en RMSD en general y de la movilidad de cada parte de la proteína. En base a estas gráficas se decidió utilizar para el análisis los últimos 30 *ns* de la simulación para el análisis que se presenta en la Tabla 5.6.2.

5.2.2. Análisis de la movilidad local de la proteína. Para estudiar de manera global la movilidad de la proteína y de cada uno de los residuos que la componen se realizó un estudio de las fluctuaciones cuadráticas medias (RMSF) por residuo considerando los últimos 20 *ns* de simulación de las trayectorias RW_b, W8D, W6D, W4D y RD (Figura 5.2.2 (a)). Dado que la proteína es homodimérica, es decir que las subunidades son idénticas en la gráfica se presentan los valores alineados en función de la numeración de los residuos en cada subunidad. Para obtener una medida general de la movilidad de cada monómero en cada trayectoria se calculó también el valor acumulado de RMSF para cada monómero, los resultados se presentan en la Figura 5.2.2 (b). De la Figura 5.2.2 (a) se puede observar que las zonas de mayor fluctuación de la proteína en cada trayectoria están localizadas en los lazos 5 (residuos 129-139) y 6 (residuos 169-179) principalmente. Este máximo en el valor de RMSF corresponde a las mismas zonas en las cinco trayectorias. En la misma Figura también se observa que el perfil de valores de RMSF es similar en todas las simulaciones así como entre monómeros. En general, los movimientos del lazo 5 parecen tener mayores fluctuaciones. Esto es razonable dado que esta región de la proteína se encuentra en contacto con el disolvente en todo momento, por lo que tiene mayor libertad de movimiento. Las hebras β que forman parte del barril central de la estructura tridimensional de la TcTIM mostraron valores bajos de RMSF cuando se compara con los correspondientes valores de RMSF en las hélices α o lazos.

En este punto llegamos a un detalle muy importante, los movimientos del lazo 6 han sido relacionados con la actividad catalítica de la enzima, dado que la entrada al sitio activo se expone y se oculta alternadamente debido al movimiento de este lazo. Se cree que de esta manera se regula el paso del sustrato al sitio activo. Se ha demostrado que este movimiento tiene lugar con un periodo entre 200-400 μ s (Rozovsky *et al.*, 2001; Rozovsky & McDermott, 2001). Para observar este movimiento por medio de simulaciones de dinámica molecular probablemente se hubiesen requerido simulaciones en escalas de los microsegundos, pero cuando se trabaja con sistemas de gran tamaño como esta proteína alcanzar dichas escalas de tiempo de simulación resulta complicado y computacionalmente muy costoso hoy en día. Sin embargo, a partir de nuestras simulaciones se observa claramente una mayor movilidad en los



(a)



(b)

FIGURA 5.2.2. 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 RW_b, W8D, W6D, W4D y RD 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.

residuos del lazo 6, además, visualmente el lazo 6 presenta cambios conformacionales que evidencian la importancia del comportamiento dinámico de esta zona. Dada la importancia de éste movimiento, más adelante se hace un análisis a fondo de este lazo.

A pesar de que no se observen grandes diferencias entre el valor acumulado de RMSF en las diferentes trayectorias, es evidente que el valor máximo corresponde a la trayectoria W8D y el mínimo a la trayectoria RD. Si se realiza una comparación entre subunidades en una misma trayectoria, en la trayectoria W6D se observa que la subunidad B presenta un RMSF acumulado ligeramente mayor a la subunidad A. En esta misma trayectoria la diferencia en la contribución de los movimientos del lazo 6 al RMSF es evidente.

5.3. Componentes principales del movimiento

Para entender a fondo el comportamiento dinámico de la proteína en cada trayectoria, se calcularon los modos principales del movimiento (ver sección 4.3). Las amplitudes de los modos obtenidos se ordenaron de manera decreciente. Dado que los valores decaen exponencialmente en la Figura 5.3 (a) se presentan únicamente los veinte primeros autovalores, considerando el resto despreciables. La gráfica permite la comparación de los autovalores correspondientes a cada trayectoria. De aquí se puede observar claramente que en presencia de agua y decano la movilidad de la enzima resulta ser mayor que en cualquiera de los dos disolventes puros. Además, a mayor cantidad de agua en presencia de decano la amplitud de los movimientos fueron más grandes. Notablemente, se observa que los autovalores obtenidos en agua pura son comparables a aquellos de la proteína en el sistema W4D, simulación con la menor cantidad de agua. Los autovalores con menores amplitudes se obtienen cuando la proteína se simuló en decano puro, lo cual sugiere una movilidad más restringida en este medio altamente hidrofóbico. Los resultados de esta gráfica concuerdan con los valores acumulados del RMSF mostrados en la Figura 5.2.2 (a). No obstante a que los resultados coincidan, el análisis de PCA es más fino que un simple análisis de RMSF. Se ha propuesto que las zonas de mayor fluctuaciones en una proteína coinciden con la zonas más vulnerables a la pérdida de estructura en un proceso de desplegamiento Daggett (2006). Así mismo, la rigidez de la proteína cuando se encuentra en decano puro puede estar relacionada con la alta estabilidad térmica usualmente encontrada de las proteínas en solventes orgánicos (Tuena de Gomez-Puyou & Gomez-Puyou, 1998). Al igual que los resultados de RMSF, el análisis de componentes principales

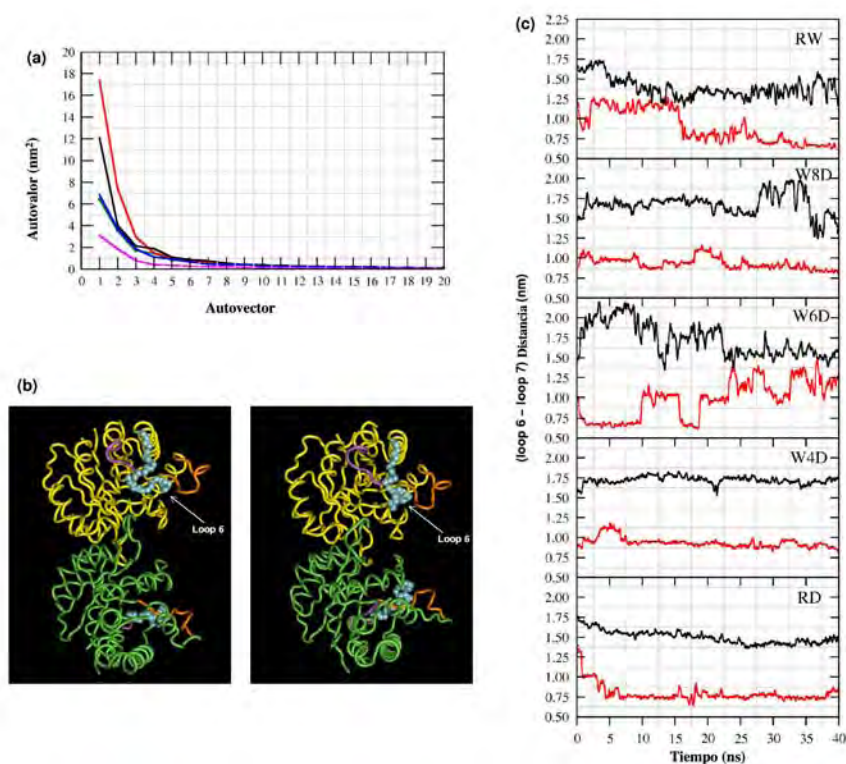


FIGURA 5.3.1. (a) Primeros 20 autovalores obtenidos por medio del análisis de componentes principales del movimiento en las simulaciones RW_b, W8D, W6D, W4D y RD en verde, rojo, negro, azul y magenta, respectivamente. (b) Representaciones de los extremos del movimiento principal de la TcTIM extraído de la simulación RW_b. Únicamente se muestra la cadena central de la estructura. Las subunidades A y B se muestran en amarillo y verde, respectivamente. Los residuos de lazos 5, 6 y 7 se muestran en color naranja, azul y magenta, respectivamente. (c) Distancia entre centros de masas de los residuos Gly70-Thr171-Glu172 (lazo 6) y el residuo Val212 (lazo 7), los resultados se presentan para las subunidades A (negro) y B (Rojo) en función del tiempo de la simulación. El nombre de la simulación se indica en la esquina superior de cada gráfica.

de nuevo señala a los lazos 5 y 6 como regiones de alta movilidad. Dado que durante 40 ns de simulación en agua pura se observaron cambios estructurales en zonas relevantes de la proteína, tales como el lazo 6, la trayectoria de la proteína en agua se extendió 100 ns más para profundizar en el análisis dinámico y estructural de la proteína.

5.4. Análisis estructural del lazo 6

Como se mencionó, la dinámica del lazo 6 ha sido relacionada con el mecanismo a través del cual se lleva a cabo la reacción catalítica. Algunos estudios experimentales han identificado un movimiento de apertura y cierre de este lazo capaz de regular el paso del sustrato hacia los residuos catalíticos en la TcTIM. Este tipo de movimientos no son nuevos, en algunos casos incluso se ha logrado caracterizar experimentalmente el periodo de tiempo con que ocurren. Recientemente, estos movimientos en algunas proteínas han sido reproducidos por medio de simulaciones de dinámica molecular en tiempos de hasta 42 *ns* (Hornak *et al.*, 2006b,a).

A lo largo de la trayectoria de 140 *ns* de simulación de la TcTIM en agua pura, partiendo de las coordenadas de la estructura cristalina (PDB 1TCD), se recorrió el camino que conecta las conformaciones abierta y cerrada del lazo 6, tal como está reportado en la literatura. En la Figura 5.4.1 se muestran, por un lado las conformaciones (a) y (b) que corresponden a las estructuras abierta (PDB 1TCD) y cerrada (5TIM) de la TIM, extraídas del PDB (www.pdb.org, PDB). Y por otro lado, dos estructuras comparables a las primeras se muestran en (c) y (d), las cuales fueron tomadas de la trayectoria de simulación a los nanosegundos 22 y 67, respectivamente. Claramente se puede observar al lazo 6 en la conformación abierta en (a) y (c) y cerrada en (b) y (d). Al inicio de la simulación el lazo 6 en el monómero B se encontraba en una conformación abierta, a lo largo de la simulación, como se puede observar en la Figura 5.4.1 (e) la línea negra que representa la distancia entre los residuos del lazo 6 y los del lazo 7, muestra varias fluctuaciones, entre lo que sería una estructura abierta (nanosegundo 22), cerrada (nanosegundo 67), o bien, una estructura semi-abierta del lazo 6 (nanosegundo 56). Para el monómero A se observa un comportamiento similar, inicialmente se parte de una estructura del lazo 6 cerrada pero a lo largo de la simulación visita una conformación semi-abierta, al igual que el monómero B. Al término de la trayectoria el lazo 6 de ambas subunidades parece estabilizarse en una conformación semi-abierta, lo cual podría ser la estructura estable en solución de la TcTIM.

5.5. Movilidad de los residuos catalíticos

Para cada sistema se analizó la movilidad de los residuos del sitio activo a lo largo de las correspondientes trayectorias. Se aplicó la metodología de análisis de clusters de estructuras, que se describe con detalle en la sección 2.2.4. Con este análisis se busca describir tanto la movilidad de los residuos catalíticos, como encontrar las

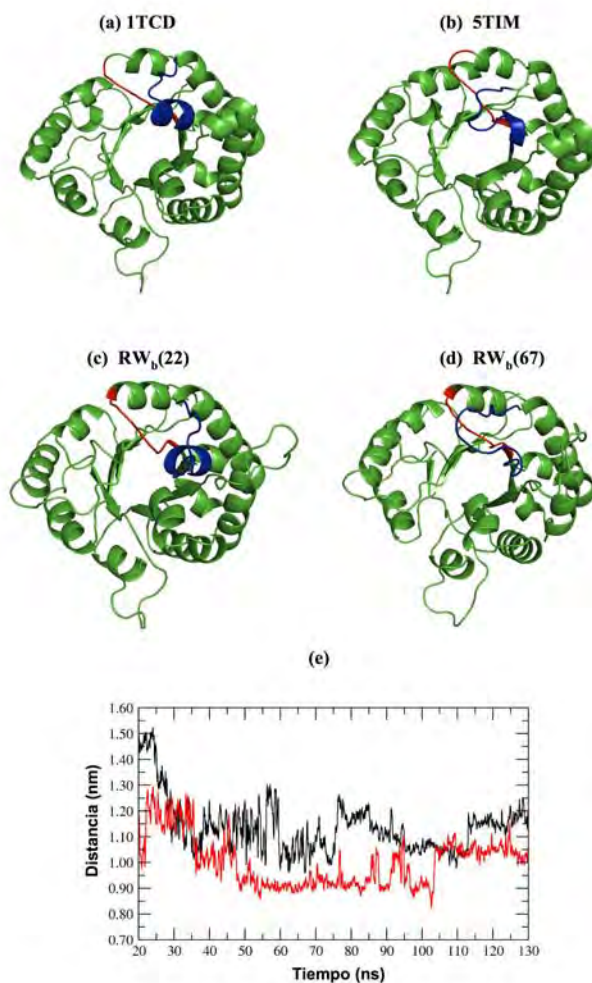


FIGURA 5.4.1. Comparación de las estructuras cristalinas mostrando conformaciones del lazo 6 abierta (a) y cerrada (b), con imágenes obtenidas por medio de simulaciones de dinámica molecular ((c) y (d)) de la trayectoria RW_b. Todas las imágenes son del monómero B. Azul: lazo 6; rojo: lazo 7. En la subfigura (e) se presenta la distancia entre los residuos que conforman al lazo 6 y el lazo 7 a lo largo de la simulación. La distancia fue calculada tanto en el monómero A (línea negra) como en el B (línea roja).

conformaciones más visitadas a lo largo de cada trayectoria. En la Figura 5.5.1 se muestra una gráfica con las estructuras más visitadas por los residuos catalíticos en la trayectoria RW_b, como se observa, los residuos catalíticos Lys14, His96 y Glu168 a lo largo de la simulación presentan bastante movilidad. La conformación en rojo corresponde a la más visitada a lo largo de la trayectoria. Es evidente que a lo largo

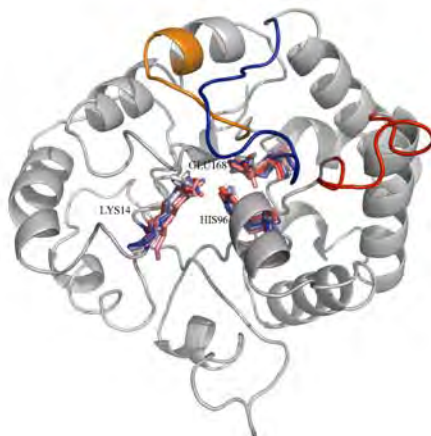


FIGURA 5.5.1. Representación de la movilidad de los residuos LyS14, His96 y Glu168 que pertenecen al sitio activo en la subunidad A. En un gradiente de rojo a azul se muestran las distintas conformaciones recorridas durante la simulación. La conformación en rojo sería la más visitada. Cada conformación es la estructura representativa de cada uno de los cluster de estructuras. Los lazos 5, 6 y 7 se muestran en color rojo, azul y naranja, respectivamente.

de la simulación los residuos visitan distintas conformaciones, lo cual nos habla de una estructura flexible de la TcTIM en agua pura. Este mismo análisis fue realizado en los sistemas con agua/decano y decano puro. En general, en las trayectorias RW_b, W8D y W6D se observó mayor movilidad en los residuos catalíticos, mientras que en RD y W4D ésta disminuye sustancialmente, mostrando unos residuos totalmente rígidos en la simulación en decano puro. Esta restricción en la movilidad es un signo de que en un medio altamente hidrofóbico la acción catalítica se ve disminuida o incluso cancelada por la baja libertad de movimiento que tienen en ausencia de agua los residuos de la proteína. En presencia de decano en las trayectorias W8D y W6D los residuos Lys14 y His96 presentaron una ligera rotación a lo largo de la cadena alifática en el caso de la Lys y en la terminal carboxílica en el Glu. No obstante, en general, la estructura promedio del sitio activo se conservó en todos los casos.

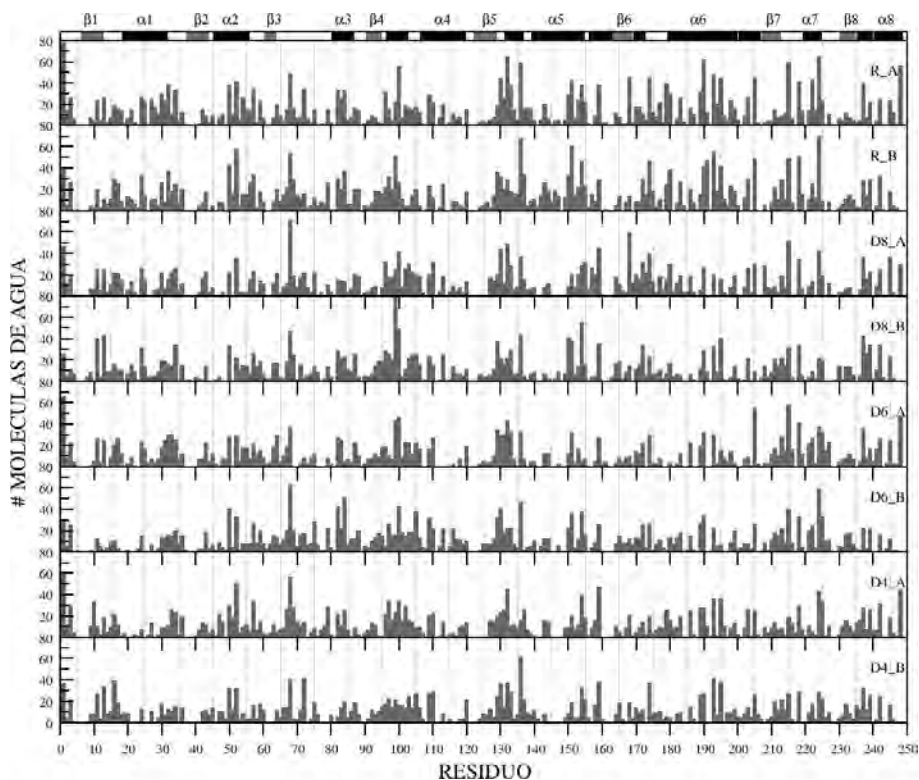
5.6. Análisis del disolvente en el entorno de la proteína

De la Figura 5.1.1 es evidente que la distribución de las moléculas de agua alrededor de la enzima al término de las simulaciones W8D, W6D y W4D es bastante

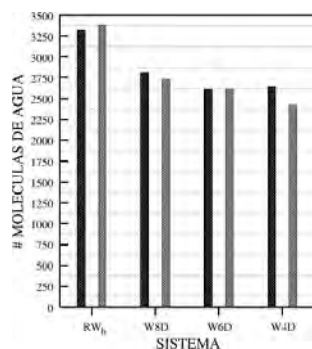
homogénea. El número de moléculas de agua cercanas a cada residuo (a una distancia menor a 4 \AA) fue casi el mismo en las trayectorias donde había agua y decano que en la simulación RW_b . A pesar de que la diferencia entre el número de moléculas de agua en cada caja de simulación es considerable, en RW_b hay 24,825 moléculas y en la simulación con menor cantidad (W4D) solo se tienen 1185. El perfil de la distribución de moléculas de agua en la superficie de la enzima en todos los casos es similar (Figura (5.6.1) (a)). Además, el número de moléculas de agua acumuladas que se encuentran a menos de 4 \AA de cualquier átomo de la proteína (Figura (5.6.1) (b)) muestra que el número total no es muy diferente entre las distintas simulaciones. El mismo análisis se realizó con las moléculas de decano.

ID	# Mol Agua	# Mol Decano	% Agua	Tiempo MID <i>ns</i>	Estruct. inicial (ID)	RoG (<i>nm</i>)	Puentes de H	Distancia A-B (<i>nm</i>)	SAS (I) (nm^2)	SAS (II) (nm^2)
RW _a	24825	0	100	20	ITCD					
RW _b	24825	0	100	140	RW _b (20)	2.46	379±9	3.58±0.03	117±2	93±2
W8D	3579	3402	11.80	40	RW _b (20)	2.50	380±11	3.66±0.03	128±2	90±2
W6D	2336	3208	8.45	40	RW _b (20)	2.49	388±10	3.63±0.02	129±2	90±1
W4D	1185	2825	5.04	40	RW _b (20)	2.392	392±9	3.35±0.02	127±2	87±1
RD	0	2587	0	40	ITCD	2.345	563±11	2.28±0.01	122±2	58.1±0.9

Tabla 5.6.2. Simulaciones realizadas, incluyendo parámetros de simulación y resultados del análisis. RoG: Radio de Giro. Puentes de Hidrógeno intramoleculares. Distancia A-B: Distancia entre los c.o.m. de los monómeros A y B a lo largo de la trayectoria. SAS: Área accesible al disolvente hidrofóbica (I) e hidrofílica (II).

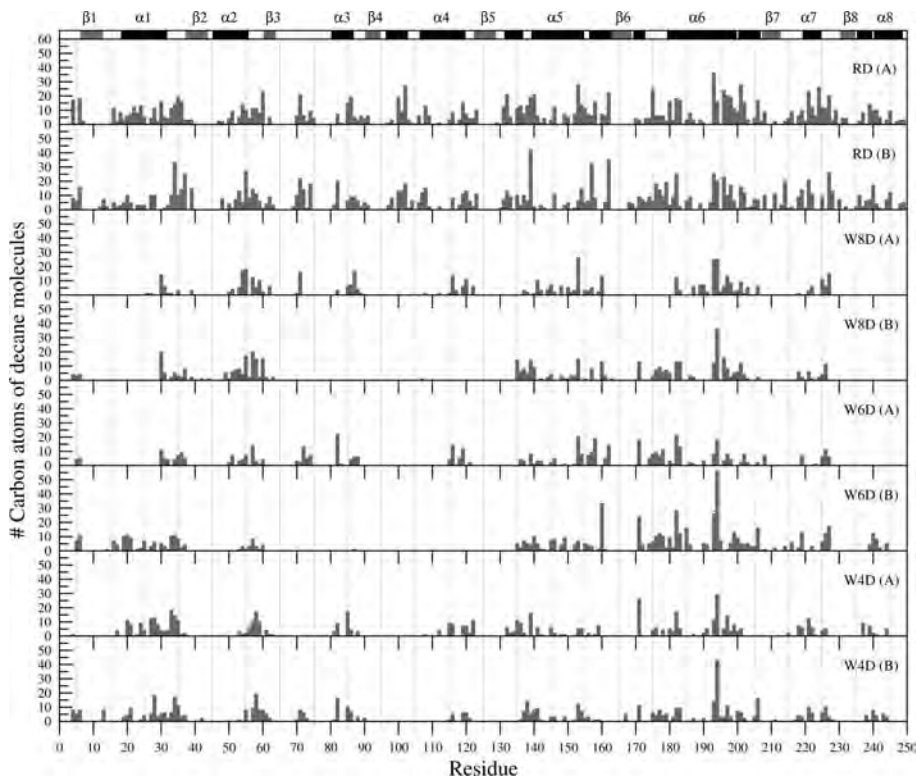


(a)

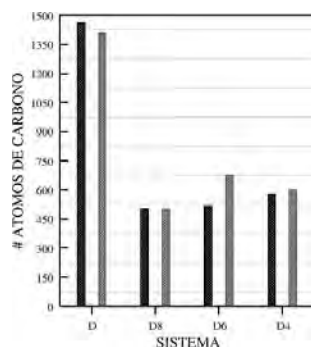


(b)

FIGURA 5.6.1. (a) Número de moléculas de agua que se encuentran a una distancia menor de 4 Å de cada residuo en la última conformación de las trayectorias RW_b, W8D, W6D y W4D. El nombre del sistema y de la subunidad (A ó B) se indican en la esquina superior derecha de cada línea. En la parte superior de la gráfica se indica el tipo de estructura secundaria correspondiente a la secuencia de aminoácidos de la TcTIM. (b) Número de moléculas de agua acumuladas calculadas sumando los valores de cada línea en la Figura (5.6.1)(b) para la subunidad A (gris) y la subunidad B (negro).



(a)



(b)

FIGURA 5.6.2. (a) Número de átomos de decano que se encuentran a una distancia menor que 4 Å de cada residuo en la última conformación de las trayectorias W8D, W6D, W4D y RD. El nombre del sistema y de la subunidad (A o B) se indican en la esquina superior derecha de cada línea. En la parte superior de la gráfica se indica el tipo de estructura secundaria correspondiente a la secuencia de aminoácidos de la TcTIM. (b) Número de átomos de decano acumulados calculados sumando los valores de cada línea en la Figura 5.6.2 (a). Los resultados se presentan para la subunidad A (gris) y la subunidad (B).

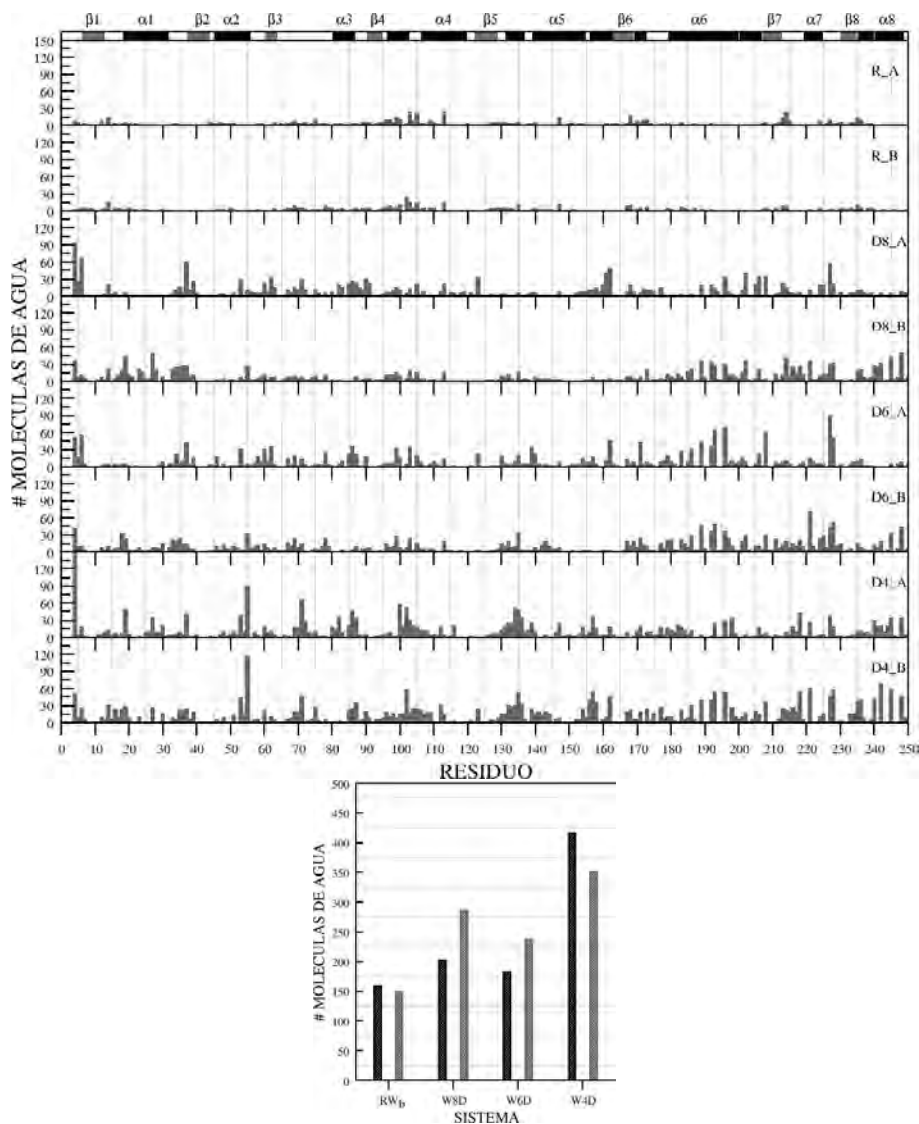


FIGURA 5.6.3. (a) Número de moléculas de agua que se encuentran a una distancia menor de 4 Å de cada residuo de la TcTIM, y que permanecieron al menos 2 ns a esa distancia durante los últimos 20 ns de simulación en las trayectorias RW_b, W8D, W6D y W4D. El nombre del sistema y de la subunidad (A o B) se indican en la esquina superior derecha de cada línea. En la parte superior de la gráfica se indica el tipo de estructura secundaria correspondiente a la secuencia de aminoácidos de la TcTIM. (b) Número de moléculas de agua acumuladas calculadas sumando los valores de cada línea en la Figura 5.6.3 (a) para la subunidad A (gris) y la subunidad B (negro).

Los resultados que se muestran en la Figura 5.6.2 corresponden al cálculo del número de átomos de decano que se encuentran a una distancia menor de 4 Å de

cada residuo de la proteína. La distribución de los átomos del decano no es tan homogénea entre trayectorias como lo fue en el caso de las moléculas de agua. Sin embargo, se puede observar que existen regiones de la proteína con una clara afinidad a las moléculas de decano, tal es el caso de los segmentos que comprenden la hélice- α_6 , el lazo 6, y el lazo $\alpha\beta_2$ (Figura 5.6.2 (a)). La ubicación de estas zonas con afinidad hidrofóbica puede ayudar a la búsqueda de posibles sitios de unión de inhibidores, tal como lo propone Mattos y Ringe (Ringe, 1995 Dec; Mattos & Ringe, 1996 May). El número de átomos de decano acumulados para cada monómero fueron calculados también y los resultados se presentan en la Figura 5.6.2 (b). Cabe aclarar que el mismo análisis fue realizado a más de una conformación de los últimos 20 *ns* en cada trayectoria, tanto para el cálculo del número de moléculas de agua como de átomos de decano, debido a que no hubieron variaciones significativas en los resultados éstos se presentan únicamente para la última conformación de cada trayectoria.

Como una estimación del tiempo de residencia de las moléculas de agua alrededor de cada residuo se determinó la cantidad de moléculas de agua permanecieron a una distancia menor a 4 Å de cada residuo durante al menos 2 *ns*. Este cálculo se realizó para los sistemas RW_b, W8D, W6D y W4D sobre últimos 20 *ns* de cada trayectoria 5.6.3. De la gráfica se puede observar que la residencia de las moléculas del agua en la superficie de la proteína es mayor cuando hay presencia de un disolvente orgánico en el medio. Lo anterior puede explicarse en base al bajo coeficiente de partición del agua en decano, de tal manera que las moléculas de agua prefieren establecer interacciones con la proteína que con las moléculas del decano. El valor acumulado de estas moléculas de agua fue también calculado (ver Figura 5.6.3 (b)). El aumento en el número de moléculas de agua con tiempos de residencia mayor a 2 *ns* en función del decremento de la concentración de agua es evidente. Un análisis similar realizado aumentando el tiempo de residencia límite para contabilizar las moléculas de agua mostró un comportamiento similar, solo que el número de moléculas de agua en cada caso fue menor.

5.7. Puentes de hidrógeno intramoleculares

En trabajos anteriores, se ha propuesto que la razón por la cual muchas enzimas mantienen su movilidad y la actividad enzimática a bajas concentraciones de agua se debe a que pueden formar puentes de hidrógeno con las moléculas de agua de su entorno (Zaks & Klibanov, 1988a). En ausencia de agua, los grupos polares de la proteína en lugar de establecer interacciones con el disolvente tratan de hacerlo con ellos mismos, aumentando así el número de puentes de hidrógeno intramoleculares.

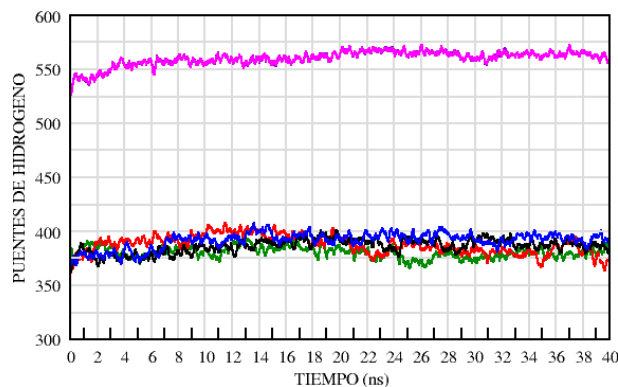


FIGURA 5.7.1. Número de puentes de hidrogeno intramoleculares a lo largo de las trayectorias. RW_b (verde), W8D (rojo), W6D (negro), W4D (azul), y RD (magenta).

Para hacer una comparación entre trayectorias el número de puentes de hidrógeno fue calculado en cada uno de ellas, los resultados se muestran en la Tabla 5.6.2.

En el caso de la simulación en decano, no es posible formar puentes de hidrógeno con el solvente, en cambio el número de puentes de hidrógeno intramoleculares fue muy grande. Un resultado similar se obtuvo si en lugar de calcular los puentes intramoleculares de todos los residuos se contabilizan sólo aquellos que forman parte de algún tipo de estructura secundaria. Los resultados fueron similares. El incremento en el número de puentes de hidrógeno observado en la trayectoria RD parece ser el causante de la aparente contracción de la proteína en esta simulación. Dicha contracción se cuantificó por medio del radio de giro (RoG) de la proteína en cada trayectoria. Los resultados del RoG calculado para la proteína en los últimos 20 ns de cada trayectoria se muestran en la Tabla 5.6.2. Las contribuciones hidrofóbicas al área accesible al solvente (SAS) no variaron mucho entre los distintos sistemas, no obstante el valor de SAS hidrofílica en la simulación RD fue significativamente más pequeño que en los demás sistemas (ver Tabla 5.6.2).

CAPÍTULO 6

Conclusiones

Los efectos de la relación entre la concentración agua/decano y la estructura y comportamiento dinámico de la triosafosfato isomerasa, así como el comportamiento de las moléculas del disolvente en la vecindad de la enzima, fueron estudiados en detalle por medio de simulaciones de dinámica molecular. En contraste con estudios recientes realizados a escalas de tiempo mucho menores, se encontró que la enzima es sensible a la concentración relativa de agua. Tanto el análisis de las fluctuaciones como de los componentes principales del movimiento mostraron, que, en presencia de un solvente orgánico, a mayor concentración de agua mayor es la movilidad de la enzima. Sin embargo, la movilidad de la enzima en agua pura mostró amplitudes de los movimientos comparables a las que se obtuvieron de la simulación W4D, la simulación con la menor concentración de agua, mientras que se observó una clara rigidez de la proteína en decano puro. Estos resultados sugieren que la enzima aumenta su movilidad, probablemente debido a la presencia de moléculas del solvente orgánico situadas en zonas específicas de la superficie de la proteína. Es importante señalar que existe una cantidad mínima de moléculas de agua necesaria para que la enzima mantenga su estructura y movilidad.

A partir del comportamiento de la triosafosfato isomerasa en solución acuosa se encontró que la enzima se comporta como una estructura flexible, en la que tienen lugar movimientos concertados entre las dos subunidades. A lo largo de la trayectoria de 140 *ns* en agua pura se observó que el lazo catalítico muestra conformaciones abiertas y cerradas similares a las obtenidas en la literatura que se relacionan directamente con el mecanismo a través del cual la proteína lleva a cabo su función. Además, se observa una conformación semiabierta que parece ser la estructura más estable a lo largo de la trayectoria en ausencia de sustrato. Este análisis proporciona nueva información relacionada con la dinámica interna de la TcTIM que puede ser utilizada para entender mejor esta enzima.

Parte 2

Estudio de la estabilidad dimérica de
triosafosfato isomerasa de *Tripanosoma*
cruzi, *Tripanosoma brucei* y humana.

CAPÍTULO 7

Introducción

El desarrollo de nuevos fármacos enfocados al tratamiento de enfermedades parasitarias es un trabajo difícil. Su complejidad radica no sólo en poder diseñar fármacos efectivos, sino que además estas moléculas deben ser selectivas para los parásitos y no afectar al organismo anfitrión. A lo largo del proceso evolutivo los sitios catalíticos entre proteínas de distintas especies son altamente conservados lo cual complica el diseño de fármacos selectivos (Olivares-Illana *et al.*, 2008). La solución a este problema podría ser diseñar fármacos efectivos contra el parásito pero no con el huésped. Esta alternativa requiere por supuesto de un conocimiento a fondo del metabolismo del parásito, y de la estructura y mecanismos dinámicos de la proteína en específico. Se sabe que la conservación de los residuos en la interfase de proteínas oligoméricas es mucho más baja que en los sitios catalíticos. Así, la variabilidad que presentan estos residuos permite la distinción entre especies (Caffrey *et al.*, 2004). Esto aunado al hecho de que en algunos casos los monómeros de una misma especie por si solos no son catalíticamente activos ofrece la posibilidad de tener un escenario útil para el diseño de fármacos selectivos (Keskin *et al.*, 2008).

La triosafosfato isomerasa (TIM) es una enzima ubicua que se encuentra presente en casi todos los organismos (Peimbert *et al.*, 2008). Su función consiste en catalizar la interconversión del gliceraldehído 3 fosfato y la dihidroxiacetona fosfato en un paso de la secuencia metabólica de la glucólisis (Albery & Knowles, 1976). Desde el punto de vista enzimático, la TIM ha sido reconocida como el catalizador perfecto dado que su actividad sólo es limitada por la difusión del sustrato (Knowles, 1991; Blacklow *et al.*, 1988). Esta característica la hace una de las enzimas más estudiadas. Hasta la actualidad se han resuelto 19 estructuras distintas de TIM por medio de difracción de rayos X (Peimbert *et al.*, 2008). La TIM se pliega en un barril $(\alpha\beta)_8$. Se estima que casi un 10% de las proteínas conocidas y que presentan actividad catalítica comparten el mismo patrón de plegamiento (Farber & Petsko, 1990). Cada barril consiste en ocho hebras beta alternadas con ocho hélices alfa. Las hebras beta se encargan de formar el barril interno mientras que las hélices forman otro barril externo (ver Figura 8.2.1 (c)). La estructura funcional de la TIM ocurre cuando ésta

se encuentra en su estado oligomérico. Típicamente es un homodímero, aunque se han encontrado algunas estructuras tetraméricas en diversas especies hipertermofílicas (Peimbert *et al.*, 2008). Algunos experimentos realizados en estructuras monoméricas de TIM obtenidas a partir de ingeniería genética soportan la hipótesis de que la dimerización induce una correcta orientación de los residuos catalíticos (Borchert *et al.*, 1994). Además, algunos estudios computacionales sugieren la existencia de un movimiento cooperativo entre monómeros que puede ser esencial para que la enzima lleve a cabo su función catalítica (Cansu & Doruker, 2008; Diaz-Vergara & Pineiro, 2008). Cada subunidad de la TIM contiene aproximadamente 250 residuos. A lo largo de la evolución tanto la estructura como la secuencia de aminoácidos ha sido altamente conservada entre distintas especies. En algunos casos el porcentaje de identidad es de hasta 98.4%, entre el humano y el conejo, mientras que en otros llega a ser tan bajo como de 28.8% entre la TIM humana y la correspondiente al *Pyrococcus woesei* (Peimbert *et al.*, 2008). La intercara entre las subunidades está formado principalmente por los lazos 1, 2, 3. El lazo 3 de cada subunidad se extiende formando estrechos contactos con los residuos de los lazos 1, 2 y 4 de la subunidad adyacente. Los residuos catalíticos (Lys13, His95 y Glu165) localizados en los extremos terminales de los lazos 5, 6 y 7 están completamente conservados. Al lazo 6 usualmente se conoce como el lazo catalítico debido a que aparentemente se abre permitiendo que el sustrato llegue al sitio activo y al cerrarse manteniendo el sustrato en el sitio activo se lleva a cabo la reacción. Se ha demostrado que algunas mutantes presentan una disminución de la actividad cuando una parte de este lazo es eliminada (Pompliano *et al.*, 1990). La función de la TIM consiste catalizar un paso en la secuencia metabólica de la glucólisis y de esta manera asegurar la producción de ATP. En parásitos anaerobios ésta es la única fuente de producción de energía por lo que la presencia de esta enzima es imperativa (Peimbert *et al.*, 2008). En base a que las unidades aisladas de la estructura nativa de la TIM no son activas y tomando en cuenta que los residuos y la estructura de las interfaces entre subunidades varían en función de la especie, esta enzima es un blanco ideal para el desarrollo de tratamientos por infecciones causados por parásitos anaerobios.

Las tripanosomiasis son enfermedades parasitarias consideradas como en la actualidad como enfermedades infecciosas emergentes (Jones *et al.*, 2008), su origen radica en la invasión del organismo anfitrión por parte de especies protozoarias del género de los tripanosomas. En muchos casos las enfermedades son asintomáticas, situación que depende en gran medida del parásito. Con frecuencia estas enfermedades no son diagnosticadas y sin tratamiento persisten hasta la muerte del organismo anfitrión.

A pesar de su prevalencia en ciertas regiones donde se consideran como endémicas y de su creciente propagación por el mundo (Schmunis, 2007), aún no existe un tratamiento efectivo para este tipo de enfermedades. Distintos grupos de investigación han enfocado sus esfuerzos al diseño de fármacos para tratar la tripanosomiasis basados en la selectividad de los inhibidores hacia distintas especies de TIM. Algunos inhibidores diseñados han demostrado su habilidad para inhibir la TIM de tres especies tripanocidas casi sin afectar la TIM humana (Olivares-Illana *et al.*, 2006). Estas moléculas han demostrado tener clara afinidad a las interfases de las distintas TIMs. Una forma de inhibir la acción catalítica es favoreciendo la monomerización de las enzimas, puesto que algunos experimentos han demostrado que la TIM es activa únicamente en su forma dimérica (Waley, 1973). En esta línea Sight y cols. (Singh *et al.*, 2001) estudió la interacción de distintas TIMs y polipéptidos con secuencias tomadas del lazo 3. El residuo en la posición 15 (según la numeración de la TcTIM), Cys en parásitos y Met en la TIM humana, representa una de las diferencias más importantes entre las especies. En todos los casos este aminoácido aparece rodeado por el lazo 3 de la subunidad opuesta (ver Figura 8.2.1 (a)) y por lo tanto parece jugar un papel muy importante en la estabilización del dímero. Los residuos que conforman el lazo 3 están altamente conservados. Doce de sus catorce residuos son idénticos en la TcTIM y en la TbTIM. Los residuos que no están conservados, la Ala70 y Lys71 en TbTIM, no mantienen contacto con la Cys15 (Zomosa-Signoret *et al.*, 2007). A pesar de la gran similitud entre la TIM de ambas especies la susceptibilidad de la Cys15 a tioles es aproximadamente dos órdenes de magnitud mayor en TcTIM que en la TbTIM. En heterodímeros formados por monómeros del tipo TcTIM-TbTIM se ha observado que el comportamiento dinámico de la enzima determina la exposición de la Cys15 a los agentes externos, por lo que en particular es el comportamiento dinámico de la subunidad adyacente quien determina la reactividad de cada Cys15 (Zomosa-Signoret *et al.*, 2007). En estudios que consistieron en la medición de la reactivación catalítica de estructuras provenientes de monómeros desplegados se encontró que la velocidad y la eficiencia de la formación de dímeros activos es mucho más grande en TcTIM que en TbTIM (Zomosa-Signoret *et al.*, 2003). Estos resultados contrastan con los obtenidos mediante el desplegamiento térmico los cuales revelaron que la desnaturalización TcTIM es varias ordenes de magnitud más lenta que en la TbTIM, indicando que la primera es una estructura cinéticamente más estable (Costas *et al.*, 2009). Así, la TIM se presenta como una proteína blanco apropiada para tratar la tripanosomiasis, basados en que se puede diseñar un inhibidor que actúe selectivamente en la intercara de la proteína. En la Figura 7.1.1 se muestra una comparación

detallada de los residuos que conforman las intercaras de la TcTIM, TbTIM y hTIM. En la Figura 7.1.2 se presenta un análisis de la conservación de residuos a lo largo de la secuencia de aminoácidos entre las tres estructuras.

A pesar de los avances realizados hasta ahora, aún quedan muchas incógnitas por resolver para entender el comportamiento de la TIM. En particular, hasta ahora no ha sido posible determinar el cambio en la energía libre asociada al proceso de monomerización de las TIMs. Este parámetro es muy importante puesto que proporciona una medida de la estabilidad de la estructura dimérica necesaria para que la enzima realice su actividad catalítica y por consiguiente es un dato esencial en el diseño de fármacos especie-selectivas de TIM.

7.1. Aproximación de Jarzynski

En este trabajo, tanto la estabilidad dimérica de la TIM de tres distintas especies (*Trypanosoma brucei*, *Trypanosoma cruzi* y *humana*), como la energía libre necesaria para separar los monómeros de la TIM, se determinó por medio de simulaciones de dinámica molecular. Para efectuar esta estimación de la energía libre entre dos moléculas (A y B) se utilizan típicamente dos técnicas. La primera, también conocida como perturbación de la energía libre, consiste en desacoplar lentamente una molécula de su entorno. En segundo lugar están los métodos basados en el cálculo del potencial de fuerza media que siente una molécula B a lo largo de una coordenada de reacción cuando es separada lentamente de otra molécula A. Ambas técnicas han demostrado obtener buenos resultados principalmente en la estimación de diferencias de energías libres de sistemas pequeños en las que es factible alcanzar el equilibrio después de algunas decenas de nanosegundos. Estas metodologías tienen el inconveniente de que para obtener resultados confiables se requiere de un número grande de simulaciones que permita un muestreo exhaustivo del espacio de configuraciones, lo cual siempre conlleva a un excesivo tiempo de cómputo, sobretodo cuando el sistema está compuesto por moléculas de gran tamaño. Considerar todos los cambios conformacionales que normalmente componen la dinámica de proteínas representa un problema para estas técnicas. Una nueva técnica que representa una alternativa para obtener energías libres de sistemas macromoleculares fue propuesta hace pocos años por C. Jarzynski. Este autor propone una relación entre el trabajo no reversible que se requiere para llevar un sistema de un estado A a otro estado final B, con la energía libre del proceso. En su ecuación propone que el exponencial de la energía libre es igual al promedio de la exponencial de los valores del trabajo (W_i) para llevar

el sistema de un estado a otro obtenido de N repeticiones independientes del proceso (Jarzynski, 1997b,a).

$$(7.1.1) \quad \exp^{-\beta\Delta F} = \langle \exp^{-\beta W_i} \rangle$$

donde $\beta = (kT)^{-1}$, k es la constante de Boltzmann y T la temperatura. ΔF es la energía libre entre dos estados termodinámicos. El lado derecho de la ecuación es el promedio del trabajo de N trayectorias del proceso, partiendo cada una de ellas de una estructura equilibrada. La velocidad con que el proceso se lleve a cabo y el número de trayectorias promediadas influirá en la precisión y la fiabilidad del proyecto.

Esta aproximación ha sido utilizada con razonable éxito en una variedad de sistemas (Bastug *et al.*, 2008; Xiong *et al.*, 2006; Hummer & Szabo, 2001). Sin embargo, su aplicación a procesos de monomerización de proteínas es escaso (Cuendet & Michielin, 2008). Una de las características de esta técnica y a su vez una ventaja, es que las trayectorias del proceso pueden encontrarse fuera del equilibrio. Es decir que la expresión propuesta por Jarzynski proporciona la herramienta para estimar la energía libre de Gibbs a partir de simulaciones fuera del equilibrio. Cuando el número de trayectorias obtenidas no es suficiente para que la igualdad se cumpla se asocia un error sistemático a los cálculos que puede ser grande. Gore y cols. (Gore *et al.*, 2003), propusieron una forma de evaluar este error, permitiendo así extraer así el valor de la energía libre a partir de un número finito de trayectorias. Basándonos en esta técnica, en este trabajo se simularon un número significativo de trayectorias correspondientes 450 ns de simulación del proceso de separación entre dos subunidades para cada una de las tres TIMs mencionadas. El proceso se realizó a dos velocidades distintas y la energía libre de Gibbs asociada a cada proceso se calculó mediante la aproximación de Jarzynski.

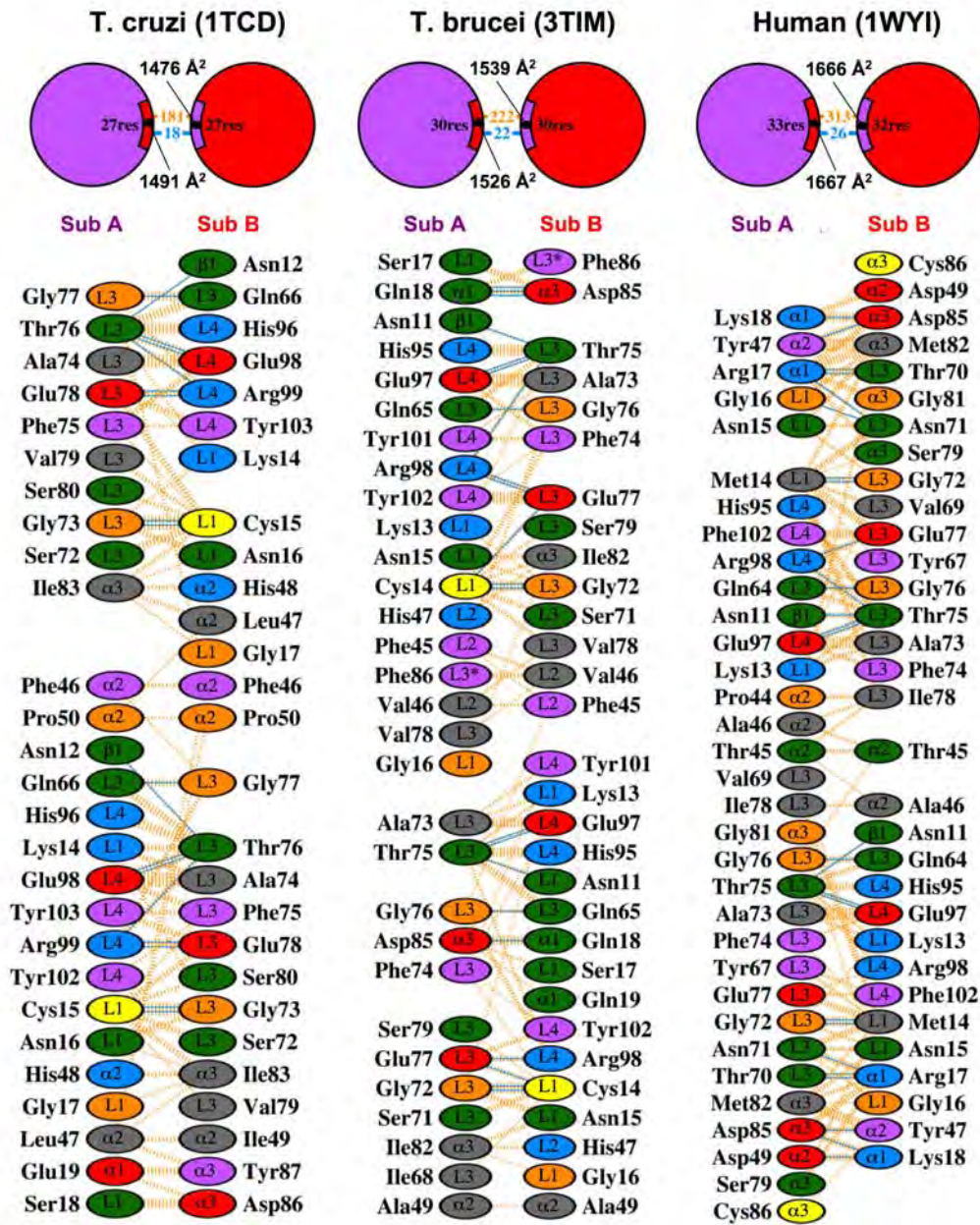


FIGURA 7.1.1. Esquema comparativo de los residuos en la interfase de la TbTIM, TcTIM y hTIM.

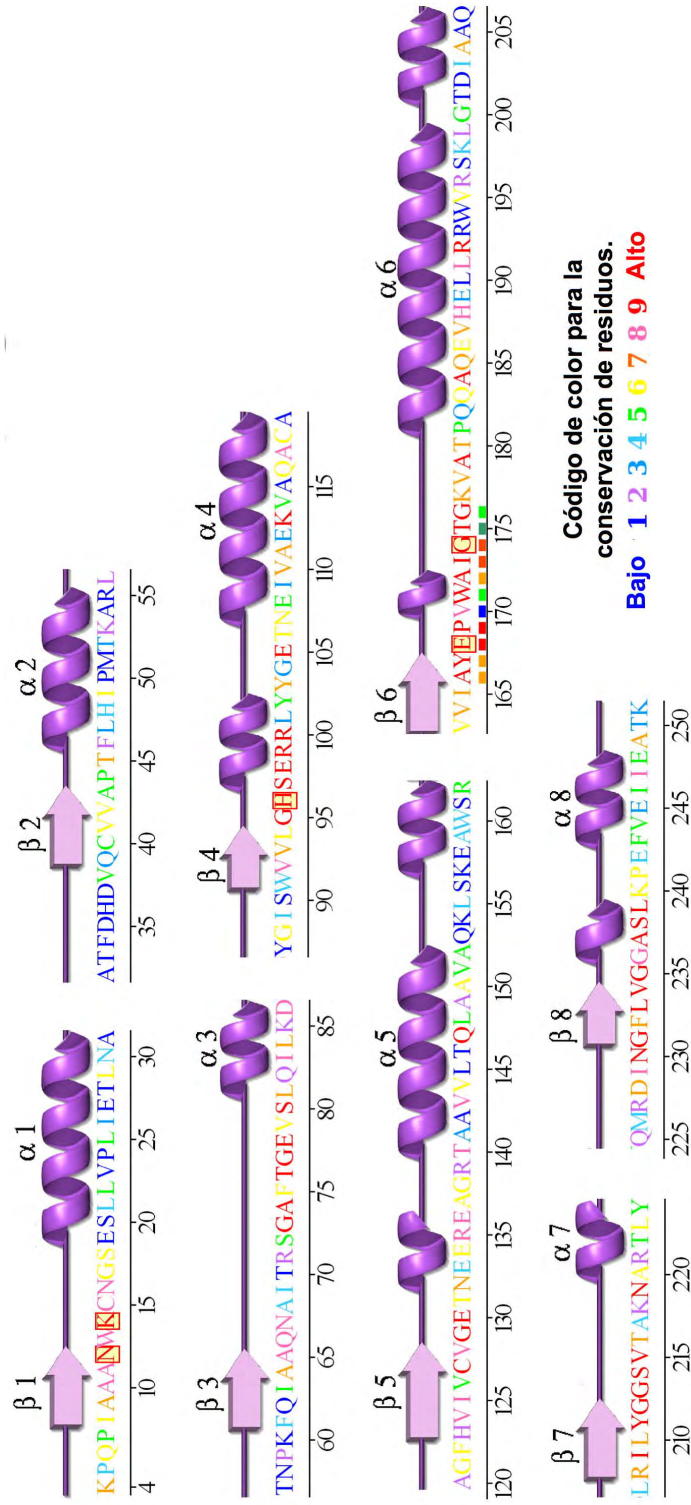


FIGURA 7.1.2. Secuencia de aminoácidos y conservación de residuos en la estructura primaria de la TbTIM, TcTIM y hTIM.

CAPÍTULO 8

Metodología

8.1. Parámetros de simulación

Todas las simulaciones se realizaron con el paquete computacional GROMACS 3.3.1 (Berendsen *et al.*, 1995; Lindahl *et al.*, 2001; Van Der Spoel *et al.*, 2005; www.gromacs.org, GROMACS) utilizando el campo de fuerza GROMOS96 (53a6) (Oostenbrink *et al.*, 2004) para la proteína y el modelo de agua SPCE por sus siglas en inglés Simple Point Charge Extended (Kusalik & Svishchev, 1994). Las interacciones entre átomos no enlazados se consideraron de la siguiente manera: para distancias menores a 0.9 nm se calcularon a cada paso de tiempo; las interacciones que se comprenden entre 0.9 nm y 1.5 nm se evaluaron cada 5 pasos de tiempo; y para distancia mayores a 1.4 nm se consideró un medio continuo con una constante dieléctrica $\epsilon = 78,0$. El paso de tiempo utilizado en cada caso fue de 2 femtosegundos (*fs*). Para evitar interacciones desfavorables antes de cada simulación se realizó una minimización de energía mediante el método Steepest Descent. Las velocidades iniciales se asignaron en base a una distribución de Maxwell-Boltzmann a 298 K. Tanto los ángulos como las distancias de enlace entre moléculas de agua se restringieron utilizando el algoritmo SETTLE (Miyamoto & Kollman, 1992). Para mantener los enlaces dentro de la proteína y el decano se utilizó el algoritmo de LINCS (Hess *et al.*, 1997). A lo largo de cada simulación las coordenadas, las velocidades y las energías se almacenaron con una frecuencia de cada 10 picosegundos (*ps*) para su posterior análisis.

8.2. Construcción de los sistemas

Trayectorias de referencia. Las estructuras de TbTIM, TcTIM y hTIM se tomaron del Protein Data Bank (www.pdb.org, PDB) con códigos PDB 3TIM, 1TCD y 1WYI, respectivamente. Cada estructura se colocó dentro de una caja de simulación con geometría rectangular de 7 x 8 x 14 nm³, con la proteína orientada de tal manera que la línea que une el centro de ambas subunidades coincida con la dirección del eje *z*. Las cajas se llenaron con aproximadamente 25,000 moléculas de agua previamente equilibradas. La carga eléctrica de cada caja de simulación se neutralizó añadiendo

iones de Sodio o Cloro en cada sistema: 6 y 12 Cl^- para las cajas con la TcTIM y la TbTIM, respectivamente, y 2 Na^+ para la caja con la hTIM. Posteriormente, se obtuvieron 40 *ns* de simulación para cada trayectoria. En estas simulaciones se aplicó una restricción orientacional evitando la rotación angular de la proteína, y además, permitiendo que el eje más largo de la proteína y de la caja permanezcan paralelos a lo largo de la simulación. Para mantener la presión a 1 bar y la temperatura a 298 K se utilizaron un barostato y un termostato de Berendsen con constantes de acoplamiento de 1 ps y 0.5 ps, respectivamente (Berendsen *et al.*, 1984). A estas trayectorias se les dio el nombre de trayectorias de referencia.

Trayectorias de monomerización. La monomerización de las distintas especies de TIM se realizó utilizando simulaciones de dinámica molecular imponiendo una contribución a la función potencial que permite separar los monómeros a una velocidad constante hasta una distancia entre sus centros de masas (c.o.m.) a partir de la distancia de equilibrio, la de la estructura cristalina (Figura 8.2.1 (b)). La separación de los monómeros se realizó a partir de los centros de masas en la dirección del eje de coordenadas z aumentando su distancia hasta 1.5 *nm*. Las velocidades de separación fueron de 10 Å/*ns* y 5 Å/*ns*. Para realizar estas simulaciones se definió un parámetro λ que varía de 0 a 1 a lo largo de la simulación con distinta rapidez entre los estados inicial y final. De estas simulaciones se realizaron un total de 108 trayectorias (162 *ns*) para la velocidad de separación más rápida y de 96 trayectorias (288 *ns*) para la más lenta (ver Tabla 8.2.1). En cada caso se determinó el trabajo W_i necesario para llevar a cabo cada transformación. Para justificar la distancia máxima de separación se monitoreó el potencial de Lennard Jones a lo largo de las distintas trayectorias. De este análisis se encontró que a pesar de que las interacciones entre monómeros se pierden a los 3 *nm*, 1.5 *nm* de separación son suficientes para asegurar que se toman en cuenta todas las interacciones que contribuyen de forma significativa a la estabilización de la intercara dimérica. Todas las simulaciones de monomerización fueron realizadas a volumen constante y a una temperatura de 298 K. Para controlar la temperatura se utilizó un termostato de Berendsen con la misma constante de acoplamiento que en las simulaciones de referencia.

8.3. Métodos de análisis

Todas las imágenes se generaron con los visualizadores PyMOL 0.99 (DeLano, 2008) y VMD (Humphrey *et al.*, 1996). Para el análisis se utilizaron tanto herramientas del paquete computacional GROMACS, como scripts de análisis en perl, awk y

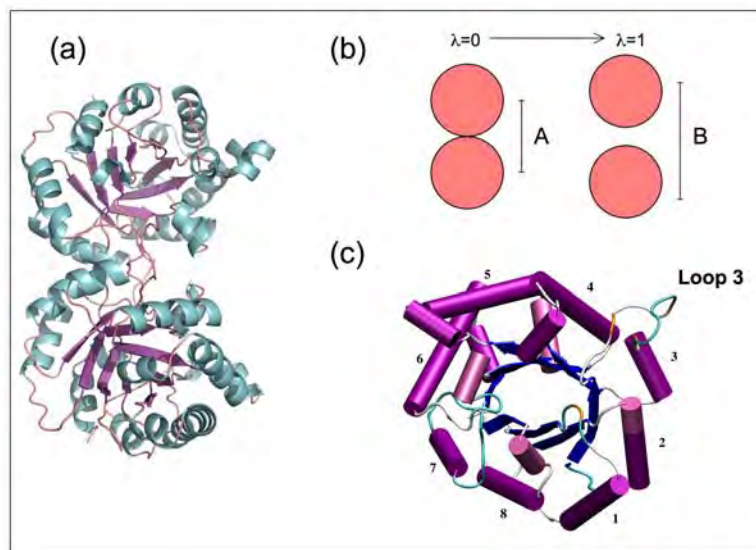


FIGURA 8.2.1. (a) Representación de la TcTIM en la configuración inicial. (b) Esquema del proceso de monomerización. (c) Vista superior de un monómero de la TcTIM, en la figura muestra la numeración de las α -hélices, y la ubicación del lazo 3.

Velocidad ($\text{\AA}/ns$)	N	Tiempo total (ns)
10	108	162
5	96	288

TABLA 8.2.1. Número total de simulaciones de monomerización realizados para cada sistema.

bash. Específicamente para el análisis de las trayectorias de monomerización utilizando el método de Jarzynski junto con las correcciones propuestas por Gore (Gore *et al.*, 2003) se desarrollaron programas en el lenguaje de programación Fortran 77.

Trayectorias de referencia. En el análisis de las trayectorias de referencia se tomaron en cuenta las consideraciones que a continuación se describen. La estructura de referencia para el cálculo del RMSD y RMSF fue la estructura cristalina de cada proteína tomada directamente del PDB. Parámetros como el radio de Giro (RoG), el área accesible al disolvente (SAS) tanto hidrofóbico como hidrofílico se calcularon a

lo largo de las trayectorias. El número de puentes de hidrógeno intramoleculares de cada proteína se calculó utilizando la convención de que la distancia donador-aceptor necesaria para la formación de un puente debe ser menor a 30 Å y con un ángulo menor a 30°. Se realizó un análisis de componentes principales del movimiento para cada una de las trayectorias de referencia; éstos se obtuvieron a través de la diagonalización de la matriz de covarianza de todos los átomos de la enzima, obteniéndose así los autovalores y autovectores correspondientes a cada movimiento desacoplado dentro de la proteína (ver sección 2.2.3). Para estudiar la dinámica interna de cada proteína se utilizaron tanto los componentes principales como las fluctuaciones cuadráticas medias. El primer nanosegundo de cada trayectoria se eliminó para todos los análisis.

Simulaciones de monomerización. La igualdad de Jarzynski (2.3.9) se utilizó para calcular el cambio en la energía libre correspondiente al proceso de monomerización. Para calcular el error asociado a cada valor de ΔF se utilizaron las ecuaciones propuestas por Gore y cols. (Gore *et al.*, 2003). El procedimiento se explica a detalle en la sección 4.3.

Resultados y discusión

9.1. Análisis estructural

El hecho de que las TIMs estudiadas poseen el mismo patrón de plegamiento y que presenten una alta coincidencia en su secuencia de aminoácidos, apunta hacia un comportamiento dinámico muy similar entre ellas. Sin embargo, como sugieren algunos estudios experimentales en los que se compara la TIM de agentes tripanocidas y la TIM de humana, deben existir ciertas características internas que ocasionen las diferentes respuestas ante la presencia de agentes externos y/o mutaciones estructurales.

En este trabajo se realizó una comparación de la estructura de las TIMs estudiadas. Como es bien sabido, las tres enzimas registran una alta similaridad, principalmente en el patrón de plegamiento (Figura 9.1). Sin embargo, a nivel de la secuencia de aminoácidos la identidad es solamente de 74% entre TbTIM y TcTIM, y de 54% entre éstas y la hTIM (identidades calculadas con el servidor

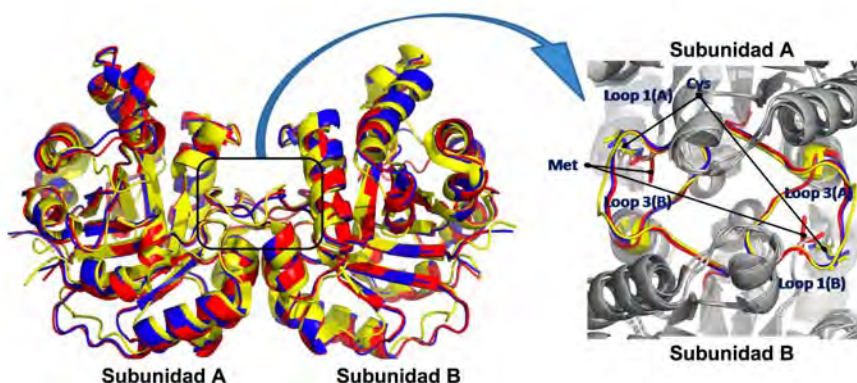


FIGURA 9.1.1. Representación de la superposición estructural de la triosafosfato isomerasa de *Tripanosoma brucei* (azul), *Tripanosoma cruzi* (amarillo) y humana (rojo). Las estructuras se tomaron directamente del PDB con los códigos 3TIM, 1TCD y 1WYI para cada una de las enzimas, respectivamente. Se incluye una vista de la interfase resaltando al lazo 3 en cada subunidad, al residuo Cys15, en TbTIM y TcTIM, y Met14 en hTIM.

Sistema	RoG (<i>nm</i>)	Dist AB (<i>nm</i>)	RMSD (<i>nm</i> ²)	SAS (I) (<i>nm</i> ²)	SAS (II) (<i>nm</i> ²)	RMSF (<i>nm</i>)	Puentes Hidrógeno
TbTIM	2.48	3.54±0.28	0.2848±0.03	124.5±3	114±2	0.1466±0.1	376±11
TcTIM	2.49	3.56±0.30	0.2489±0.03	123.7±3	113±2	0.1294±0.1	388±10
hTIM	2.48	3.52±0.37	0.2987±0.04	128.3±2	115±3	0.1573±0.1	367±10

TABLA 9.2.1. Propiedades de la proteína promediadas a lo largo de las simulaciones TbTIM, TcTIM y hTIM. RoG, RMSD, SAS hidrofóbica (I) e hidrofílica (II), RMSF y puentes de Hidrógeno se calcularon y promediaron sobre 39 *ns* de simulación. La distancia AB representa la distancia entre c.o.m. entre monómeros.

LALIGN:http://www.ch.embnet.org/software/LALIGN_form.html). Por lo que a nivel de interfase, las diferencias deben marcar un comportamiento distinto entre las enzimas de estudio. La comparación de la region interfacial en las TIMs se describe en la Figura 7.1.1. Los resultados demuestran que entre 27, 30 y 32-33 residuos de cada subunidad se encuentran en contacto directo con su monómero adyacente para TcTIM, TbTIM y hTIM, respectivamente. Ésto, traducido a área de contacto nos dice que hay una superficie de 1476-1491, 1529-1539, y 1666-1667 Å² en contacto en la interfase para cada subunidad de la TcTIM, TbTIM y hTIM. A priori, este resultado podria indicar una mayor estabilidad dimérica en la hTIM al poseer una mayor área de contacto entre subunidades, sin embargo esta aseveración no puede ser afirmada a partir tan sólo de esta información estructural que no tiene en cuenta ni el método efecto del disolvente ni el comportamiento dinámico de la enzima.

9.2. Análisis de las trayectorias de referencia

Inicialmente se analizaron los resultados de las tres trayectorias TbTIM, TcTIM y hTIM que contienen 40 *ns* de simulación cada una. La estructura global de las tres TIMs se mantuvo estable a lo largo de estas las trayectorias tal como indican las gráficas de RMSD (Figura 9.2). En la Tabla 9.2.1 se presenta, entre otros parámetros, el RMSD promedio a lo largo de cada trayectoria, de ahí se observa que el RMSD correspondiente a la trayectoria de hTIM es ligeramente mayor al que se obtiene para TbTIM y TcTIM. La información que contiene el RMSD puede sugerirnos que la hTIM experimenta, a lo largo de la simulación, mayores cambios estructurales y/o dinámicos, por lo que más adelante se estudia este comportamiento a detalle.

El radio de giro (RoG) de cada enzima a lo largo de las simulaciones fue prácticamente constante para las 3 enzimas estudiadas tomando valores entre 2.48 y 2.49

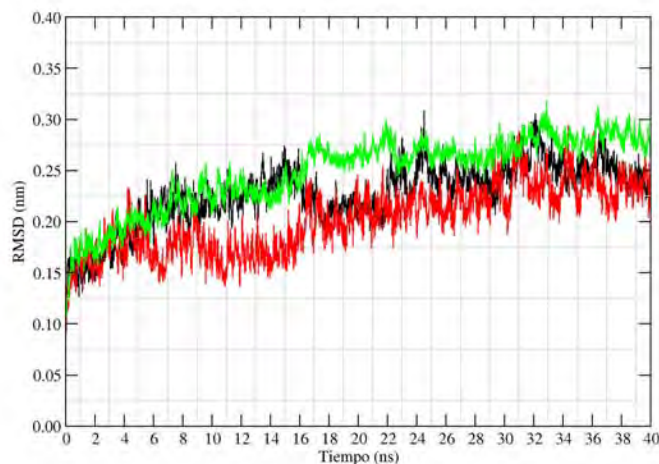


FIGURA 9.2.1. Desviaciones cuadráticas medias (RMSD) de la TbTIM (negro), TcTIM (rojo) y hTIM (verde), tomando en cuenta únicamente los átomos de la cadena central. Como referencia en todos los casos se utilizó la estructura cristalina correspondiente a cada TIM.

nm. Tanto la distancia entre centros de masas de los monómeros, como el área hidrofóbica e hidrofílica accesible al disolvente fue muy similar entre los sistemas (Tabla 9.2.1). Los puentes de Hidrógeno intramoleculares que proporcionan estabilidad a la estructura se calcularon también para cada trayectoria (Tabla 9.2.1). Los resultados muestran que la enzima hTIM tiene aproximadamente 20 puentes de hidrógeno menos que TbTIM y TcTIM. Además, la TcTIM es la enzima que presenta el mayor número de puentes de hidrógeno intramoleculares con un total de 338. En promedio el número de puentes de Hidrógeno decrece en orden TcTIM >TbTIM >hTIM.

9.3. Análisis estructural

La movilidad de cada residuo se comparó a través del cálculo del RMSF por residuo para cada proteína (Figura 9.3.1). Los perfiles de RMSF, como se puede observar son similares en las tres especies. Al igual que se reporta en otro conjunto de simulaciones realizadas con TcTIM (Diaz-Vergara & Pineiro, 2008), los segmentos que comprenden los lazos 5 y 6 presentan las máximas fluctuaciones en ambas subunidades. Esta característica se repite en las tres proteínas estudiadas (Figura 9.3.1). Particularmente, L3*, L5*- α 5, L6* y L6 (donde el asterisco (*) denota a los lazos $\beta\alpha$) corresponden a las regiones con mayor movilidad en cada subunidad en las trayectorias. Se observa que los extremos de las cadenas de aminoácidos de la hTIM poseen

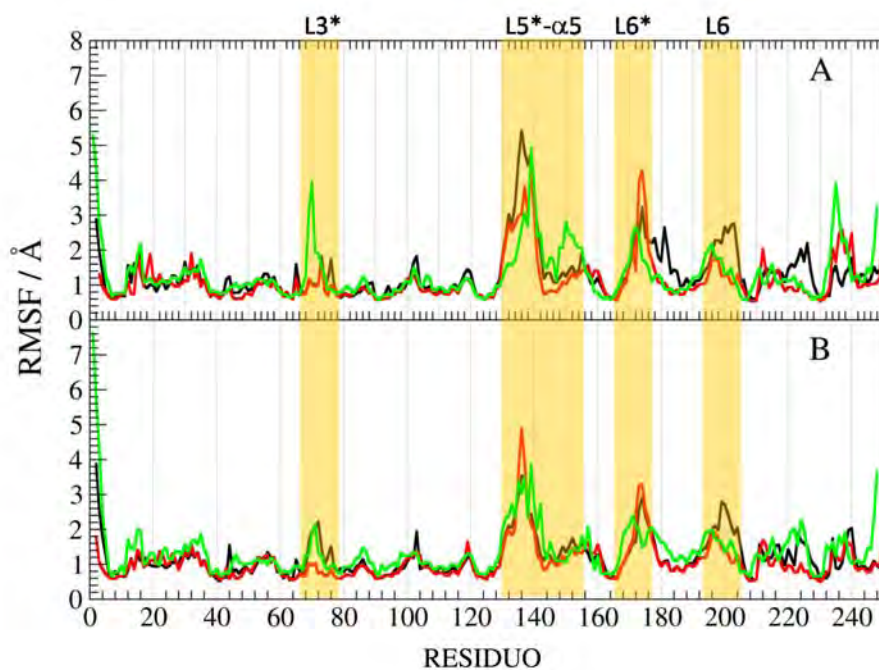


FIGURA 9.3.1. Fluctuaciones cuadráticas medias (RMSF) por residuo calculadas sobre 39 *ns* de trayectoria de cada proteína. La gráfica correspondiente a cada subunidad de la proteína está identificada con una letra (A o B) en la esquina superior derecha. Se muestran los valores de RMSF para TbTIM (negro), TcTIM (rojo) y hTIM (verde). En todos los casos se utilizó como referencia la estructura cristalina correspondiente a cada TIM.

valores más grandes de RMSF que sus homólogos en TbTIM y TcTIM. Además, los valores de RMSF en las zonas correspondientes a L3* y $\alpha 5$ de la subunidad A en la hTIM son claramente más grandes que en las otras dos estructuras. Es importante observar que excluyendo las zonas del lazo 5 y lazo 6 el resto de los residuos de la TcTIM presenta fluctuaciones menores que en la TbTIM y la hTIM.

La disposición espacial de los residuos en la interfase no es la única característica relevante en la estabilidad de la proteína sino que también la movilidad de esos residuos podría contribuir a la estabilidad dimérica. A través de un análisis de los contactos entre residuos se logró identificar la interacción entre la Ser71 y la Ser79 (numeración de la TcTIM) presente tanto en la TbTIM y la TcTIM pero carente en la hTIM en donde la posición 71 es ocupada por una Asp. De tal manera que en la hTIM no es posible la formación de ese puente de hidrógeno, lo cual en cierta manera

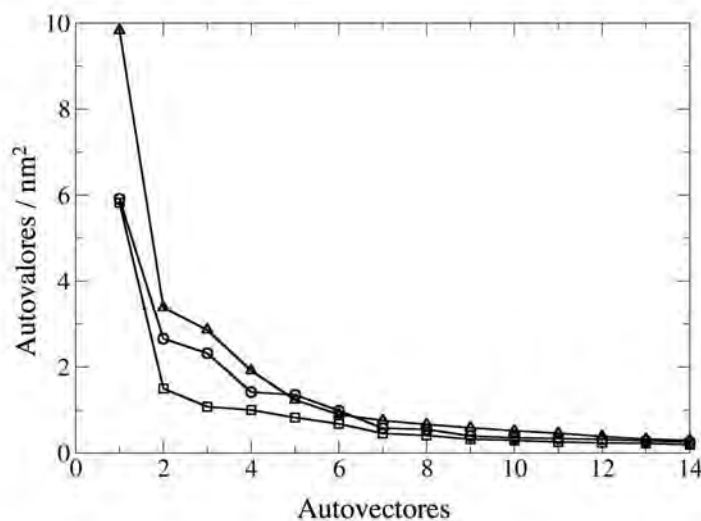


FIGURA 9.3.2. Autovalores correspondientes a los primeros movimientos extraídos mediante análisis de componentes principales. Los resultados se presentan para TbTIM (círculos), TcTIM (cuadrados) y hTIM (triángulos).

puede contribuir a la inestabilidad del lazo y a la mayor movilidad registrada en el RMSF de dicho segmento de la proteína.

Para cuantificar las diferencias entre movilidad de cada enzima se calcularon los componentes principales de la proteína en cada trayectoria (ver secciones 2.2.3 y 8.3). Los autovalores obtenidos de cada trayectoria que se grafican en la Figura 9.3.2 decrecen en el orden $hTIM > TbTIM > TcTIM$. La traza correspondiente a la matriz de covarianzas diagonalizada, que puede interpretarse como una medida de la movilidad global de la enzima, es de 19.18 nm^2 para TcTIM, 24.72 nm^2 para TbTIM y 31.61 nm^2 para hTIM. En algunos trabajos se defiende la hipótesis de que a mayor flexibilidad de la proteína menor será la estabilidad estructural y que esta combinación de factores se reflejan en la habilidad que la estructura tendrá para realizar su función. El desplegamiento térmico de TcTIM y TbTIM ha sido estudiado recientemente por calorimetría diferencial de barrido, indicando que la TcTIM es más estable que la TbTIM. La energía de activación para TcTIM es el doble que la de TbTIM mientras que la temperatura de transición es 4 K mayor para TcTIM (Costas *et al.*, 2009).

La velocidad máxima de conversión catalítica, V_{\max} , podría considerarse como una medida de la habilidad de la enzima para realizar su función. Por lo que una comparación de este parámetro podría apuntar a algunas diferencias entre especies.

Sin embargo entre las distintas TIMs estudiadas este valor no es muy diferente y no se observan claras correlaciones con la flexibilidad de la proteína. Específicamente, valores de V_{\max} de 7498 y 9670 $\mu\text{mol min}^{-1}\text{mol}^{-1}$ se han reportado para la TcTIM, de 9300 $\mu\text{mol min}^{-1}\text{mol}^{-1}$ para la TbTIM (Zárate-Pérez *et al.*, 2009), y de 7080 $\mu\text{mol min}^{-1}\text{mol}^{-1}$ para la hTIM (Rodríguez-Almazan *et al.*, 2008). Es claro que la flexibilidad de la enzima no es el único factor a considerar para comparar la función de las distintas especies. A pesar de que el análisis realizado a partir de las simulaciones de las proteínas en agua no proporciona una explicación de la estabilidad estructural y/o diferencias entre especies, éste muestra algunas de las diferencias en el comportamiento dinámico que puede estar relacionado con la habilidad catalítica y estabilidad en las diferentes estructuras.

9.4. Cálculo de la Energía libre

Como se explicó en la sección de metodología se generaron dos conjuntos de simulaciones para cada proteína. Cada uno de ellos consistió en simulaciones del proceso de monomerización de la TIM a una determinada velocidad de separación (10 Å/ns y 5 Å/ns) dando un total de 465 *ns* (Tabla 8.2.1). Para cada trayectoria se calculó el trabajo W_i requerido para aumentar la distancia de separación de los centros de masa de las dos subunidades en 1.5 nm (Figura 9.4.1).

La obtención de la diferencia de energía libre se realizó a través de la ecuación de Jarzynski. Además de este valor, se calculó también, para cada sistema, el error asociado. La metodología utilizada para evaluar el error, fue la propuesta por Gore y cols. (Gore *et al.*, 2003). Debido a la complejidad de su evaluación, ésta se realizó de tres maneras distintas para cada sistema. En general, para cada grupo de simulaciones se obtuvieron: el trabajo promedio $\langle W \rangle$ (kJ/mol), el *bias* B_J (kJ/mol), varianza σ_J^2 , estimador de Jarzynski ΔF_J y estimador de Jarzynski corregido ΔF_{Jc} (Tabla 9.4.1). Los valores $B_{J,1}$, $\Delta F_{Jc,1}$ y $B_{J,2}$, $\Delta F_{Jc,2}$ se calcularon mediante las ecuaciones propuestas por Gore y cols. para evaluar el error asociado (Sección 2.3.1).

Los resultados obtenidos se muestran en la Tabla 9.4.1 y las distribuciones correspondientes a los valores de trabajo se presentan en la Figura 9.4.1. Como se puede observar en las dos velocidades de separación empleadas el trabajo promedio $\langle W \rangle$ disminuye en el orden TcTIM > TbTIM > hTIM. Lo cual se puede apreciar en la Figura 9.4.2 al comparar las gráficas de distribuciones para cada velocidad y cada proteína. Si las transformaciones se hubiesen llevado a cabo con velocidades infinitamente lentas de separación, de tal manera que haya dado lugar a un proceso cuasiestático, la diferencia de energía libre quedaría descrita únicamente con el

promedio del trabajo $\langle W \rangle$ (kJ/mol). Sin embargo, al haberse realizado a una velocidad finita el estimador de Jarzynski resulta la manera más apropiada para extraer dicha propiedad. Para que estrictamente se obtenga el valor de energía usando este estimador, se requiere que la distribución de los valores del trabajo se ajusten a una gaussiana. Como se observa en la Figura 9.4.2 las gráficas de las distribuciones están bien definidas, además de que muestran diferencias entre sistemas. En este trabajo asumiremos que esas distribuciones son gaussianas y por tanto aplicaremos las ecuaciones de Jarzynski y de Gore. La ecuación de Jarzynski (Ec. 2.3.10) pondera con mayor intensidad los valores más bajos de W_i , por tanto, proporciona un mejor estimado de la diferencia de energía libre que el promedio $\langle W_i \rangle$. De la aplicación de esta ecuación, se observa los valores obtenidos de ΔF_J no corregidos decrecen en el mismo orden que $\langle W \rangle$ (ver Tabla 9.4.1). Asumiendo que la distribución de trabajo es gaussiana, y que las simulaciones se realizaron en un régimen cercano al equilibrio podemos utilizar las correcciones propuestas por Gore. Los valores de ΔF_c que se obtienen se presentan en la Tabla 9.4.1, puesto que el número de simulaciones realizadas es pequeño y que el trabajo disipado es grande (Ec. 2.3.14), se obtiene entonces que el *bias* (Ec. 2.3.18) es la principal fuente de error. Debido a esta sobreestimación del $\langle W_{dis} \rangle$ y del B_J , se obtiene un ΔF_c (Ec. 2.3.22) con valores negativos. Tal y como se describe en el artículo original de Gore y cols., las correcciones dadas por $B_{J,1}$ y $B_{J,2}$ representan una opción más apropiada para el cálculo del error en este caso. Los valores corregidos de energía libre se expresan entonces como $\Delta F_{Jc,2}$. Estos valores de la diferencia de energía libre, al igual que $\langle W \rangle$ y ΔF_J decrecen en el orden TcTIM >TbTIM >hTIM a ambas velocidades de separación. Lo anterior, puede interpretarse como que es más fácil separar las dos subunidades de la hTIM que las de TbTIM y TcTIM. Por tanto, la resistencia a la separación es menor en TbTIM que en TcTIM. Estos resultados resultan un tanto inesperados porque un estudio de la interfase, demuestra que el número de contactos es mayor en hTIM. No obstante, algunos resultados experimentales soportan los resultados aquí obtenidos. Recientemente, Cháñez-Cardenas y cols. (Cháñez-Cárdenas *et al.*, 2005) estimaron que la diferencia de energía libre del proceso de monomerización de la TcTIM es de $65.8 kJ/mol$. Mainfroid y cols. lograron estimar la diferencia de energía libre para la hTIM, obteniendo un valor de $59.8 kJ/mol$ (Mainfroid *et al.*, 1996). El cálculo del valor correspondiente para TbTIM no ha sido estimado con éxito, pese a varios intentos (Cháñez-Cárdenas *et al.*, 2002). A pesar de que los resultados obtenidos por simulaciones moleculares coinciden con lo que se ha reportado en la literatura desde el punto de vista cualitativo, los valores numéricos obtenidos experimentalmente son

Vel. de separación	Sistema	$\langle W \rangle$	ΔF_J	B_J	σ_J^2	ΔF_{Jc}	$B_{J,1}$	$\Delta F_{Jc,1}$	$B_{J,2}$	$\Delta F_{Jc,2}$
10 Å/ns	TcTIM	1531	1282	2025	10041	-753	207	1075	409	873
	TbTIM	1386	1170	1238	6138	-68	176	994	347	823
	hTIM	1304	1103	1550	7687	-447	161	942	317	787
5 Å/ns	TcTIM	1308	1116	1485	7364	-369	153	964	300	816
	TbTIM	1134	970	1448	7179	-478	127	843	249	720
	hTIM	1080	856	1564	7754	-708	185	671	364	491

TABLA 9.4.1. Trabajo promedio $\langle W \rangle$ (kJ/mol), Trabajo disipado $\langle W_{dis} \rangle$ (kJ/mol), Bias B_J (kJ/mol), Varianza σ_J^2 (kJ/mol^2), Estimador de Jarzynski ΔF_J y Estimador de Jarzynki corregido ΔF_{Jc} (kJ/mol).

mucho menores. La razón de esta diferencia radica principalmente en la metodología aplicada, ya que los valores absolutos del trabajo aumentan linealmente con la velocidad de separación y nos encontramos muy lejos de una velocidad comparable al equilibrio. Además, es necesario tomar en cuenta el diseño del experimento. Mientras que la disociación dimérica en experimentos se lleva usualmente hasta la separación total de las dos subunidades en este trabajo se incrementó la separación entre las dos subunidades en 15 Å. Una gráfica de los valores de ΔF_{J2} en función de la distancia de separación entre monómeros muestra que entre 12 Å y 13 Å la pendiente comienza a disminuir. Sin embargo, la estabilización de ΔF_{J2} parece ocurrir más allá de los 15 Å (Figura 9.4.3). Lograr una separación mayor es posible en un proyecto ambicioso, dado que se requerirían cajas de simulación más grandes con mayor cantidad de disolvente, y también sería necesario la utilización de más recursos de cómputo. No obstante, como se observa en la Figura 9.4.3 en el intervalo de separación entre monómeros estudiado se observa para algunas proteínas cierta convergencia como en el caso de la hTIM. Además de que se pueden encontrar diferencias entre especies. Cabe señalar que todos los resultados presentados en esta sección demuestran el buen funcionamiento de la metodología aplicada, por lo que cualitativamente pueden ser utilizados para describir la separación monomérica de las especies estudiadas.

9.5. Análisis visual de las trayectorias

Para ilustrar los cambios estructurales que se llevan a cabo a nivel intramolecular a lo largo del proceso de monomerización se muestran imágenes de la enzima hTIM

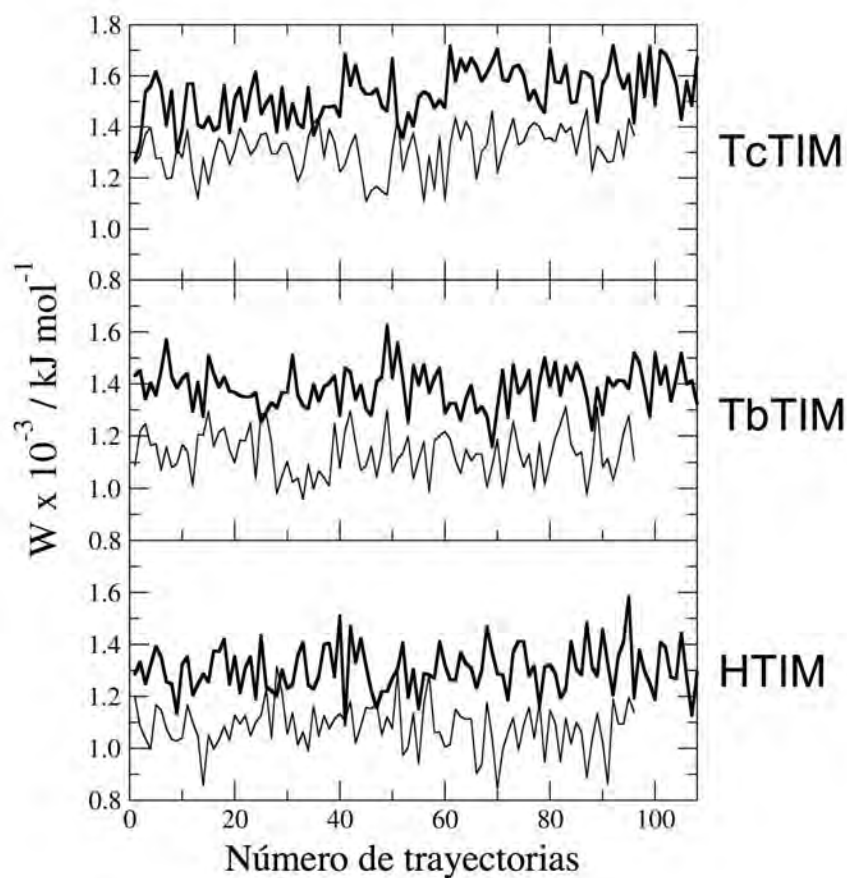


FIGURA 9.4.1. Trabajo W del proceso de monomerización obtenido de la separación de las subunidades a 10 \AA/ns (negro) y 5 \AA/ns (gris) para cada enzima.

en función de λ (Figura 9.5.1). Durante el proceso de disociación los residuos de la interfase cambian su posición hasta que los monómeros se separan. El proceso comienza con la elongación del lazo 3 mientras el resto del barril se aleja de forma casi uniforme de la subunidad adyacente. En el estado final a pesar de que la distancia entre los c.o.m. de ambos barriles se incrementó en 1.5 nm , algunos residuos del lazo 3 aún continúan en contacto con el otro monómero. Esto es posible gracias a la interacción de los residuos del lazo 3 con el residuo en la posición 14, ocupado por una Cys en TcTIM y TbTIM y Met en la hTIM, como se puede observar en la Figura 9.5.1.

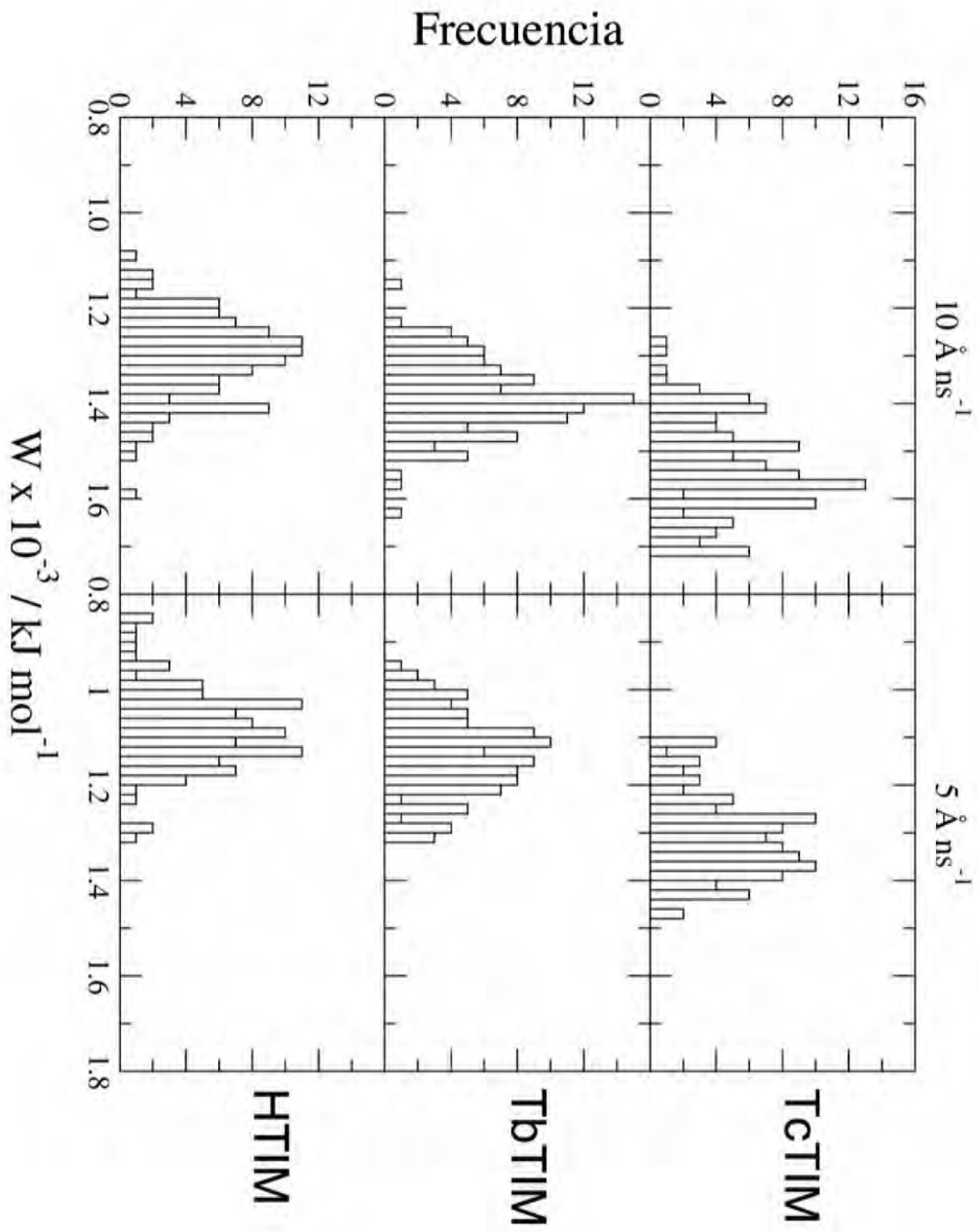


FIGURA 9.4.2. Distribuciones del trabajo W obtenido de la separación de los monómeros a 10 \AA/ns y 5 \AA/ns para cada enzima.

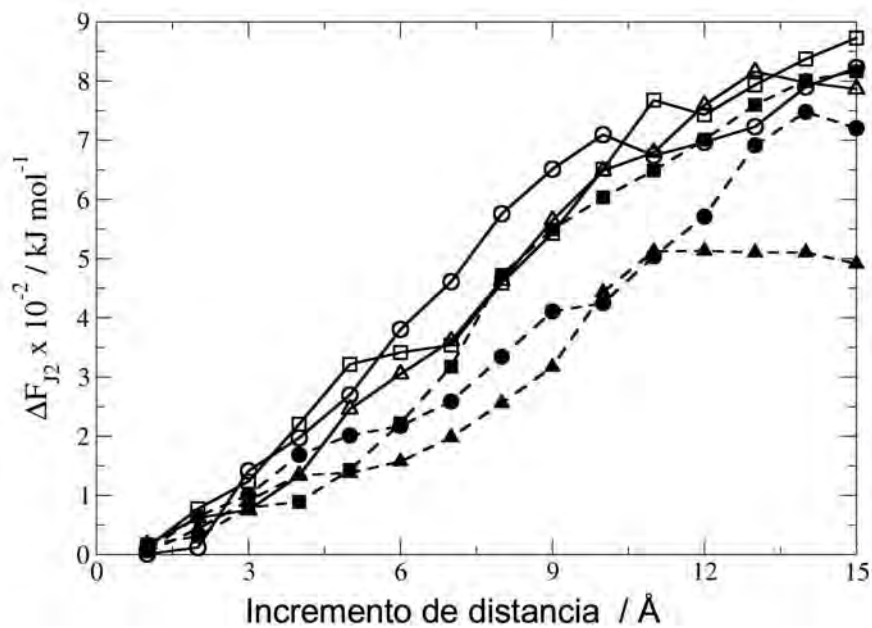


FIGURA 9.4.3. Diferencia de energía libre ΔF_{J_2} en función del incremento de la distancia de separación entre centros de masas de los monómeros de cada proteína a velocidades de 10 \AA/ns (líneas sólidas) y 5 \AA/ns (líneas punteadas). Los resultados se presentan para TbTIM (círculos), TcTIM (cuadrados) y hTIM (triángulos).

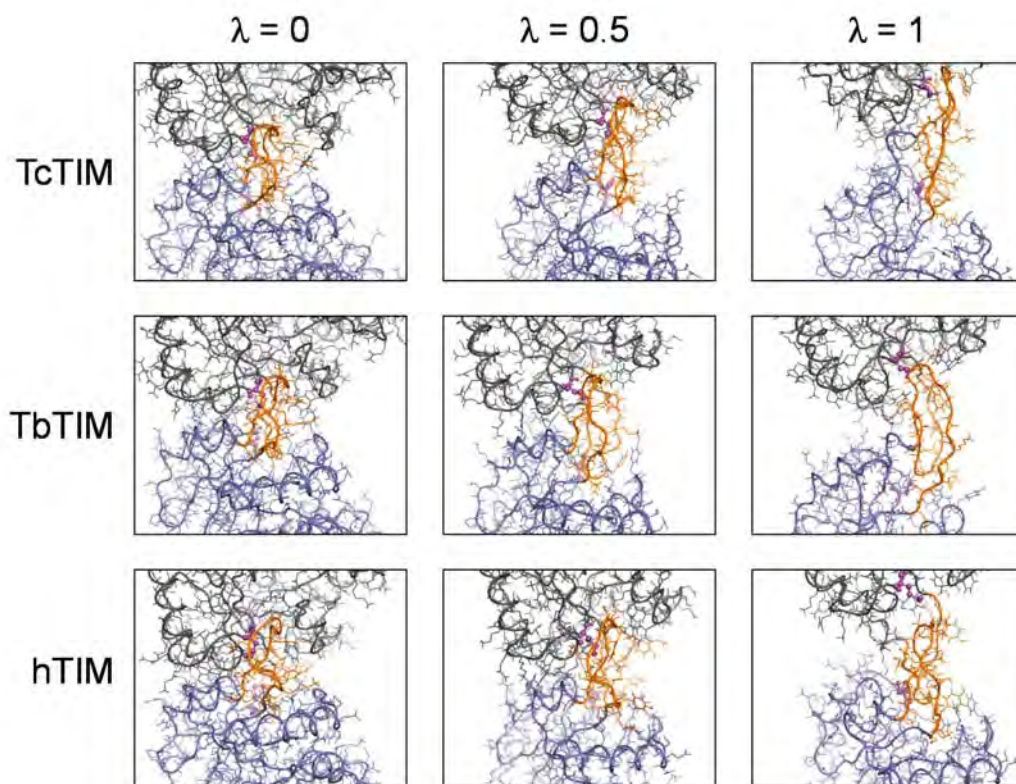


FIGURA 9.5.1. Imagen transversal de la interfase de cada proteína durante el proceso de separación $\lambda = 0$ y $\lambda = 1$, corresponden a las configuraciones inicial y final de la trayectoria de monomerización, respectivamente. Se muestra a la subunidad A (gris) y B (azul), lazo 3 (naranja) para cada proteína, y en color magenta el residuo Met14 en hTIM, y, Cys15 en TbTIM y TcTIM.

CAPÍTULO 10

Conclusiones

Se calculó la energía libre asociada al proceso de disociación de las subunidades de la TIM de 3 especies (TcTIM, TbTIM y hTIM) a través de la aproximación de Jarzynski. Además, se realizó un estudio comparativo de la movilidad global y por residuos fue realizado en este trabajo. Se evaluaron los cambios conformacionales experimentados en la interfase de la TcTIM, TbTIM y hTIM a lo largo del proceso de monomerización y se compararon los resultados obtenidos para las distintas especies. Los resultados sugieren que la TcTIM, es la proteína que presenta una menor movilidad en solución, y que, es la hTIM quien mayor flexibilidad presentó bajo las condiciones de estudio. Por otro lado, la energía libre obtenida de la monomerización de la TcTIM es mayor que las especies homólogas estudiadas, mientras que la hTIM es quien menor resistencia opone ante la separación monomérica. En general, la técnica empleada permite obtener valores relativos de energía libre y de estudiar las interacciones relevantes entre monómeros a lo largo del proceso. El análisis estructural, permitió ubicar la interacción entre el lazo 3 y la Cys15 (Met14 en hTIM) como una de las más estables dentro de la proteína, en congruencia con lo que se reporta en la literatura, siendo una interacción responsable en gran medida de la estabilización dimérica. En este sentido, también la interacción Ser71-Ser79 parece ser relevante para mantener la integridad estructural del lazo 3, al estar ausente en la hTIM, se espera una mayor vulnerabilidad del lazo, por tanto, se sugiere que podría contribuir a la desestabilización del dímero.

CAPÍTULO 11

Conclusiones Generales

Con este trabajo se logró describir el comportamiento dinámico de la triosafosfato isomerasa de *Trypanosoma cruzi* en un medio orgánico como el decano y en mezclas agua/decano en escalas de tiempo del orden de varias decenas de nanosegundos con resolución atómica. Se encontró que en presencia de un solvente apolar la proteína modifica su dinámica y su estructura para adaptarse al medio en el que se encuentra. En general se observó que en estas condiciones la proteína incrementa su flexibilidad, y que ésta es comparable con la que presenta en agua pura siempre y cuando se mantenga un número mínimo de moléculas de agua en contacto con la parte externa de su estructura. El análisis de componentes principales para la triosafosfato isomerasa de *Trypanosoma cruzi*, *Trypanosoma brucei* y *humana*, permitió identificar los movimientos de mayor amplitud dentro de la dinámica de la enzima. Los resultados, que concuerdan para las tres proteínas de estudio, demostraron que las zonas de mayor movilidad se localizan en los residuos que conforman los lazos 5 y 6. De forma general, se encontró que existe un movimiento concertado entre las dos subunidades de la estructura dimérica. Las investigaciones realizadas permitieron hacer una comparación de la movilidad global y por residuos de las tres especies de TIM. Se logró indentificar que la hTIM presenta una movilidad significativamente mayor que las enzimas de los dos parásitos. Las estimaciones de las diferencias en la energía libre de monomerización calculadas mediante el método de Jarzynski indican que la TcTIM presenta más resistencia a la ruptura de los dímeros que las estructuras de las otras dos especies, siendo la hTIM la enzima para la que es más sencillo separar las dos subunidades. El análisis detallado de los residuos en la interfase, permitió ubicar la interacción entre el lazo 3 de un monómero con la Cys15 (Met14 en hTIM) del monómero adyacente como una de las más estables y que más contribuye para la estabilidad dimérica dentro de la proteína, en congruencia con lo que se reporta en la literatura. En este sentido, también la interacción Ser71-Ser79 parece ser relevante para mantener la integridad estructural del lazo 3. Al estar ausente en la hTIM, se espera una mayor vulnerabilidad del lazo, por tanto, se sugiere que podría contribuir a la desestabilización del dímero. En general, los resultados

que aquí se presentan aportan detalles del comportamiento dinámico y de la estabilidad de la enzima triosafosfato isomerasa para las diferentes especies y en diferentes condiciones.

Bibliografía

- ALBERY, W.J. Y KNOWLES, J.R. (1976). Evolution of enzyme function and the development of catalytic efficiency. *Biochemistry*, **15**, 5631–40.
- AMADEI, A., LINSSEN, A.B. Y BERENDSEN, H.J. (1993). Essential dynamics of proteins. *Proteins*, **17**, 412–425.
- BASTUG, T., CHEN, P.C., PATRA, S.M. Y KUYUCAK, S. (2008). Potential of mean force calculations of ligand binding to ion channels from jarzynski equality and umbrella sampling. *J. Chem. Phys.*, **128**, 155104.
- BELL, G., HALLING, P.J., MOORE, B.D., PARTRIDGE, J. Y REES, D.G. (1995). Biocatalyst behaviour in low-water systems. *Trends in Biotechnol.*, **13**, 468–473.
- BERENDSEN, H., POSTMA, J., VAN GUNSTEREN, W. Y HERMANS, J. (1981). Interaction models for water in relation to protein hydration. *Intermolecular Forces*, **342**, 331.
- BERENDSEN, H.J.C., POSTMA, J.P.M., VAN GUNSTEREN, W.F., DINOLA, A. Y HAAK, J.R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, **81**, 3684–3690.
- BERENDSEN, H.J.C., VAN DER SPOEL, D. Y VAN DRUNEN, R. (1995). Gromacs: A message-passing parallel molecular dynamics implementation. *Phys. Comm.*, **95**, 43–56.
- BLACKLOW, S.C., RAINES, R.T., LIM, W.A., ZAMORE, P.D. Y KNOWLES, J.R. (1988). Triosephosphate isomerase catalysis is diffusion controlled. appendix: Analysis of triose phosphate equilibria in aqueous solution by ^{31}P nmr. *Biochemistry*, **27**, 1158–67.
- BORCHERT, T.V., ABAGYAN, R., JAENICKE, R. Y WIERENGA, R.K. (1994). Design, creation, and characterization of a stable, monomeric triosephosphate isomerase. *Proc. Natl. Acad. Sci. USA.*, **91**, 1515–8.
- CAFFREY, D.R., SOMAROO, S., HUGHES, J.D., MINTSERIS, J. Y HUANG, E.S. (2004). Are protein-protein interfaces more conserved in sequence than the rest of the protein surface? *Protein Sci.*, **13**, 190–202.

- CANSU, S. Y DORUKER, P. (2008). Dimerization affects collective dynamics of triosephosphate isomerase. *Biochemistry*, **47**, 1358–68.
- CHÁNEZ-CÁRDENAS, M.E., FERNÁNDEZ-VELASCO, D.A., VÁZQUEZ-CONTRERAS, E., CORIA, R., SAAB-RINCÓN, G. Y PÉREZ-MONTFORT, R. (2002). Unfolding of triosephosphate isomerase from trypanosoma brucei: identification of intermediates and insight into the denaturation pathway using tryptophan mutants. *Arch. Biochem. Biophys.*, **399**, 117–29.
- CHÁNEZ-CÁRDENAS, M.E., PÉREZ-HERNÁNDEZ, G., SÁNCHEZ-REBOLLAR, B.G., COSTAS, M. Y VÁZQUEZ-CONTRERAS, E. (2005). Reversible equilibrium unfolding of triosephosphate isomerase from trypanosoma cruzi in guanidinium hydrochloride involves stable dimeric and monomeric intermediates. *Biochemistry*, **44**, 10883–92.
- COLOMBO, G., TOBA, S. Y MERZ, K.M. (1999). Rationalization of the enantioselectivity of subtilisin in DMF. *J. Am. Chem. Soc.*, **121**, 3486–3493.
- CONNOLLY, M.L. (1983). Analytical molecular surface calculation. *J. Appl. Crystallogr.*, **16**, 548–558.
- COSTAS, M., RODRIGUEZ-LARREA, D., DE MARIA, L., BORCHERT, T.V., GOMEZ-PUYOU, A. Y SANCHEZ-RUIZ, J.M. (2009). Between-species variation in the kinetic stability of tm proteins linked to solvation-barrier free energies. *J. Mol. Biol.*, **385**, 924–937.
- CUENDET, M.A. Y MICHIELIN, O. (2008). Protein-protein interaction investigated by steered molecular dynamics: the tcr-pmhc complex. *Biophys. J.*, **95**, 3575–90.
- DAGGETT, V. (2006). Protein folding-simulation. *Chem Rev*, **106**, 1898–1916.
- DAURA, X., MARK, A.E. Y GUNSTEREN, W.F.V. (1998). Parametrization of aliphatic chn united atoms of GROMOS96 force field. *J. Comput. Chem.*, **19**, 535–547.
- DAURA, X., VAN GUNSTEREN, W.F. Y MARK, A.E. (1999). Folding-unfolding thermodynamics of a beta-heptapeptide from equilibrium simulations. *Proteins*, **34**, 269–280.
- DELANO, W.L. (2008). The pymol molecular graphics system.
- DIAZ-VERGARA, N. Y PINEIRO, A. (2008). Molecular dynamics study of triosephosphate isomerase from trypanosoma cruzi in water/decane mixtures. *J. Phys. Chem. B*, **112**, 3529–39.
- DÍAZ-VERGARA, N.D. (2004). Estudio de la triosafosfato isomerasa de tripanosoma cruzi en mezclas agua-decano por medio de dinámica molecular.
- DORDICK, J.S. (1988). Biocatalysis in nonaqueous media. patents and literature. *Appl. Biochem. Biotech.*, **19**, 103–112, PMID: 3060012.

- FARBER, G.K. Y PETSKO, G.A. (1990). The evolution of alpha/beta barrel enzymes. *Trends in Biochem. Sci.*, **15**, 228–34.
- FINNEY, J.L. Y POOLE, P.L. (1984). Protein hydration and enzyme activity: the role of hydrogen-induced conformation and dynamic changes in the activity of lysozyme. *Comments Mol. Cell. Biophys.*, **2**, 129–151.
- FRENKEL, D. Y SMIT, B. (2002). *Understanding molecular simulation*. Academic Press.
- GAO, X.G., MALDONADO, E., PEREZ-MONTFORT, R., GARZA-RAMOS, G., DE GOMEZ-PUYOU, M.T., GOMEZ-PUYOU, A. Y RODRIGUEZ-ROMERO, A. (1999). Crystal structure of triosephosphate isomerase from trypanosoma cruzi in hexane. *Proc. Natl. Acad. Sci. USA.*, **96**, 10062–10067.
- GARZA-RAMOS, G., TUENA DE GOMEZ-PUYOU, M., GOMEZ-PUYOU, A. Y GRACY, R.W. (1992). Dimerization and reactivation of triosephosphate isomerase in reverse micelles. *Eur. J. Biochem.*, **208**, 389–395.
- GARZA-RAMOS, G., CABRERA, N., SAAVEDRA-LIRA, E., TUENA DE GOMEZ-PUYOU, M., OSTOA-SALOMA, P., PEREZ-MONTFORT, R. Y GOMEZ-PUYOU, A. (1998). Sulfhydryl reagent susceptibility in proteins with high sequence similarity—triosephosphate isomerase from trypanosoma brucei, trypanosoma cruzi and leishmania mexicana. *Eur. J. Biochem.*, **253**, 684–691.
- GONNELLI, M. Y STRAMBINI, G.B. (1988). Protein dynamical structure by tryptophan phosphorescence and enzymatic activity in reverse micelles. 2. alkaline phosphatase. *J. Phys. Chem.*, **92**, 2854–2857.
- GORE, J., RITORT, F. Y BUSTAMANTE, C. (2003). Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. *Proc. Natl. Acad. Sci. USA.*, **100**, 12564–9.
- HESS, B., BEKKER, H., BERENDSEN, H.J.C. Y FRAAIJE, J.G.E.M. (1997). Lincs: A linear constraint solver for molecular simulations. *J. Comput. Chem.*, **18**, 1463–1472.
- HORNAK, V., OKUR, A., RIZZO, R.C. Y SIMMERLING, C. (2006a). Hiv-1 protease flaps spontaneously close to the correct structure in simulations following manual placement of an inhibitor into the open state. *J. Am. Chem. Soc.*, **128**, 2812–3.
- HORNAK, V., OKUR, A., RIZZO, R.C. Y SIMMERLING, C. (2006b). Hiv-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA.*, **103**, 915–20.
- HUMMER, G. Y SZABO, A. (2001). Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proc. Natl. Acad. Sci. USA.*, **98**, 3658–61.

- HUMPHREY, W., DALKE, A. Y SCHULTEN, K. (1996). Vmd: visual molecular dynamics. *J. Mol. Grap.*, **14**, 33–8, 27–8.
- JARZYNSKI, C. (1997a). Equilibrium free-differences from equilibrium measurements: a master-equation approach. *Phys. Rev. E*, **56**, 5018–5035.
- JARZYNSKI, C. (1997b). Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, **78**, 2690–2693.
- JONES, K.E., PATEL, N.G., LEVY, M.A., STOREYGARD, A., BALK, D., GITTLEMAN, J.L. Y DASZAK, P. (2008). Global trends in emerging infectious diseases. *Nature*, **451**, 990–993.
- KESKIN, O., GURSOY, A., MA, B. Y NUSSINOV, R. (2008). Principles of protein-protein interactions: what are the preferred ways for proteins to interact? *Chem. Rev.*, **108**, 1225–1244.
- KLIBANOV, A. (1997). Why are enzymes less active in organic solvents than in water? *Trends in Biotechnol.*, **15**, 97–101.
- KLIBANOV, A.M. (1989). Enzymes that work in organic solvents. *Chemtech*, 354–359.
- KNOWLES, J.R. (1991). Enzyme catalysis: not different, just better. *Nature*, **350**, 121–4.
- KUSALIK, P.G. Y SVISHCHEV, I.M. (1994). The spatial structure in liquid water. *Science*, **265**, 1219–1221.
- LEVY, Y. Y ONUCHIC, J.N. (2004). Water and proteins: A love–hate relationship. *Proc. Natl. Acad. Sci. USA.*, **101**, 3325–3326.
- LINDAHL, E., HESS, B. Y VAN DER SPOEL, D. (2001). Gromacs 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.*, **7**, 306–317.
- MAINFROID, V., TERPSTRA, P., BEAUREGARD, M., FRERE, J.M., MANDE, S.C., HOL, W.G., MARTIAL, J.A. Y GORAJ, K. (1996). Three htim mutants that provide new insights on why tim is a dimer. *J. Mol. Biol.*, **257**, 441–56.
- MALDONADO, E., SORIANO-GARCIA, M., MORENO, A., CABRERA, N., GARZARAMOS, G., DE GOMEZ-PUYOU, M., GOMEZ-PUYOU, A. Y PEREZ-MONTFORT, R. (1998). Differences in the intersubunit contacts in triosephosphate isomerase from two closely related pathogenic trypanosomes. *J. Mol. Biol.*, **283**, 193–203.
- MATTOS, C. Y RINGE, D. (1996 May). Locating and characterizing binding sites on proteins. *Nat. Biotechnol.*, **14**, 595–599.
- MICAELO, N.M. Y SOARES, C.M. (2007). Modeling hydration mechanisms of enzymes in nonpolar and polar organic solvents. *FEBS Journal*, **274**, 2424–2436.
- MICAELO, N.M., TEIXEIRA, V.H., BAPTISTA, A.M. Y SOARES, C.M. (2005).

- Water dependent properties of cutinase in nonaqueous solvents: A computational study of enantioselectivity. *Biophys. J.*, **89**, 999–1008.
- MIYAMOTO, S. Y KOLLMAN, P.A. (1992). Settle: An analytical version of the shake and rattle algorithm for rigid water models. *J. Comput. Chem.*, **13**, 952–962.
- NORIN, M., HAEFFNER, F., HULT, K. Y EDHOLM, O. (1994). Molecular dynamics simulations of an enzyme surrounded by vacuum, water, or a hydrophobic solvent. *Biophys. J.*, **67**, 548–559.
- OLIVARES-ILLANA, V., PEREZ-MONTFORT, R., LOPEZ-CALAHORRA, F., COSTAS, M., RODRIGUEZ-ROMERO, A., TUENA DE GOMEZ-PUYOU, M. Y GOMEZ PUYOU, A. (2006). Structural differences in triosephosphate isomerase from different species and discovery of a multitypanosomatid inhibitor. *Biochemistry*, **45**, 2556–60.
- OLIVARES-ILLANA, V., ARREOLA, R., PUYOU, A.G. Y PEREZ-MONTFORT, R. (2008). *Advances in Protein Physical Chemistry*.
- OOSTENBRINK, C., VILLA, A., MARK, A.E. Y VAN GUNSTEREN, W.F. (2004). A biomolecular force field based on the free enthalpy of hydration and solvation: the gromos force-field parameter sets 53a5 and 53a6. *J. Comput. Chem.*, **25**, 1656–76.
- ORNSTEIN, R.L. Y YA-JUN, Z. (1996). A molecular dynamics study of the effect of carbon tetrachloride on enzyme structure and dynamics: subtilisin. *Protein Eng.*, **9**, 485–492.
- PAL, S.K. Y ZEWAİL, A.H. (2004). Dynamics of water in biological recognition. *Chem. Rev.*, **104**, 2099–2124.
- PEIMBERT, M., DOMÍNGUEZ-RAMÍREZ, L., TELLEZ, L.A., BLANCAS, L.M. Y FERNÁNDEZ-VELASCO, D.A. (2008). *Advances in Protein Physical Chemistry*.
- PEREZ-MONTFORT, R., GARZA-RAMOS, G., ALCANTARA, G.H., REYES-VIVAS, H., GAO, X.G., MALDONADO, E., DE GOMEZ-PUYOU, M.T. Y GOMEZ-PUYOU, A. (1999). Derivatization of the interface cysteine of triosephosphate isomerase from trypanosoma brucei and trypanosoma cruzi as probe of the interrelationship between the catalytic sites and the dimer interface. *Biochemistry*, **38**, 4114–4120.
- PETERS, G., VAN AALTEN, D., EDHOLM, O., TOXVAERD, S. Y BYWATER, R. (1996). Dynamics of proteins in different solvent systems: analysis of essential motion in lipases. *Biophys. J.*, **71**, 2245–2255.
- POMPLIANO, D.L., PEYMAN, A. Y KNOWLES, J.R. (1990). 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.

- REYES-VIVAS, H., HERNANDEZ-ALCANTARA, G., LOPEZ-VELAZQUEZ, G., CABRERA, N., PEREZ-MONTFORT, R., DE GOMEZ-PUYOU, M.T. Y GOMEZ-PUYOU, A. (2001). Factors that control the reactivity of the interface cysteine of triosephosphate isomerase from trypanosoma brucei and trypanosoma cruzi. *Biochemistry*, **40**, 3134–3140.
- RINGE, D. (1995 Dec). What makes a binding site a binding site? *Curr. Opin. Struct. Biol.*, **5**, 825–829.
- RODRIGUEZ-ALMAZAN, C., ARREOLA, R., RODRIGUEZ-LARREA, D., AGUIRRE-LOPEZ, B., DE GOMEZ-PUYOU, M.T., PEREZ-MONTFORT, R., COSTAS, M., GOMEZ-PUYOU, A. Y TORRES-LARIOS, A. (2008). Structural basis of human triosephosphate isomerase deficiency: mutation e104d is related to alterations of a conserved water network at the dimer interface. *J. Biol. Chem.*, **283**, 23254–63.
- ROZOVSKY, S. Y MCDERMOTT, A.E. (2001). The time scale of the catalytic loop motion in triosephosphate isomerase. *J. Mol. Biol.*, **310**, 259–270.
- ROZOVSKY, S., JOGL, G., TONG, L. Y MCDERMOTT, A.E. (2001). Solution-state nmr investigations of triosephosphate isomerase active site loop motion: ligand release in relation to active site loop dynamics. *J. Mol. Biol.*, **310**, 271–280.
- RUPLEY, J.A. Y CARERI, G. (1991). Protein hydration and function. *Adv. Protein Chem.*, **41**, 37–172, PMID: 2069077.
- RUPLEY, J.A., GRATTPN, E. Y CARERI, G. (1983). Water and globular proteins. *Trends Biochem. Sci.*, **8**, 18.
- SCHMUNIS, G.A. (2007). The globalization of chagas disease. *ISBT Science Series*, **2**, 6–11.
- SCHULER, L.D. Y GUNSTEREN, W.F.V. (2000). On the choice of dihedral angle potential energy functions for n-Alkanes. *Mol. Simulat.*, **25**, 301.
- SINGH, S.K., MAITHAL, K., BALARAM, H. Y BALARAM, P. (2001). Synthetic peptides as inactivators of multimeric enzymes: inhibition of plasmodium falciparum triosephosphate isomerase by interface peptides. *FEBS Letters*, **501**, 19–23.
- SOARES, C.M., TEIXEIRA, V.H. Y BAPTISTA, A.M. (2003). Protein structure and dynamics in nonaqueous solvents: Insights from molecular dynamics simulation studies. *Biophys. J.*, **84**, 1628–1641.
- STRAMBINI, G.B. Y GONNELLI, M. (1988). Protein dynamical structure by tryptophan phosphorescence and enzymatic activity in reverse micelles. 1. liver alcohol dehydrogenase. *J. Phys. Chem.*, **92**, 2850–2853.
- TELLEZ-VALENCIA, A., AVILA-RIOS, S., PEREZ-MONTFORT, R., RODRIGUEZ-ROMERO, A., TUENA DE GOMEZ-PUYOU, M., LOPEZ-CALAHORRA, F. Y

- GOMEZ-PUYOU, A. (2002). Highly specific inactivation of triosephosphate isomerase from trypanosoma cruzi. *Biochem. Biophys. Res. Commun.*, **295**, 958–963.
- TELLEZ-VALENCIA, A., OLIVARES-ILLANA, V., HERNANDEZ-SANTOYO, A., PEREZ-MONTFORT, R., COSTAS, M., RODRIGUEZ-ROMERO, A., LOPEZ-CALAHORRA, F., TUENA DE GOMEZ-PUYOU, M. Y GOMEZ-PUYOU, A. (2004). Inactivation of triosephosphate isomerase from trypanosoma cruzi by an agent that perturbs its dimer interface. *J. Mol. Biol.*, **341**, 1355–1365.
- TOBA, S., HARTSOUGH, D.S. Y MERZ, K.M. (1996). Solvation and dynamics of chymotrypsin in hexane. *J. Am. Chem. Soc.*, **118**, 6490–6498.
- TUENA DE GOMEZ-PUYOU, M. Y GOMEZ-PUYOU, A. (1998). Enzymes in low water systems. *Crit. Rev. Biochem. Mol. Biol.*, **33**, 53–89.
- VAN DER SPOEL, D., LINDAHL, E., HESS, B., GROENHOF, G., MARK, A.E. Y BERENDSEN, H.J. (2005). Gromacs: fast, flexible, and free. *J. Comput. Chem.*, **26**, 1701–18.
- WALEY, S.G. (1973). Refolding of triose phosphate isomerase. *Biochem. J.*, **135**, 165–72.
- WWW.GROMACS.ORG (GROMACS).
- WWW.PDB.ORG (PDB).
- WWW.WHO.ORG (WHO).
- XIONG, H., CRESPO, A., MARTI, M., ESTRIN, D. Y ROITBERG, A.E. (2006). Free energy calculations with non-equilibrium methods: applications of the jarzynski relationship. *Theor. Chem. Acc.*, **116**, 338–346.
- YANG, L., DORDICK, J.S. Y GARDE, S. (2004). Hydration of enzyme in nonaqueous media is consistent with solvent dependence of its activity. *Biophys. J.*, **87**, 812–821.
- ZAKS, A. Y KLIBANOV, A. (1984). Enzymatic catalysis in organic media at 100 degrees c. *Science*, **224**, 1249–1251.
- ZAKS, A. Y KLIBANOV, A. (1985). Enzyme-catalyzed processes in organic solvents. *Proc. Natl. Acad. Sci. USA.*, **82**, 3192–3196.
- ZAKS, A. Y KLIBANOV, A. (1988a). The effect of water on enzyme action in organic media. *J. Biol. Chem.*, **263**, 8017–8021.
- ZAKS, A. Y KLIBANOV, A. (1988b). Enzymatic catalysis in nonaqueous solvents. *J. Biol. Chem.*, **263**, 3194–3201.
- ZAKS, A. Y KLIBANOV, A.M. (1986). Substrate specificity of enzymes in organic solvents vs. water is reversed. *J. Am. Chem. Soc.*, **108**, 2767–2768.
- ZÁRATE-PÉREZ, F., CHÁNEZ-CÁRDENAS, M.E., ARREOLA, R., TORRES-LARIOS,

- A. Y VÁZQUEZ-CONTRERAS, E. (2009). Different catalytic properties of two highly homologous triosephosphate isomerase monomers. *Biochem. Biophys. Res. Commun.*, **382**, 626–30.
- ZHENG, Y. Y ORNSTEIN, R.L. (1996a). A molecular dynamics and quantum mechanics analysis of the effect of DMSO on enzyme structure and dynamics: Subtilisin. *J. Am. Chem. Soc.*, **118**, 4175–4180.
- ZHENG, Y. Y ORNSTEIN, R.L. (1996b). Molecular dynamics of subtilisin carlsberg in aqueous and nonaqueous solutions. *Biopolymers*, **38**, 791–799.
- ZOMOSA-SIGNORET, V., HERNANDEZ-ALCANTARA, G., REYES-VIVAS, H., MARTINEZ-MARTINEZ, E., GARZA-RAMOS, G., PEREZ-MONTFORT, R., TUENA DE GOMEZ-PUYOU, M. Y GOMEZ-PUYOU, A. (2003). Control of the reactivation kinetics of homodimeric triosephosphate isomerase from unfolded monomers. *Biochemistry*, **42**, 3311–3318.
- ZOMOSA-SIGNORET, V., AGUIRRE-LOPEZ, B., HERNANDEZ-ALCANTARA, G., PEREZ-MONTFORT, R., DE GOMEZ-PUYOU, M.T. Y GOMEZ-PUYOU, A. (2007). Crosstalk between the subunits of the homodimeric enzyme triosephosphate isomerase. *Proteins*, **67**, 75–83.
- ZWANZIG, R.W. (1954). High-temperature equation of state by a perturbation method. i. nonpolar gases. *J. Chem. Phys.*, **22**, 1420–1426.

Parte 3

Artículo publicado

Molecular Dynamics Study of Triosephosphate Isomerase from *Trypanosoma cruzi* in Water/Decane Mixtures. Norma Díaz-Vergara y Ángel Piñeiro. *Journal of Physical Chemistry B*, 2008, 112, 3529-3539.

Molecular Dynamics Study of Triosephosphate Isomerase from *Trypanosoma cruzi* in Water/Decane Mixtures

Norma Díaz-Vergara and Ángel Piñeiro*

Departamento de Fisicoquímica, Facultad de Química, Universidad Nacional Autónoma de México, UNAM, Ciudad Universitaria 04510, México D.F., Mexico

Received: October 22, 2007; In Final Form: December 10, 2007

A comprehensive study of the triosephosphate isomerase from the parasite *Trypanosoma cruzi* (TcTIM) in water, in decane, and in three water/decane mixtures was performed using molecular dynamics (MD) simulations in a time scale of 40 ns. The structure and dynamics of the enzyme, as well as the solvent molecules' distribution and mobility, were analyzed in detail. In the presence of decane, the amplitudes of the most important internal motions of the enzyme backbone were observed to depend on the solvent concentration: the higher the water concentration, the greater the amplitudes. Contrary to this trend, the amplitudes of the TcTIM motions in pure water were similar to those of the simulation with the lowest water concentration. The enzyme was observed to be almost motionless in pure decane due to a sharp increase of the number of intramolecular hydrogen bonds. This caused a contraction of the enzyme structure accompanied by a loss of secondary structure and of a decrease of the hydrophilic solvent accessible surface. A similar behavior, although to a lesser extent, was observed in the simulation at the lowest water concentration. Our results suggest that the presence of decane molecules located at specific sites of the enzyme might accelerate its internal movements, although a minimum number of water molecules is needed for the protein to keep its structure and dynamics. Altogether, this work provides new insight into protein and water behavior in organic solvents as well as into the dynamics of TcTIM itself.

1. Introduction

An interesting relationship connecting rigidity, thermal stability, and catalytic activity of enzymes has been proposed by several authors in the past: the higher the flexibility and activity of the enzyme, the lower its unfolding temperature is.¹ Therefore, it appears that, throughout evolution, proteins have sought a balance between flexibility and thermal stability, which allows them to achieve their biological function and, at the same time, to preserve a well-defined average structure. Water plays an essential role in this context.^{2,3} It is well-known that water molecules located in the vicinity of a solute show differences in their organization and hence in their thermodynamic, dynamic, and electrostatic properties, when compared to bulk waters.^{4–6} The water behavior at a molecular level can be strongly affected by the presence of cosolvents, and this is critical for the dynamics of macromolecules.

In the middle 1980s, Zaks and Klibanov found that the catalytic activity of several enzymes, when they are suspended in hydrophobic solvents at low water concentrations, is comparable or even higher than in aqueous solution.^{7–11} On the basis of their studies, they concluded that the more hydrophobic the main solvent, the less water is needed to maintain protein activity.¹⁰ Moreover, the same authors claimed that water acts as a lubricant, forming hydrogen bonds with accessible groups of enzymes and providing them with the flexibility necessary for catalysis.¹⁰ Such explanations are generally accepted, although no direct evidence has been provided. The activity of enzymes under those conditions, or when they are trapped in

reverse micelles,^{12,13} became more than a curiosity when several biotechnological applications were proposed, taking advantage of this alternative.^{14–18} To date, no experimental technique has been able to provide a detailed structural–dynamics description of this phenomenon. To survey this issue, a number of MD simulation studies of proteins in water/organic solvent mixtures have been performed.^{19–29} However, the short time scale employed (7 ns for the longest trajectories) did not allow the observation of clear protein structural–dynamic differences between the several conditions considered. In 2003, Soares et al.¹⁹ reported the first systematic study of two model proteins (ubiquitin and cutinase) in hexane, under varying hydration conditions. Although they found correlations between root-mean-square deviations (RMSD), root-mean-square fluctuations (RMSF), and the percentage of water in the system, their analysis was performed only over the last 2 ns of 4 ns long trajectories. No significant protein backbone motions are expected at that time scale, so the observed fluctuations should be mainly due to the dynamics of the amino acid side chains. A high degree of dispersion, which the authors justified by the probable unstability of the initial structures, was observed in these simulations. Yang et al.²⁷ performed MD simulations of surfactant-solubilized subtilisin BPN' in octane, tetrahydrofuran, and acetonitrile at two hydration conditions and at a time scale of ~4 ns. The enzyme did not show significant structural or dynamic solvent-dependent differences. Then, the authors focused their analysis on the partition of water molecules between the surface of the protein and the organic phase and on the hydration of the active site. Very recently, Micaelo and Soares²⁸ published a comprehensive study of the hydration mechanism of cutinase in hexane, diisopropyl ether, 3-pentanone, ethanol, and acetonitrile with increasing hydration levels

* Corresponding author. E-mail: fangel@usc.es. Current address: Department of Applied Physics, Faculty of Physics, University of Santiago de Compostela, E-15782 Santiago de Compostela, Spain.

TABLE 1: List of MD Simulations Including Parameters and Results of Analysis

trajectory	no. of	No. of	% water	simulation	initial	total				
	water	decane				intramolecular	radii of	inter-subunit	hydrophobic	hydrophilic
	molecules	molecules		time (ns)	structure ID	H-bonds	gyration (nm)	distances (nm)	SAS (nm ²)	SAS (nm ²)
RW _a	24825	0	100	20	1TCD					
RW _b	24825	0	100	40	RW _a (20)	379 ± 9	2.46 ± 0.01	3.58 ± 0.03	117 ± 2	93 ± 2
W8D	3579	3402	11.80	40	RW _a (20)	380 ± 11	2.50 ± 0.01	3.66 ± 0.03	128 ± 2	90 ± 2
W6D	2336	3208	8.45	40	RW _a (20)	388 ± 10	2.49 ± 0.01	3.63 ± 0.02	129 ± 2	90 ± 1
W4D	1185	2825	5.04	40	RW _a (20)	392 ± 9	2.392 ± 0.008	3.35 ± 0.02	127 ± 2	87 ± 1
RD	0	2587	0	40	1TCD	563 ± 11	2.345 ± 0.006	2.28 ± 0.01	122 ± 2	58.1 ± 0.9

at a time scale of 7 ns. Slight structural changes were observed in the enzyme, but no analysis of its dynamics was performed. This latter work was mainly focused on the water behavior. These representative studies revealed the lack of MD simulations at longer time scales that could provide new insight into the structure and dynamics of enzymes at low water concentrations as well as into the role of the solvent for protein function.

In the present work, 40 ns long MD simulations of triose-phosphate isomerase from the parasite *Trypanosoma cruzi* (TcTIM) in several water/decane mixtures were performed. As a reference, two simulations of the same enzyme in pure decane and pure water, respectively, at the same time scale were carried out. TcTIM is a homodimeric enzyme with 251 residues per subunit, which folds in the so-called α/β -barrel motif with eight central β -strands surrounded by eight α -helices.³⁰ This enzyme was chosen because it was experimentally proven³¹ that TIMs from other species show a high catalytic activity at low water concentrations. Besides, it is a well-liked target for drugs against American trypanosomiasis, better known as Chagas' disease.^{30–37} A detailed analysis of the protein structure and dynamics, as well as of the solvent molecules' behavior, was performed under different conditions. The main objective of this work was to look for concentration-dependent differences in the protein and solvent behavior that could provide new insight into the understanding of the catalytic activity of enzymes in organic media. Additionally, the dynamic study of TcTIM in pure water by itself, one of the reference simulations in this work, could be helpful to better understand this enzyme and then to design or improve inhibitors. No MD studies of enzymes in organic solvents at low water concentrations, and either of TIM in aqueous media, at the time scale of tens of nanoseconds have been performed previously.

2. Materials and Methods

2.1. Setup of the Simulation Boxes. Six simulation systems were built up with one TcTIM molecule and different amounts of water as well as decane, forming a nonpolar phase. Decane was chosen due to its high hydrophobicity, which ensures that practically all water molecules remain close to the protein in the simulations that involve a mixture of both solvents. The systems, as well as the corresponding trajectories, were named RW_a, RW_b, W8D, W6D, W4D, and RD. RW_a consisted of 24 825 pre-equilibrated water molecules solvating the enzyme structure as taken from the Protein Data Bank (PDB ID 1TCD),³⁰ with no decane present. The protonation states of ionizable residues corresponded to biological pH (~7.4). The size of the simulation unit cell was chosen so that the distance between every protein atom and box walls was larger than 2 nm. This system was first simulated for 20 ns, the resulting final configuration being denoted by RW_a(20). RW_a was then extended for 40 ns more, producing the trajectory named RW_b. To build up the system W8D, waters at a distance larger than 8 Å from any of the protein atoms were removed from RW_a(20), and the resulting arrangement was enclosed into a

dodecahedron simulation box with the walls at a minimum of 2 nm from any of the water molecules. The free space of such a box was then filled with pre-equilibrated decane molecules. The same procedure as for W8D was repeated using distances of 6 and 4 Å to give the systems W6D and W4D, respectively. This procedure assures that both the protein and the water molecules in its immediate vicinity are in equilibrium to each other in the initial conformations of these three systems. As in previous studies,^{19,27} pH memory was assumed for the protein when transferred from pure water to the water/decane solutions. The system RD is analogous to RW_a but employing decane instead of water as the unique solvent. As in the case of RW_b, W8D, W6D, W4D, and RD were simulated for 40 ns. Table 1 specifies the number and relative concentration of water and decane molecules, as well as the time scale and the initial protein structure employed in each simulation.

2.2. MD Simulation Parameters. All MD simulations were performed using the GROMACS package^{38–41} version 3.2.1 with the GROMOS96 (43a2) force field^{42,43} for the protein and decane and the simple point charge (SPC) model for water.⁴⁴ Periodic boundary conditions with rhombic dodecahedron boxes as the basic unit cell in the NTP ensemble were used. The pressure was maintained by weak coupling⁴⁵ to a reference value of 1 bar, with a coupling time of 0.5 ps. The isothermal compressibility utilized in each simulation was the average of the water and decane values, weighted by the concentration of each solvent component. Water, decane, and protein were coupled separately to a temperature bath⁴⁵ at 298 K, with a coupling constant of 0.1 ps. Nonbonded interactions were evaluated using a twin range cutoff of 0.9 and 1.4 nm; interactions within the shorter and longer cutoffs were updated every step and every five steps, respectively. Beyond the 1.4 nm cutoff, a reaction field correction with a dielectric constant ϵ of 78 was employed for all the simulations, including those where the main solvent was decane. Since the dielectric constant applies only at large distances, no artificial effects were expected in the protein structure/dynamics nor in the water behavior. Before every MD simulation, an energy minimization using the steepest descent method was performed to avoid the most unfavorable interactions. Random initial velocities were assigned to the systems from a Maxwell–Boltzmann distribution at 298 K. The equations of motion were integrated using the leapfrog method with a time step of 2 fs. The bond lengths and angles in water were constrained using the SETTLE algorithm,⁴⁶ while the LINCS algorithm⁴⁷ was used to constrain bond lengths within the protein and decane molecules. During the MD simulations, coordinates, velocities, and energies were stored every 10 ps for further analysis.

2.3. Analysis of MD Trajectories. The viewers RASMOL⁴⁸ 2.7, VMD⁴⁹ 1.8.2, and PyMOL⁵⁰ 0.99 were employed to roughly inspect the behavior of both the enzyme and the solvent. This first screening was used to design a more detailed analysis. RMSD, taking always as a reference the X-ray structure of the protein, were calculated for all the trajectories. To study the

mobility of every residue, RMSF values were calculated over the last 20 ns of the trajectories RW_b, W8D, W6D, W4D, and RD, taking also the X-ray structure as a reference. The principal components (PC) of the protein motion were determined from the diagonalization of the variance–covariance matrix, providing the amplitude and the direction of each concerted motion (i.e., the eigenvalues and the corresponding eigenvectors). This analysis was carried out for the last 20 ns of all the trajectories except for RW_a, employing only atoms of the protein backbone. The trajectories were projected on the eigenvectors with the largest eigenvalues, and images of the two extreme enzyme–backbone conformations for each of these motions were generated. Additionally, the same analysis was also performed over the 40 ns of the whole RW_b trajectory. Cluster analysis was carried out over the last 20 ns of all the trajectories using the GROMOS method.⁵¹ The RMSD cutoff employed to define each cluster was 0.15 nm for trajectories RW_b, W8D, W6D, and W4D and 0.10 nm for RD. A detailed screening of the protein secondary structure was also performed for all the simulations, by using the definitions of the Dictionary of Protein Secondary Structure (DPSS).⁵² RMSD, RMSF, and secondary structure as a function of time were analyzed by means of programs contained in the GROMACS package^{38–41} version 3.2.1, while the analysis of essential protein motions as well as the cluster determinations were performed by using version 3.3.1 of the same package. A number of simple programs were developed to study the distribution and dynamics behavior of the solvent molecules around the enzyme. For several frames of every trajectory, the number of oxygens belonging to water molecules as well as decane carbons at less than 4 Å from every TcTIM atom were calculated. Additionally, for the last 20 ns of the trajectories that correspond to systems involving water (RW_b, W8D, W6D, and W4D), the distances between every oxygen atom of water molecule and every protein atom were monitored. Such a computationally expensive analysis allows the calculation of the water molecules' residence time in the vicinity of every TcTIM residue. The number of intramolecular hydrogen bonds of the protein were also calculated for all the trajectories using a donor–acceptor atom cutoff distance of 3.5 Å and a donor–hydrogen–acceptor angle of less than 30°. The radii of gyration (RoG), as well as hydrophilic and hydrophobic contributions to the solvent accessible surface (SAS) of the enzyme, were determined throughout all the trajectories. The obtained values were averaged over the last 20 ns. Tools from the GROMACS package were employed to determine the H-bonds, RoG, and SAS.

3. Results and Discussion

3.1. Visual Analysis of Trajectories. In the time scale and under the simulation conditions employed in this work, the two TcTIM subunits remained together, and their tertiary structures were apparently conserved. However, the enzyme conformation arising from the trajectory RD after 40 ns of MD simulations seems to be smaller than the rest of the conformations, and it missed several α -helices (Figure 1). The size of the final structure obtained from W4D also appears to be slightly smaller than those obtained from RW_b, W8D, and W6D. Notably, most of the water molecules in W8D, W6D, and W4D formed a cluster in the cavity next to the interface of the two protein subunits. The fact that the cluster formation was observed in the three different trajectories suggests that it is not accidental. Several reasons could explain this grouping; the most evident arises from the combination of the high hydrophobicity of decane, the main solvent in those simulations, together with

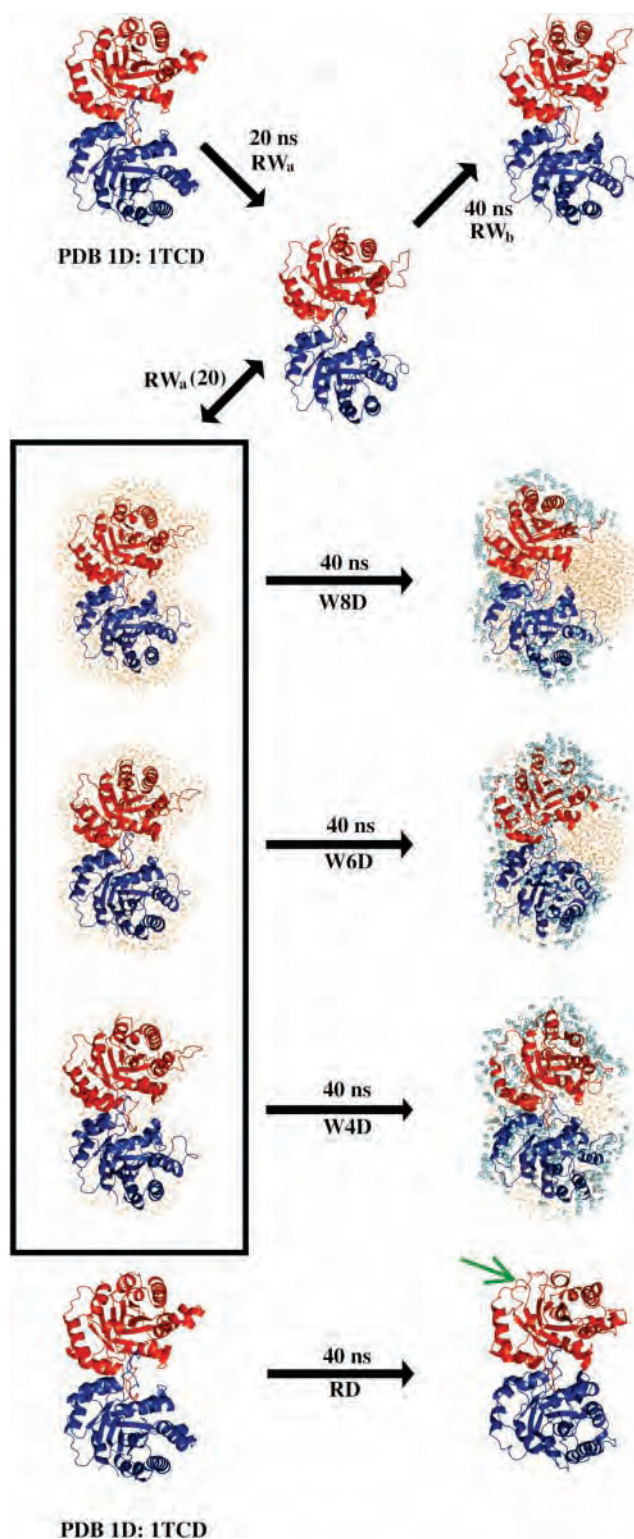


Figure 1. Initial and final conformations of the TcTIM for all the trajectories. The time scales and the name of the systems are indicated. For W8D, W6D, and W4D, all waters (orange sticks) and decane-carbon atoms (green balls) at a distance of less than 4 Å are displayed. Subunits A and B (as referred to in the text) are drawn in blue and red, respectively. The green arrow in the last conformation of RD points to the region of the enzyme that lost its secondary structure.

the geometry of the protein that provides a nucleation point around which a water cluster can grow. The enzyme internal motions could also contribute to the water aggregation. The region of the protein that is in contact with the water cluster can be considered to be relatively hydrophobic but in no way

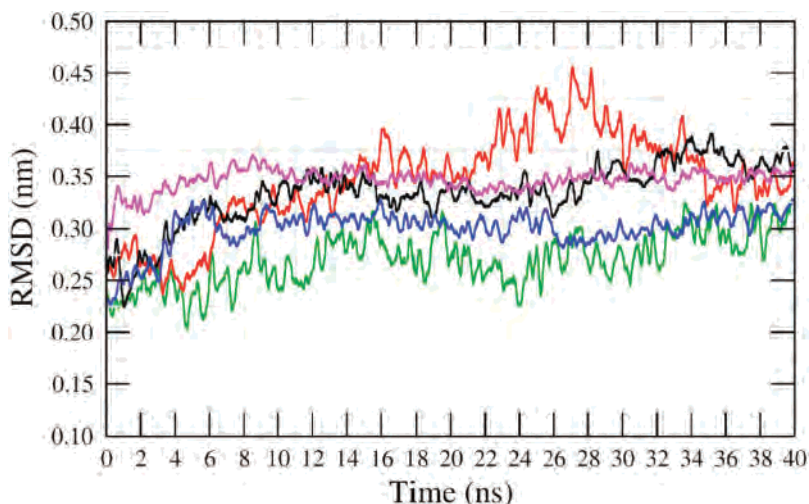


Figure 2. RMSD of the TcTIM backbone as a function of time for the trajectories: RW_b (green), W8D (red), W6D (black), W4D (blue), and RD (magenta). The crystal structure of the enzyme was employed as a reference in all cases.

more than decane. So, it seems that the major reason for the water cluster formation, precisely in the protein cavity, is the minimization of the exposure of water molecules to decane. In addition to the cluster, a number of water molecules seem to exist, individually or in small groups, fixed to specific sites of the protein.

3.2. Protein Structure and Dynamics. As described in the Materials and Methods, the simulations RW_b, W8D, W6D, and W4D started from the same initial conformation: RW_a(20) (Figure 1 and Table 1). However, RW_b was a direct extension of RW_a, while in the case of W8D, W6D, and W4D, the solvent was partially substituted by decane, giving a starting configuration that is not fully equilibrated. After a few nanoseconds, the RMSD of all the trajectories exhibited fluctuations with no clear trends (Figure 2). For instance, the RMSD of W8D increased between ~ 22 and ~ 30 ns and then decreased to recover, at the end of the trajectory, a value comparable to that at 10 ns. Notably, the RMSD corresponding to the simulation with decane as the unique solvent (RD) displayed a behavior very similar to that of the other four systems, although it appeared to present less fluctuations. RMSD values are important but not specific enough because the structure deviation of the protein at every time step is represented by just one number. RMSF values for every TcTIM residue were calculated from the last 20 ns of the trajectories RW_b, W8D, W6D, W4D, and RD (Figure 3a). Accumulated RMSF values (RMSF_{acc}) (i.e., the sum of amino acid RMSF values for each subunit and trajectory) were also calculated (Figure 3b). It can be observed that, in general, loops 5 and 6 (between residues 129–139 and 169–179, respectively) presented the highest fluctuations. RMSF values of loop 5 are important for all subunits and trajectories, while loop 6 of subunits B in W8D and A in W6D exhibited a much higher mobility than the same loop in the other subunit (Figure 3a). The activity of this enzyme has been correlated with a movement of such a loop in just one subunit, with a period of 200–400 μ s.^{53,54} The movements identified in our trajectories are probably just a flavor of the complete route covered by loop 6 because only local movements of relatively high frequency could be observed in our simulations. However, the higher fluctuations in W8D and W6D when compared to those in pure water, with similar RMSD values, suggest that the presence of decane might speed up the enzyme dynamics. In general, RMSF residue profiles were similar from one simulation or subunit to another. Loops 5 and 6 were the amino

acid segments with maximum fluctuations in the two subunits for all the simulations, while the β -strand segments were the regions of lowest fluctuations. Although the differences in the RMSF_{acc} values were not extremely large, it is noteworthy that the maximum and minimum values corresponded to the systems W8D and RD, respectively. Except for the trajectory W6D, the second subunit always presented RMSF_{acc} values slightly higher than the first one. It is evident from data shown in Figure 3a that the higher mobility of the subunit A in W6D was due to the motion of its loop 6 in that simulation.

The components with the highest amplitude of the TcTIM motion during the last 20 ns of all the trajectories were identified as explained in the Materials and Methods. These amplitudes were sorted in decreasing order, and the corresponding values were found to decay in an exponential way in all the systems (Figure 4a). Comparison between different systems revealed that, in the presence of decane, the higher the water concentration, the higher the mobility of the enzyme. Remarkably, TcTIM in pure water showed amplitudes comparable to those in W4D, the simulation with the lowest water concentration. The enzyme was clearly more rigid in the pure hydrophobic solvent. The accumulated RMSF values, when compared between the different systems, show a trend similar to that of the amplitudes of the principal motions of the enzyme. However, the PC analysis is finer than simple RMSF calculations, and the analysis was performed considering only the backbone atoms of the enzyme, whose movements are expected to be more directly related to protein function. It has been proposed in the past that the amino acid segments with high RMSF values coincide with the protein regions that first lose their structure in the unfolding processes.⁵⁵ Thus, the rigidity of the protein observed in pure decane could be related to the higher thermal stability that enzymes usually present in organic solvents.¹ In agreement with RMSF values, PC analysis showed that the most important motions of the enzyme in all the simulations involve loops 5 and 6. Probably due to the length of the trajectory, no great structural differences could be appreciated between the extremes of the projections of the trajectories on the most important eigenvectors. Attempting to obtain further insight into the protein motion, and taking advantage of the fact that the trajectory RW_b was a continuation of RW_a, PC analysis was performed for this whole trajectory (40 ns long). Again, the PC of the enzyme dynamics involved loops 5 and 6. Movie S1 shows the motion with the highest eigenvalue obtained from this simulation.

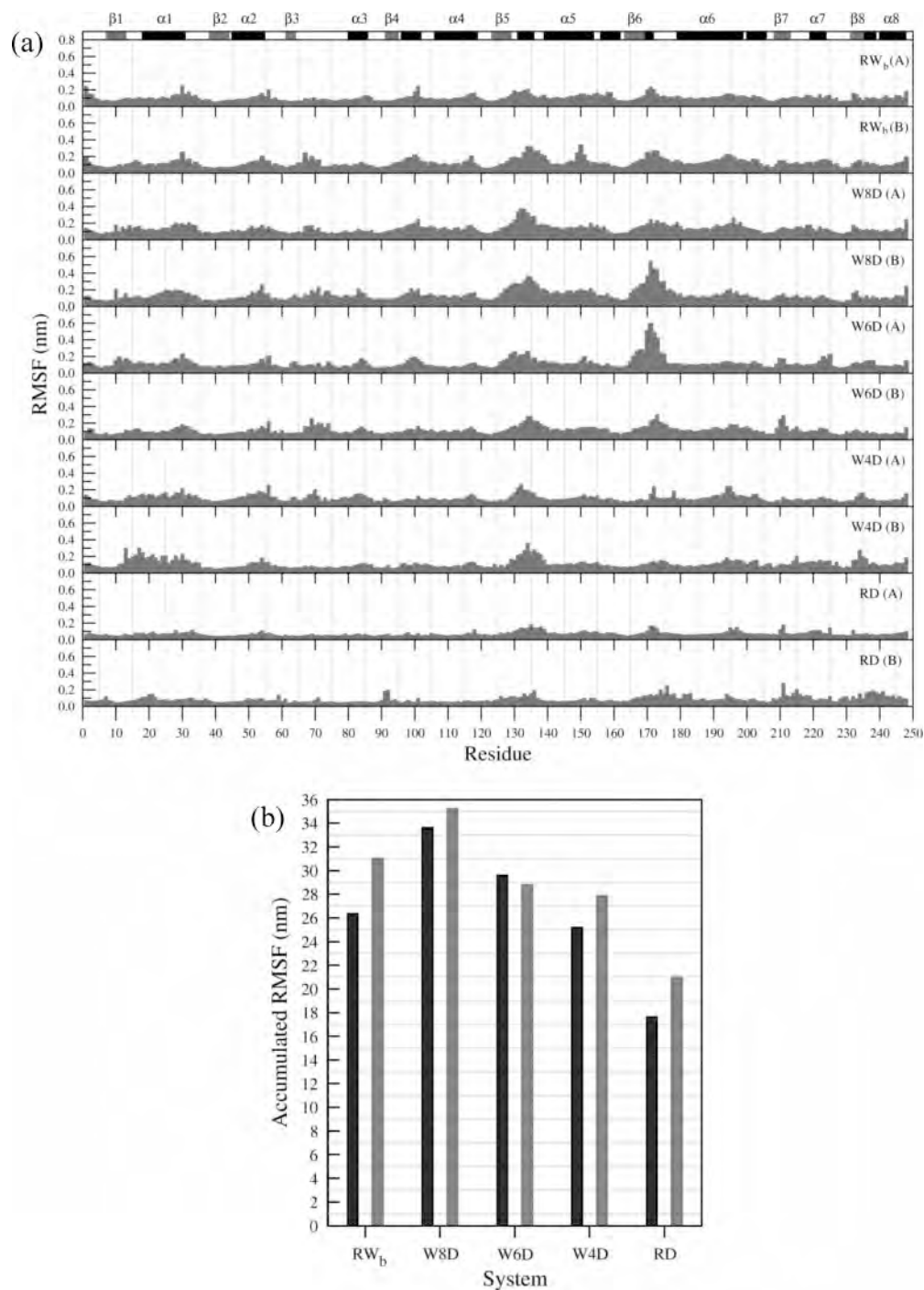


Figure 3. (a) RMSF of every TcTIM residue, calculated over the last 20 ns of each trajectory. The name of the system and subunit (A or B) are indicated at the right upper corner of every line. Segments of residues that form α -helices or β -strands are labeled at the top of the plot. (b) Accumulated RMSF values, calculated by summing the values of each line in panel a for subunits A (black) and B (gray).

Besides that of the loops, a relative movement of the enzyme subunits can be observed. Such a movement could be described as a shrinking–stretching of the enzyme toward the cavity that hosts the water cluster in W8D, W6D, and W4D, and it could contribute to the aggregation of water molecules in that region in the simulations where the two solvents are present. The extreme conformations corresponding to the motion observed in the movie are shown in Figure 4b, where the open and closed conformations of loop 6 can be clearly appreciated (Movie S1 shows precisely the transition pathway). Similar conformations of loop 6 have been previously reported for other TIMs, although the period of the transition has been estimated to be of the order of hundreds of microseconds.⁵⁶ The observation of this move-

ment in our simulations could be favored by the initial enzyme structure or/and by the simulation conditions. To study the dynamics of this movement, the distance between the center of mass of the three amino acids of loop 6 that present the highest displacement, and one amino acid of loop 7 that is relatively static and located in the direction of the movement, were calculated (Figure 4c). It can be seen that, in general, the distance in subunit A is higher than the distance in subunit B. The two distances are almost constant for the trajectories W4D and RD, while they are much more variable for the other three trajectories. In particular, for RW_b, loop 6 of the second subunit seems to start from an open conformation, and it changes to a closed conformation at ~ 15 ns. The net displacement of the chosen

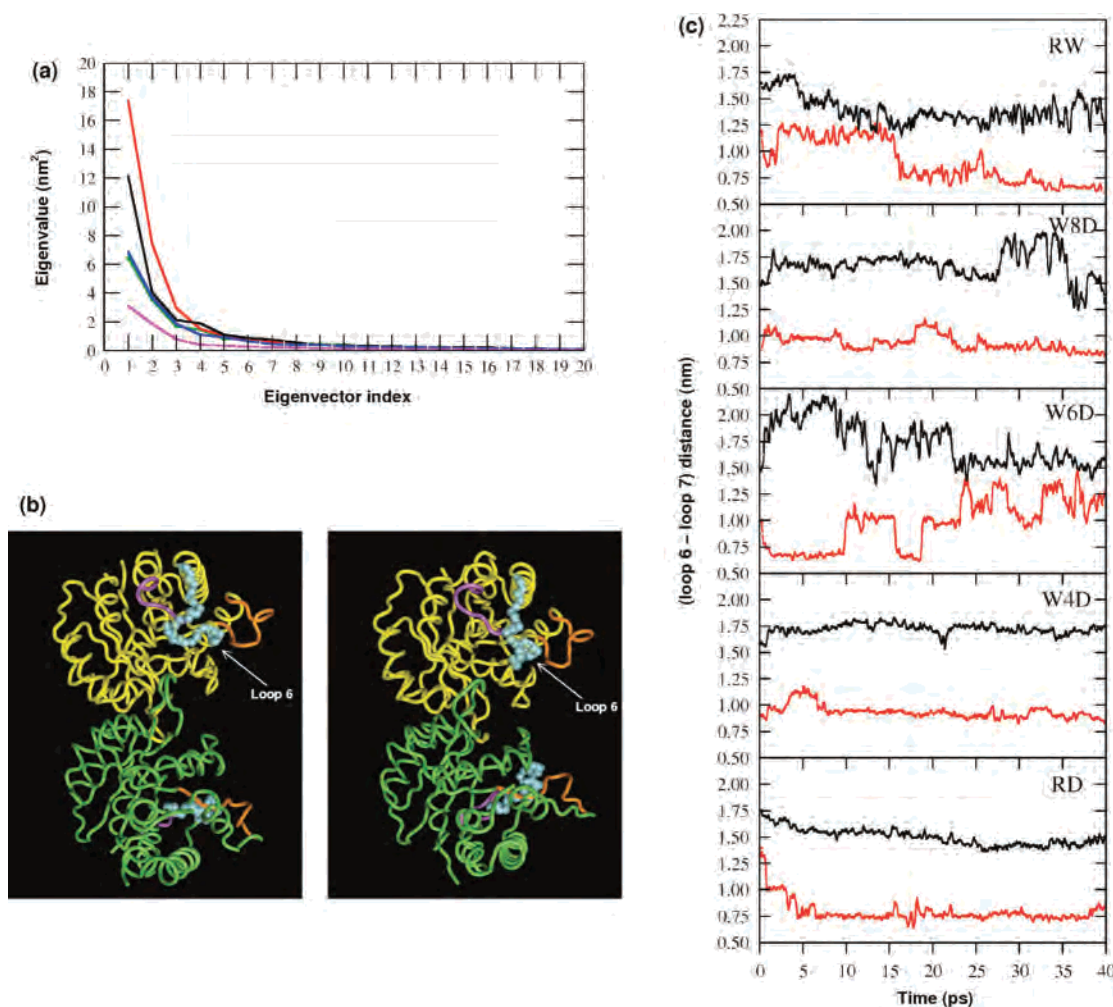


Figure 4. (a) Highest eigenvalues obtained from principal component analysis of the simulations RW_b, W8D, W6D, W4D, and RD in green, red, black, blue, and magenta, respectively. (b) The two extreme conformations of the principal TcTIM motion obtained from the trajectory RW_b. Only the backbones of subunits A (yellow) and B (green) are displayed. The residues of loops 5–7 are in orange, cyan spheres, and magenta, respectively. (c) Distance between the center of mass of the amino acids Gly170–Thr171–Gly172 (loop 6) and Val212 (loop 7) of the subunits A (black) and B (red). The name of the system is indicated at the right upper corner of every plot.

group of amino acids is ~ 0.6 nm in the direction of the measurement. At least two transitions (closed–open) were observed for the same subunit in the trajectory W6D.

As described in the Materials and Methods, an analysis of structural clusters over the last 20 ns of all the trajectories was performed. The representative structures of every cluster for each trajectory were superimposed on each other to show the differences between the most visited conformations along the simulation pathways. The catalytic residues (Lys 14, His 96, and Glu 168) showed a higher mobility in the trajectories RW_b, W8D, and W6D than in RD and W4D. Particularly, Lys 14 and His 96 were found to be particularly flexible in W8D and W6D, where they exposed a slight rotation around the aliphatic chain in Lys and of the carboxylic head in Glu. In spite of these features, the average structure of the site appeared to be conserved in all cases.

A detailed analysis of the protein secondary structure was performed for all the trajectories as a function of time. The number of residues that form part of the β -strands was nearly constant (~ 38 for each subunit) for all the simulations. Even when the global structure of the protein after 40 ns of MD simulations was preserved, a lack of the number of residues that constitute α -helices took place in the trajectory RD (Figure 1). In this simulation, such a number fluctuated between 65 and 80 for each enzyme subunit, while for the systems containing

water, it took values between 90 and 100. In particular, the shortest helices of the α/β -barrel structure ($\alpha 3$ and $\alpha 7$) in one of the TcTIM subunits, as well as $\alpha 8$ in the other subunit, unfolded in the trajectory RD (unfold in this context does not necessarily mean a transition to a random coil but a lack of stable H-bonds leading to a stretching or distortion of the helix. As seen in Figure 1 for the final conformation of the trajectory RD, certain order is still present in the unfolded segments, although the structure cannot be defined as a helix on the basis of the DPSS.⁵²). $\alpha 3$ was observed to be unstable in the same subunit along the remaining four trajectories, including RW. $\alpha 5$ also unfolded during the trajectory W4D, again for the same subunit as for the system RD. The lack of well-defined α -helices for these segments did not appear to be critical for the stability of the protein because of their small size (5–8 residues). However, during the simulations of RD, helix $\alpha 1$ also unfolded in one of the protein subunits. Such a helix consists of 18 residues at the beginning of the simulations, which represent a significant percentage of the overall protein secondary structure (Figure 1).

3.3. Solvent Molecule Distribution and Dynamics. It is evident from Figure 1 that the distribution of solvent molecules around the enzyme structure was not homogeneous in the final conformation of the trajectories W8D, W6D, and W4D. However, the number of water molecules in the vicinity of every

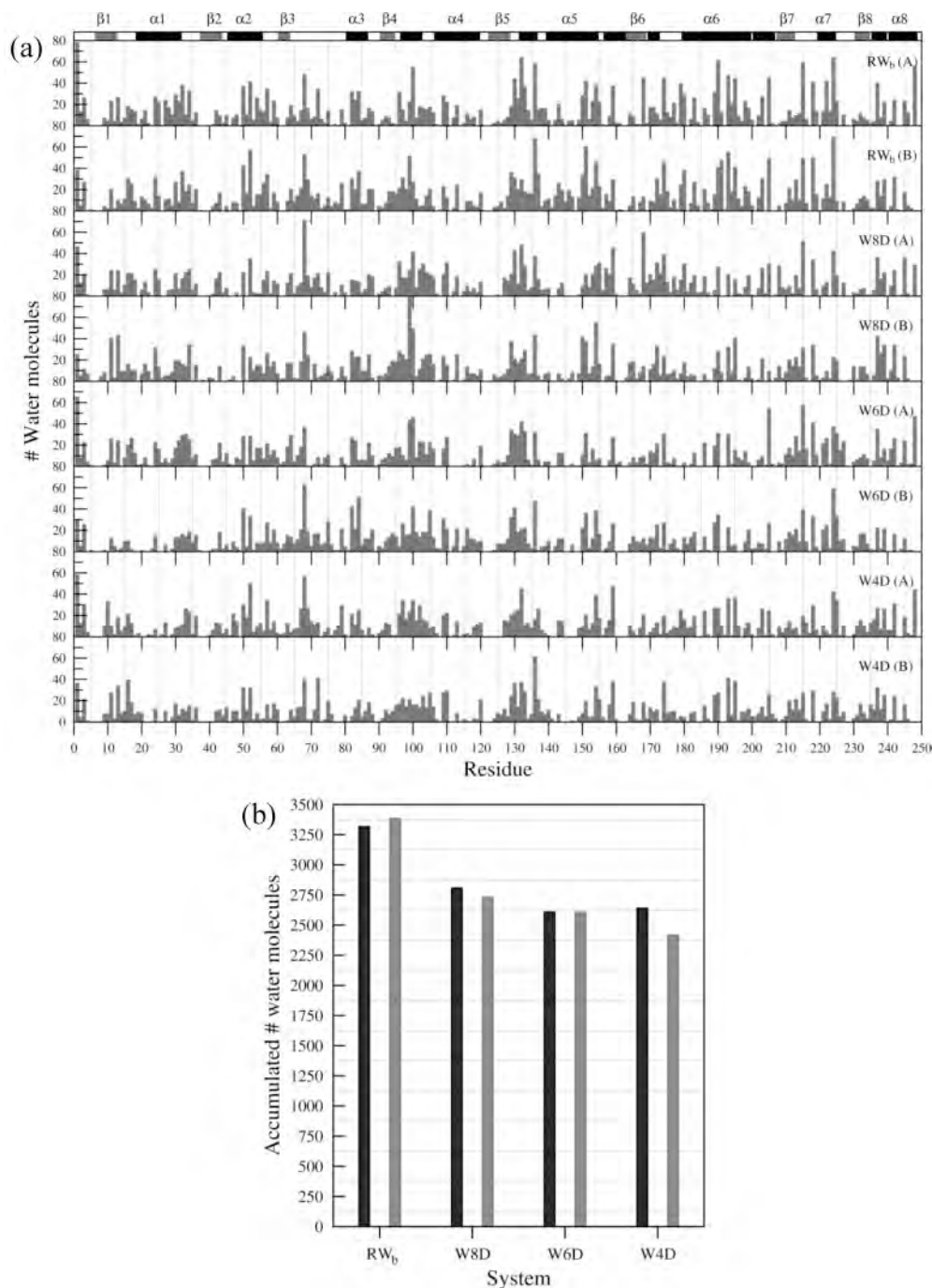


Figure 5. (a) Number of water molecules that are at a distance of less than 4 Å in the last conformation of the trajectories RW_b, W8D, W6D, and W4D. The name of the system and subunit (A or B) are indicated at the right upper corner of every line. Segments of residues that form α -helices or β -strands are labeled at the top of the plot. (b) Accumulated number of water molecules calculated by summing values of each line in panel a for subunits A (black) and B (gray).

protein residue was almost the same as that in RW_b, even though the difference in the total number of waters is important, ranging from 24 825 to 1185 (Figure 5a). The profiles for both subunits in the same system were also very similar. The accumulated number of water molecules was calculated (Figure 5b). A similar analysis was performed for the number of carbon atoms of decane molecules at less than 4 Å of every protein residue (Figure 6a). The decane atom distribution was less similar among systems and protein subunits than the water distribution. Nevertheless, several protein regions with a consistent affinity for the hydrophobic solvent could be identified such as segments of helix 6, loop 6, and loop $\alpha\beta 2$ (Figure 6a). This result could be used to propose potential binding sites for hydrophobic

ligands, as previously suggested by Mattos and Ringe.^{57,58} The accumulated values of the number of decane atoms were also calculated (Figure 6b). By comparing Figures 5 and 6, it is clear that the water molecule distribution around the enzyme was more specific than the distribution of the nonpolar solvent. It is worth mentioning that, although the results presented in Figures 5 and 6 were obtained from a single conformation of each system, the same analysis was performed for a number of frames through the last 20 ns of all the trajectories with no significant variations.

The residence time values of water molecules were filtered to determine the number of waters that remained at a distance of less than 4 Å of every residue for more than 2 ns. This

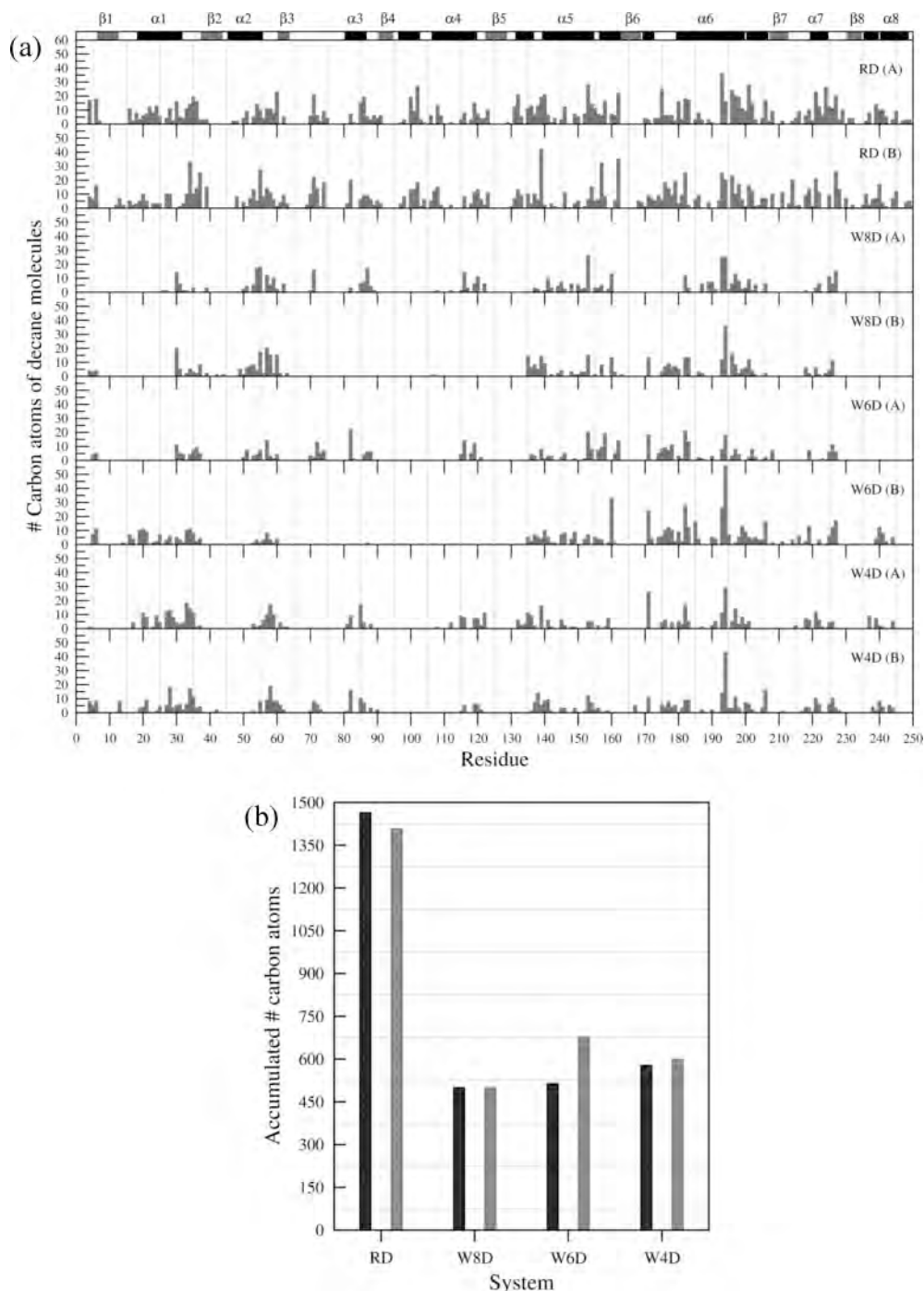


Figure 6. (a) Number of carbon atoms of decane molecules that are at a distance of less than 4 Å in the last conformation of the trajectories W8D, W6D, W4D, and RD. The name of the system and subunit (A or B) are indicated at the right upper corner of every line. Segments of residues that form α -helices or β -strands are labeled at the top of the plot. (b) Accumulated number of carbon atoms of decane molecules calculated by summing values of each line in panel a for subunits A (black) and B (gray).

calculation was performed for the systems RW_b, W8D, W6D, and W4D over the last 20 ns of every trajectory (Figure 7a). It could be appreciated that there were less water molecules with residence times higher than 2 ns with pure water than for the simulations with water/decane mixtures. This is reasonable due to the low partition coefficient of water in decane and also because the systems W8D, W6D, and W4D do not have many water molecules available to interact with the residues that indeed need to be hydrated. The corresponding accumulated numbers of water molecules were also calculated (see Figure 7b). The increment of the number of water molecules with long residence times when the water concentration decreases is clear.

An analogous analysis was performed with a longer cutoff for the residence time, the results being similar although with a lower number of water molecules in the vicinity of each residue (data not shown).

Aproximately 15 water molecules with different residence times were found to be located in the dimer interface of the enzyme in the trajectories RW_b, W8D, W6D, and W4D, and a variable number of water molecules seeped into the active site. The presence of water molecules in contact with the catalytic residues was associated with an open conformation of loop 6. This supports the theory that such a loop modulates the entrance of the substrate to the catalytic residues. In the simulation of

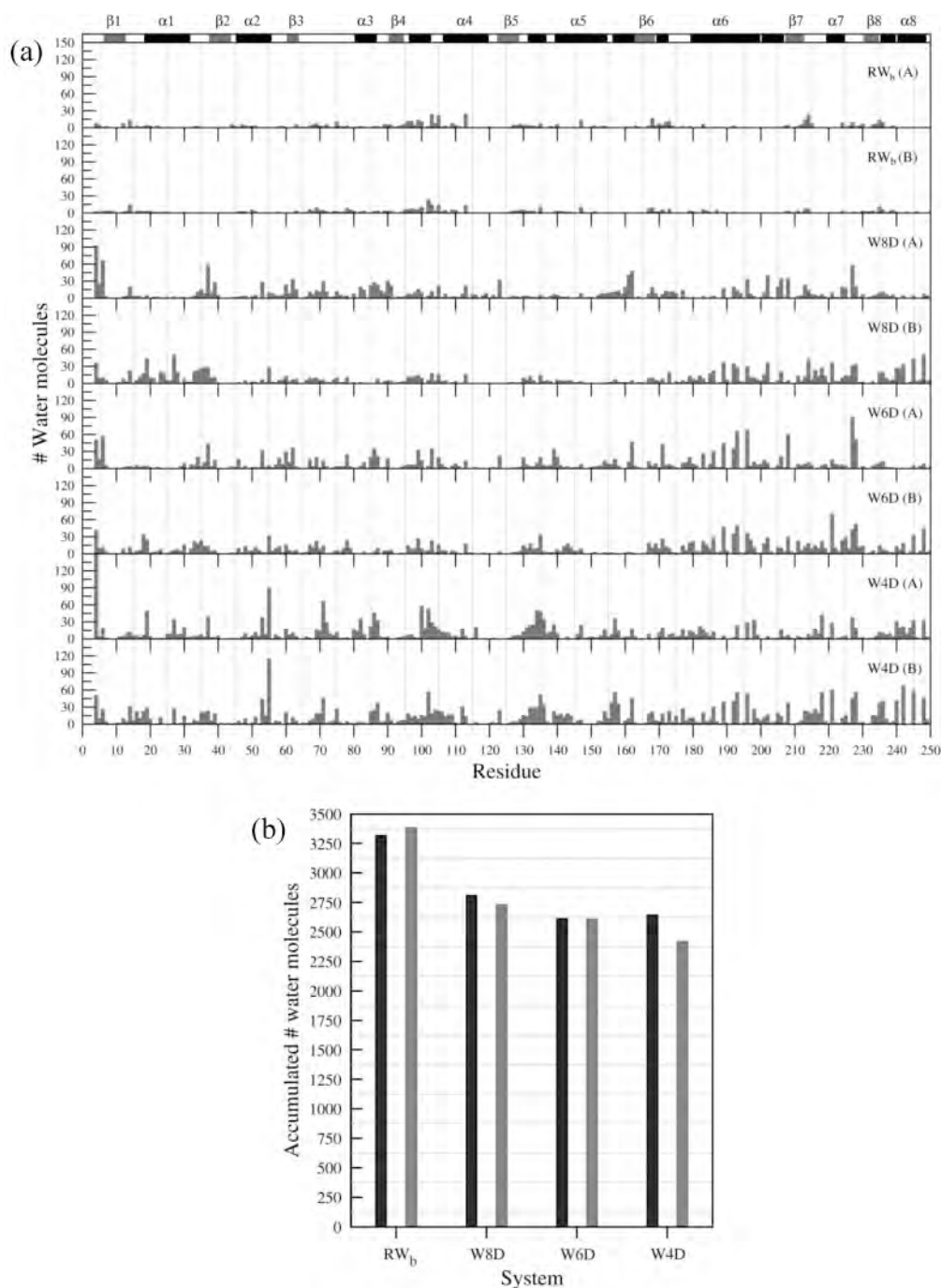


Figure 7. (a) Number of water molecules that are at a distance of less than 4 Å from every TcTIM residue, which remain for at least 2 ns during the last 20 ns of the trajectories RW_b, W8D, W6D, and W4D. The name of the system and subunit (A or B) are indicated at the right upper corner of every line. Segments of residues that form α -helices or β -strands are labeled at the top of the plot. (b) Accumulated number of water molecules calculated by summing values of each line in panel a for subunits A (black) and B (gray).

RD, no decane molecules were found to be close to the interface nor to the active site.

3.4. Intramolecular Hydrogen Bonds. It has been proposed that many enzymes maintain their mobility and activity at low water concentrations because, even under those conditions, they can form hydrogen bonds with the water molecules of their environment.¹⁰ In the absence of water, the polar groups of the protein switch to interact with each other, forming intramolecular H-bonds. The numbers of these H-bonds for trajectories RW_b, W8D, W6D, W4D, and RD are shown in Figure 8a. The plot is very clear, showing that all the simulations involving water exhibited almost the same number of intramolecular H-bonds (see average values in Table 1). However, for the simulation

with decane as a unique solvent, there were no water molecules available, and many more intramolecular H-bonds were formed. The same calculation was performed considering only the amino acids that form part of β -strands or α -helices, obtaining similar results (Figure 8b). Also in the case of RD, the increment in the number of H-bonds resulted in a contraction of the TcTIM structure, which was quantified by measuring the RoG. The average values of these radii over the last 20 ns of every trajectory are reported in Table 1. The hydrophobic contributions to the SAS did not vary significantly from one system to another, but the hydrophilic SAS was much lower in RD than in any of the other systems. Thus, the contraction of the protein structure in RD involved the formation of H-bonds between the polar

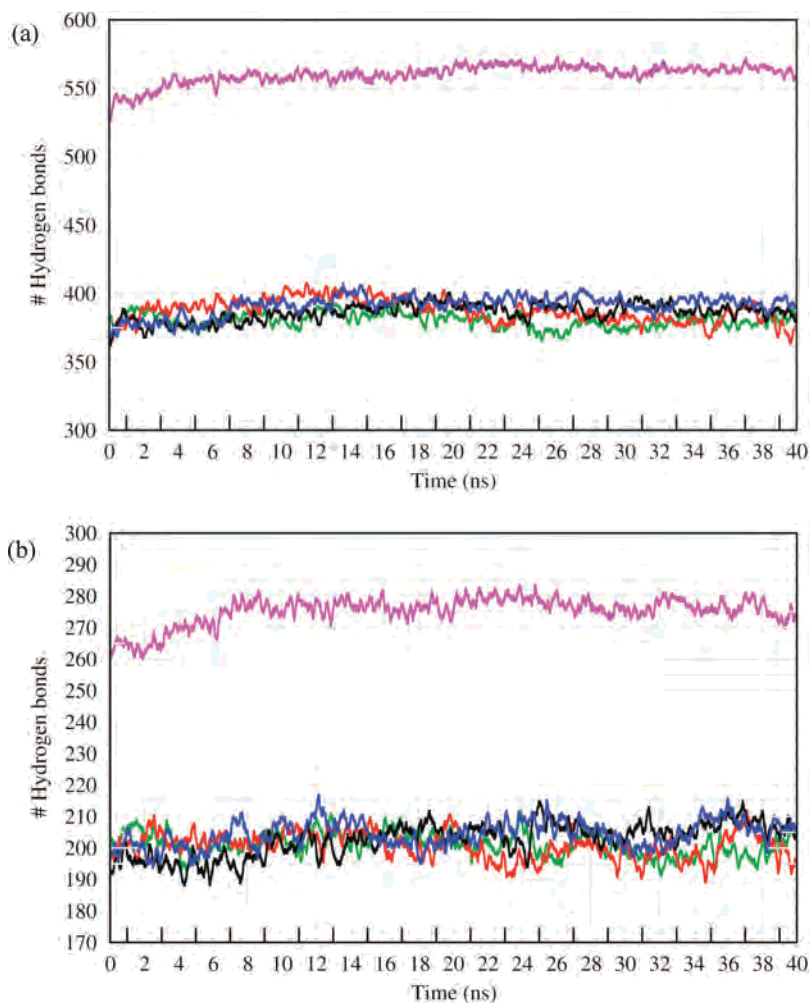


Figure 8. Number of intramolecular hydrogen bonds of (a) all TcTIM residues and (b) the TcTIM residues that form part of α -helices and β -strands, as a function of time, for the trajectories RW_b (green), W8D (red), W6D (black), W4D (blue), and RD (magenta).

hydrophilic groups. Simultaneously, those groups internalized in the protein to escape from the hydrophobic solvent. Unexpectedly, the enzyme also reduced its size throughout the simulation of W4D. Such a contraction cannot be explained in the same terms as for RD because the reduction of hydrophilic SAS in this system was almost negligible (see Table 1), and the numbers of H-bonds, as well as the distribution of water molecules in the vicinity of the residues, were similar to the values found for RW_b, W8D, and W6D. The distance between the centers of mass of both subunits was calculated for all the systems, their values being consistently lower for W4D and RD than for RW_b, W8D, and W6D (Table 1). Thus, the reduction of RoG in W4D was due to the two subunits approaching each other, maintaining the hydrophilic and hydrophobic SAS values. This approach also contributed to the decrease of RoG in RD, and it could be related to the previously mentioned shrinking–stretching of the enzyme structure toward the cavity. Such a motion could not be reversible due to the pressure of the decane molecules and to the low number of waters present in this simulation.

4. Conclusion

The effects of the water on the decane concentration ratio in the structure and dynamics of TcTIM, as well as the behavior of the solvent molecules in the vicinity of the enzyme, were studied in detail through MD simulations at a time scale of 40 ns. In contrast to recent MD studies performed at much shorter

time scales,²⁷ the enzyme was found to be sensitive to the water concentration. The analysis of the protein fluctuations and of the PC of the backbone motion revealed that, in the presence of organic solvents, the higher the water concentration, the higher the mobility of the enzyme. Nevertheless, TcTIM in pure water showed amplitudes comparable to those in W4D, the simulation with the lowest water concentration, while it was clearly more rigid in pure decane. These results suggest that the enzyme dynamics speeds up at relatively low water concentrations, probably due to the presence of organic solvent molecules attached to specific sites of the protein surface, although there is a threshold in the number of water molecules needed for the protein to keep its structure and dynamics. Our analysis also provides new insight into the dynamics of TcTIM that could be helpful in better understanding this enzyme.

Acknowledgment. We thank Prof. Armando Gómez-Puyou, Universidad Nacional Autónoma de México (UNAM), Dr. Pilar Brocos (USC), Prof. Miguel Costas (UNAM), and Dr. Hao Fan (UCSF) for their critical reading of the manuscript and for very useful suggestions. This work was supported by Grants 41328-Q and J49811-Q from CONACyT and Grant IN105107 from PAPIIT-UNAM, México. We are grateful to the Dirección General de Servicios de Cómputo Académico (DGSCA) of the UNAM for computer time employed in this work and for their excellent services. N.D.-V. is grateful for financial support from CONACyT.

Supporting Information Available: Movie S1, showing the projection of the trajectory RW_b on the eigenvector with the highest eigenvalue, obtained by PC analysis. This information is available free of charge via the Internet at <http://pubs.acs.org>.

References and Notes

- (1) Tuena de Gomez-Puyou, M.; Gomez-Puyou, A. *Crit. Rev. Biochem. Mol. Biol.* **1998**, *33*, 53.
- (2) Finney, J. L.; Poole, P. L. *Commun. Mol. Cell. Biophys.* **1984**, *2*, 129.
- (3) Levy, Y.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 3325.
- (4) Pal, S. K.; Zewail, A. H. *Chem. Rev.* **2004**, *104*, 2099.
- (5) Rupley, J. A.; Careri, G. *Adv. Protein Chem.* **1991**, *41*, 37.
- (6) Rupley, J. A.; Gratton, E.; Careri, G. *Trends Biochem. Sci.* **1983**, *8*, 18.
- (7) Zaks, A.; Klibanov, A. M. *Science (Washington, DC, U.S.)* **1984**, *224*, 1249.
- (8) Zaks, A.; Klibanov, A. M. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 3192.
- (9) Zaks, A.; Klibanov, A. M. *J. Am. Chem. Soc.* **1986**, *108*, 2767.
- (10) Zaks, A.; Klibanov, A. M. *J. Biol. Chem.* **1988**, *263*, 8017.
- (11) Zaks, A.; Klibanov, A. M. *J. Biol. Chem.* **1988**, *263*, 3194.
- (12) Gonnelli, M.; Strambini, G. B. *J. Phys. Chem.* **1988**, *92*, 2854.
- (13) Strambini, G. B.; Gonnelli, M. *J. Phys. Chem.* **1988**, *92*, 2850.
- (14) Bell, G.; Halling, P. J.; Moore, B. D.; Partridge, J.; Rees, D. G. *Trends Biotechnol.* **1995**, *13*, 468.
- (15) Dordick, J. S. *Appl. Biochem. Biotechnol.* **1988**, *19*, 103.
- (16) Dordick, J. S. *Enzyme Microb. Technol.* **1989**, *11*, 194.
- (17) Klibanov, A. M. *CHEMTECH* **1986**, *16*, 354.
- (18) Klibanov, A. M. *Trends Biotechnol.* **1997**, *15*, 97.
- (19) Soares, C. M.; Teixeira, V. H.; Baptista, A. M. *Biophys. J.* **2003**, *84*, 1628.
- (20) Micaelo, N. M.; Teixeira, V. H.; Baptista, A. M.; Soares, C. M. *Biophys. J.* **2005**, *89*, 999.
- (21) Colombo, G.; Toba, S.; Merz, K. M. *J. Am. Chem. Soc.* **1999**, *121*, 3486.
- (22) Norin, M.; Haeffner, F.; Hult, K.; Edholm, O. *Biophys. J.* **1994**, *67*, 548.
- (23) Peters, G. H.; van Aalten, D. M. F.; Edholm, O.; Toxvaerd, S.; Bywater, R. *Biophys. J.* **1996**, *71*, 2245.
- (24) Toba, S.; Hartsough, D. S.; Merz, K. M. *J. Am. Chem. Soc.* **1996**, *118*, 6490.
- (25) Zheng, Y. J.; Ornstein, R. L. *J. Am. Chem. Soc.* **1996**, *118*, 4175.
- (26) Zheng, Y. J.; Ornstein, R. L. *Protein Eng.* **1996**, *9*, 485.
- (27) Yang, L.; Dordick, J. S.; Garde, S. *Biophys. J.* **2004**, *87*, 812.
- (28) Micaelo, N. M.; Soares, C. M. *FEBS Lett.* **2007**, *274*, 2424.
- (29) Zheng, Y. J.; Ornstein, R. L. *Biopolymers* **1996**, *38*, 791.
- (30) Maldonado, E.; Soriano-Garcia, M.; Moreno, A.; Cabrera, N.; Garza-Ramos, G.; de Gomez-Puyou, M. T.; Gomez-Puyou, A.; Perez-Montfort, R. *J. Mol. Biol.* **1998**, *283*, 193.
- (31) Garza-Ramos, G.; Tuena de Gomez-Puyou, M.; Gomez-Puyou, A.; Gracy, R. W. *Eur. J. Biochem.* **1992**, *208*, 389.
- (32) Gag, X. G.; Maldonado, E.; Perez-Montfort, R.; Garza-Ramos, G.; Tuena de Gomez-Puyou, M.; Gomez-Puyou, A.; Rodriguez-Romero, A. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 10062.
- (33) Garza-Ramos, G.; Cabrera, N.; Saavedra-Lira, E.; Tuena de Gomez-Puyou, M.; Ostoa-Saloma, P.; Perez-Montfort, R.; Gomez-Puyou, A. *Eur. J. Biochem.* **1998**, *253*, 684.
- (34) Perez-Montfort, R.; Garza-Ramos, G.; Alcantara, G. H.; Reyes-Vivas, H.; Gao, X. G.; Maldonado, E.; Tuena de Gomez-Puyou, M.; Gomez-Puyou, A. *Biochemistry* **1999**, *38*, 4114.
- (35) Reyes-Vivas, H.; Hernandez-Alcantara, G.; Lopez-Velazquez, G.; Cabrera, N.; Perez-Montfort, R.; Tuena de Gomez-Puyou, M.; Gomez-Puyou, A. *Biochemistry* **2001**, *40*, 3134.
- (36) Tellez-Valencia, A.; Avila-Rios, S.; Perez-Montfort, R.; Rodriguez-Romero, A.; Tuena de Gomez-Puyou, M.; Lopez-Calahorra, F.; Gomez-Puyou, A. *Biochem. Biophys. Res. Commun.* **2002**, *295*, 958.
- (37) Tellez-Valencia, A.; Olivares-Illana, V.; Hernandez-Santoyo, A.; Perez-Montfort, R.; Costas, M.; Rodriguez-Romero, A.; Lopez-Calahorra, F.; Tuena de Gomez-Puyou, M.; Gomez-Puyou, A. *J. Mol. Biol.* **2004**, *341*, 1355.
- (38) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43.
- (39) <http://www.gromacs.org>.
- (40) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306.
- (41) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701.
- (42) Schuler, L. D.; van Gunsteren, W. F. *Mol. Simul.* **2000**, *25*, 301.
- (43) Daura, X.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **1998**, *19*, 535.
- (44) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. *Interaction Models for Water in Relation to Protein Hydration*; Reidel D. Publishing Company: 1981.
- (45) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
- (46) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952.
- (47) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. *J. Comput. Chem.* **1997**, *18*, 1463.
- (48) Sayle, R. A.; Milnerwhite, E. J. *Trends Biochem. Sci.* **1995**, *20*, 374.
- (49) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33.
- (50) <http://pymol.sourceforge.net/>.
- (51) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. *Angew. Chem., Int. Ed.* **1999**, *38*, 236.
- (52) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577.
- (53) Rozovsky, S.; Jogl, G.; Tong, L.; McDermott, A. E. *J. Mol. Biol.* **2001**, *310*, 271.
- (54) Rozovsky, S.; McDermott, A. E. *J. Mol. Biol.* **2001**, *310*, 259.
- (55) Daggett, V. *Chem. Rev.* **2006**, *106*, 1898.
- (56) Wierenga, R. K.; Noble, M. E.; Davenport, R. C. *J. Mol. Biol.* **1992**, *224*, 1115.
- (57) Mattos, C.; Ringe, D. *Nat. Biotechnol.* **1996**, *14*, 595.
- (58) Ringe, D. *Curr. Opin. Struct. Biol.* **1995**, *5*, 825.