

**UNIVERSIDAD DE CHILE**  
**FACULTAD DE CIENCIAS QUÍMICAS Y FARMACÉUTICAS**  
**DEPARTAMENTO DE BIOQUÍMICA Y BIOLOGÍA MOLECULAR**



**ESTUDIO DE DINÁMICA MOLECULAR CON SOLVENTE IMPLÍCITO DE LA  
INFLUENCIA DE LA INTERFASE DE INTERACCIÓN ENTRE LOS DOMINIOS  
DE LA PROTEINA FTSZ EN LA ESTABILIDAD Y PLEGAMIENTO.**

**Memoria para optar al Título de Bioquímico**

**Leonardo Andrés Sepúlveda Durán**

**PATROCINANTE**

**Dr. Davor Cotorás**

**DIRECTORES TESIS**

**Dr. Octavio Monasterio Opazo**

**Dr. Tomás Pérez-Acle**

**2007**

**Universidad de Chile**  
**Facultad de Ciencias Químicas y Farmacéuticas**  
**Departamento de Bioquímica y Biología Molecular**



**ESTUDIO DE DINÁMICA MOLECULAR CON SOLVENTE IMPLÍCITO DE LA  
INFLUENCIA DE LA INTERFASE DE INTERACCION ENTRE LOS DOMINIOS  
DE LA PROTEINA FTSZ EN LA ESTABILIDAD Y PLEGAMIENTO**

Memoria de Título entregada a la Universidad de Chile en cumplimiento parcial de los  
requisitos para optar al Título de Bioquímico

**Leonardo Andrés Sepúlveda Durán**

Dr. Octavio Monasterio Opazo  
Director de Memoria de Título

---

Dr. Tomás Pérez-Acle  
Co-Director de Memoria de Título

---

**Comisión de Evaluación de la Memoria de Título**

Dr. Davor Cotorás Tadic  
Profesor patrocinante

---

Dr. Claudio Olea Azar  
Corrector

---

Dr. Víctor Calvo Pérez  
Corrector

---

**Santiago – Chile - 2007**

*“...Model makers are storytellers ...”*

Ken Dill  
*Molecular Driving Forces*

## Agradecimientos

Quiero agradecer a Tomás Pérez-Acle, quien me acogió en el otrora Centro de Genómica y Bioinformática (CGB, ahora CBUC) de la Universidad Católica cuando era un alumno que ignoraba completamente las artes ocultas de la programación y la Bioinformática. Su apoyo fue fundamental para aprender un arte que requiere dedicación, concentración, y mucha práctica. Asimismo, a los impetuosos integrantes del CGB, cuyo grito de guerra puede resumirse en “¡¡Vamos que se puede!!”

En segundo lugar, quiero agradecer a Octavio Monasterio por acogerme en el Laboratorio de Biología Estructural y Molecular (BEM) de la Facultad de Ciencias de la Universidad de Chile donde fue posible encarnar mis deseos de entender y ver el plegamiento de una proteína y despertar en mi el deseo de aprender las artes de la ciencia experimental. Asimismo quiero agradecer a los entrañables miembros del BEM, A José Jaime, por su ayuda y consejo, a Alexis, María Jesús, Luis, Felipe, Cesar, Verónica y Macarena, por estar ahí y tocar mi sensibilidad.

A María Gabriela Mercado y Mario Tello, por las innumerables tardes de compañía y comprensión en la Soledad de la Ciencia, por su agudeza científica y sobretodo, por el cariño y apoyo emocional en los momentos difíciles.

Y antes de que todo esto fuera posible, al amor incondicional de mi familia, a mi tía Gladis y tía Betty por su cariño, a mi hermana Loreto por sostenerme en mis caídas, a mi padre por enseñarme a tocar guitarra y cantar, y encender la llama de mi curiosidad científica con nuestras discusiones acerca de lo profano y lo divino. Y finalmente a aquella de quien aprendí la disciplina para trabajar y estudiar, aquella que me propulsó en los momentos que mi voluntad claudicaba, gracias a Margarita, mi Madre.

## ÍNDICE GENERAL

<b>ÍNDICE GENERAL</b> .....	<b>I</b>
<b>ÍNDICE DE FIGURAS</b> .....	<b>III</b>
<b>ÍNDICE DE TABLAS</b> .....	<b>V</b>
<b>ABREVIATURAS</b> .....	<b>VI</b>
<b>RESUMEN</b> .....	<b>VII</b>
<b>SUMMARY</b> .....	<b>IX</b>
<b>INTRODUCCIÓN</b> .....	<b>1</b>
LA PROTEÍNA FTSZ. ....	2
PLEGAMIENTO DE PROTEÍNAS.....	5
DETERMINACIÓN DEL PLEGAMIENTO PROTEICO POR MEDIO DE SIMULACIÓN MOLECULAR .....	6
<b>HIPOTESIS</b> .....	<b>8</b>
OBJETIVO GENERAL .....	8
OBJETIVOS ESPECÍFICOS .....	9
<b>MATERIALES Y MÉTODOS</b> .....	<b>10</b>
ELECCIÓN DE LAS ESTRUCTURAS CRISTALINAS USADAS EN LAS SIMULACIONES .....	10
DEFINICIÓN DE LOS DOMINIOS DE FTSZ .....	12
PROTOCOLO DE SIMULACIÓN CON EEF1. ....	14
SIMULACIONES A 300 K Y A TEMPERATURAS DE DESPLEGAMIENTO .....	17
ALGORITMO DE ANÁLISIS DE LA ESTRUCTURA SECUNDARIA .....	17
ANÁLISIS DEL CAMBIO DE ESTRUCTURA SECUNDARIA DURANTE EL DESPLEGAMIENTO .....	18
DEFINICIÓN DE LOS NÚCLEOS HIDRÓFOBOS EN PaFtsZ Y MjFtsZ .....	21
DEFINICIÓN DE CONTACTOS NATIVOS .....	22
RMSD (ROOT MEAN SQUARE DEVIATION) .....	22
RMSF (ROOT MEAN SQUARE FLUCTUATION) .....	23
SASA (SOLVENT ACCESIBLE SURFACE AREA) .....	23
<b>RESULTADOS</b> .....	<b>25</b>
ESTABILIDAD DE LAS ESTRUCTURAS EN SIMULACIONES A 300 K .....	25
<i>Cambio de estructura secundaria durante la simulación</i> .....	27
<i>Cálculo de fluctuaciones de los carbonos <math>\alpha</math></i> .....	29
SIMULACIONES DE DESPLEGAMIENTO TÉRMICO: ASPECTOS CUALITATIVOS.....	30
<i>Simulaciones del desplegamiento térmico de PaFtsZ</i> .....	30
<i>Simulaciones del desplegamiento térmico de MjFtsZ</i> .....	32
SIMULACIONES DE DESPLEGAMIENTO: ASPECTOS CUANTITATIVOS.....	35
<i>Comportamiento de los residuos de aminoácidos durante el desplegamiento de las estructuras secundarias: análisis de matrices R</i> .....	35
INFLUENCIA DE LA INTERFASE EN LA ESTABILIDAD DE LOS DOMINIOS DE FTSZ: VALORES $S_{1/2}^M$ .....	38
<i>Estabilidad de la estructura secundaria de los dominios aislados de PaFtsZ</i> .....	39

<i>Estabilidad de la estructura secundaria de los dominios aislados de MjFtsZ</i> .....	40
CONTACTOS NATIVOS DE LOS NÚCLEOS HIDRÓFOBOS .....	42
<i>Estabilidad relativa de los núcleos hidrófobos</i> .....	45
<i>Influencia de la interfase en la estabilidad de los núcleos hidrófobos</i> .....	48
<i>Análisis de accesibilidad al solvente de los residuos cambiados en las mutantes de triptófano de EcFtsZ: F40W, F135W, F275W e I294W</i> .....	49
<b>DISCUSIÓN</b> .....	<b>54</b>
EVALUACIÓN DEL MÉTODO DE SOLVENTE IMPLÍCITO .....	54
CAMBIOS CONFORMACIONALES A 300 K.....	57
COMPARACIÓN DE LA ESTABILIDAD DE MJFtsZ CON LA DE PAftsZ.....	60
INFLUENCIA DE LA INTERFASE DE INTERACCIÓN EN LA ESTABILIDAD DE LOS DOMINIOS.....	62
<b>CONCLUSIONES</b> .....	<b>67</b>
<b>PERSPECTIVAS FUTURAS</b> .....	<b>67</b>
<b>REFERENCIAS</b> .....	<b>69</b>
<b>ANEXO I</b> .....	<b>74</b>
<b>ANEXOII</b> .....	<b>83</b>
<b>ANEXO III</b> .....	<b>86</b>
<b>ANEXO IV</b> .....	<b>87</b>

## ÍNDICE DE FIGURAS

<b>Figura 1.</b>	Comparación estructural de FtsZ y tubulina	<b>4</b>
<b>Figura 2.</b>	Comparación estructural de PaFtsZ y MjFtsZ y definición de dominios de FtsZ	<b>13</b>
<b>Figura 3.</b>	Protocolo no conservativo aumenta la velocidad del desplegamiento de Cl2 sin cambiar la vía de desplegamiento	<b>16</b>
<b>Figura 4</b>	Protocolo para comparar el desplegamiento de las proteínas PaFtsZ y sus dominios aislados	<b>20</b>
<b>Figura 5</b>	Núcleos hidrófobos de MjFtsZ	<b>21</b>
<b>Figura 6</b>	Estabilidad de las simulaciones a 300 K de MjFtsZ y PaFtsZ	<b>26</b>
<b>Figura 7</b>	Comportamiento temporal de la estructura secundaria de simulaciones a 300 K	<b>28</b>
<b>Figura 8</b>	Fluctuaciones de carbonos $\alpha$ de simulaciones a 300 K	<b>29</b>
<b>Figura 9</b>	Efecto de la temperatura sobre la estructura de PaFtsZ	<b>31</b>
<b>Figura 10</b>	Efecto de la temperatura sobre el desplegamiento de PaFtsZ	<b>34</b>
<b>Figura 11</b>	Matrices $R$ de estructura secundaria para las simulaciones del desplegamiento de PaFtsZ a 300, 400, 450 y 500 K	<b>36</b>
<b>Figura 12</b>	Matrices $R$ de la estructura secundaria para simulaciones del desplegamiento de MjFtsZ a 300, 400, 450 y 500 K	<b>37</b>
<b>Figura 13</b>	Comparación del valor de estructura secundaria medio para PaFtsZ silvestre ( $S_{1/2}^m$ ) y el de sus dominios aislados ( $S_{1/2}^{dom}$ ).	<b>40</b>
<b>Figura 14</b>	Comparación del valor de estructura secundaria medio para MjFtsZ silvestre ( $S_{1/2}^m$ ) y el de los dominios aislados ( $S_{1/2}^{dom}$ )	<b>42</b>
<b>Figura 15</b>	Contactos en la estructura nativa de MjFtsZ y PaFtsZ	<b>44</b>
<b>Figura 16</b>	Conducta cinética del porcentaje de contactos nativos <sup>6</sup> Q de los núcleos hidrófobos de PaFtsF y MjFtsZ	<b>46</b>

<b>Figura 17</b>	Efecto de la temperatura en el desplegamiento de los núcleos hidrófobos de PaFtsZ y MjFtsZ	<b>47</b>
<b>Figura 18</b>	Efecto de la interfase de interacción en la estabilidad de los núcleos hidrófobos	<b>49</b>
<b>Figura 19</b>	Cambio en el porcentaje de accesibilidad al solvente de los residuos de PaFtsZ y MjFtsZ que se reemplazaron en las mutantes de EcFtsZ	<b>51</b>
<b>Figura 20</b>	Superposición estructural de las estructuras iniciales de PaFtsZ y MjFtsZ con las obtenidas después de 5ns de simulación, por medio de ajuste del dominio amino	<b>57</b>
<b>Figura 21</b>	Contactos cristalográficos entre el dominio carboxilo de PaFtsZ y el inhibidor SulA.	<b>59</b>
<b>Figura 22</b>	Agentes caotrópicos y sondas fluorescentes utilizados en los estudios de desplegamiento de FtsZ	<b>64</b>
<b>Figura 23</b>	Modelo de desplegamiento de FtsZ basado en simulaciones y datos experimentales	<b>66</b>



## ÍNDICE DE TABLAS

<b>Tabla 1</b>	Resumen de las características de las estructuras cristalinas de FtsZ disponibles en la base de datos PDB.	<b>11</b>
<b>Tabla 2</b>	Rango de los residuos que comprenden los dominios amino y carboxilo de PaFtsZ y MjFtsZ	<b>14</b>
<b>Tabla 3</b>	Simulaciones de desplegamiento	<b>17</b>
<b>Tabla 4</b>	Número de contactos nativos de los núcleos hidrófobos	<b>45</b>
<b>Tabla 5</b>	Resumen de los resultados del análisis de accesibilidad al solvente de los residuos equivalentes a las mutantes de triptófano F40W, F135W, F275W e I294W de EcFtsZ	<b>53</b>

## ABREVIATURAS

<b>Fts</b>	Proteína que genera filamentación asociada a la temperatura ( <i>Filamentous temperature sensitive protein</i> ).
<b>RMSD</b>	Raíz de la desviación cuadrática media ( <i>Root Mean Square Deviation</i> ).
<b>GdmCl</b>	Cloruro de guanidinio ( <i>Guanidinium Chloride</i> ).
<b>EcFtsZ</b>	FtsZ de <i>Escherichia coli</i> .
<b>MjFtsZ</b>	FtsZ de <i>Methanococcus jannaschii</i> .
<b>CD</b>	Dicroísmo circular ( <i>circular dichroism</i> ).
<b>CI2</b>	Inhibidor de quimiotripsina 2 ( <i>Chymotrypsin Inhibitor 2</i> ).
<b>ENCAD</b>	Cálculos de energía y dinámica ( <i>ENergy CACalculations and Dynamics</i> ).
<b>PaFtsZ</b>	FtsZ de <i>Pseudomonas aeruginosa</i> .
<b>PDB</b>	Base de datos de proteínas (Protein Data Bank).
<b>AMBER</b>	Construcción asistida de modelos con refinamiento energético ( <i>Assisted Model Building with Energy Refinement</i> ).
<b>GROMOS</b>	Programa de simulación molecular de Groningen ( <i>GROningen Molecular Simulation Software</i> ).
<b>CHARMM</b>	Química en Harvard y mecánica macromolecular ( <i>Chemistry at HARvard Macromolecular Mechanics</i> ).
<b>EEF1</b>	Función de energía efectiva 1 ( <i>Effective Energy Function 1</i> ).
<b>ABNR</b>	Método de Newton-Raphson con base adaptada ( <i>Adapted Basis Newton Raphson</i> ).
<b>STRIDE</b>	Identificación de estructura secundaria ( <i>secondary STRucture IDentification</i> ).
<b>BLAST</b>	Herramienta de búsqueda de alineamientos locales básicos ( <i>Basic Local Alignment Search Tool</i> ).
<b>Matriz S</b>	Matriz de estructura Secundaria.
<b>Matriz R</b>	Matriz de estructura secundaria Residual.
<b>VMD</b>	Dinámica molecular visual. ( <i>Visual Molecular Dynamics</i> ).
<b>SASA</b>	Área superficial accesible a solvente ( <i>Solvent Accesible Surface Area</i> ).
<b>DSSP</b>	Definir estructura secundaria de proteínas ( <i>Define Secondary Structure of Proteins</i> ).
<b>FITC</b>	Isocianato de fluoresceína. ( <i>Fluorescein Isothiocyanate</i> ).
<b>ANS</b>	ácido 8-anilino-1-naphtaleno-sulfónico ( <i>8-anilino-1-naphthalene-sulfonic acid</i> ).

## RESUMEN

FtsZ es una proteína bacteriana esencial en el proceso de división celular. Se localiza en la membrana interna de la célula generando un anillo en la mitad de su eje longitudinal, donde actúa como una proteína de andamiaje reclutando a las demás proteínas del mecanismo de división. Esta estructura denominada “anillo Z” esta compuesta de polímeros de FtsZ, los cuales son capaces de interactuar lateralmente generando sabanas o dobles filamentos. La polimerización se produce por la unión de GTP al monómero de FtsZ. Dentro del filamento, FtsZ adquiere actividad GTPásica, y tras la hidrólisis del GTP el filamento de FtsZ se desestabiliza y desensambla.

Los cambios estructurales producidos tras la hidrólisis del GTP pueden explicarse al analizar la estructura de FtsZ, que esta compuesta por dos dominios, el amino que contiene el sitio de unión a nucleótido del tipo Rossman y el carboxilo que presenta un plegamiento de tipo  $\alpha/\beta$ . Ambos interactúan estrechamente por medio de una interfase hidrofóbica generada entre las sabanas plisadas de ambos dominios y la helice H7, la que se encuentra entre ambos dominios. Esta interfase no solo tiene relevancia funcional en el mecanismo de polimerización, sino que también juega un rol estabilizador entre los dominios de FtsZ. Mientras que dominios de una FtsZ de *Thermotoga maritima* son estables en forma independiente, los dominios de FtsZ de *Escherichia coli* solo muestran estructura secundaria en presencia de agentes estabilizantes.

Con el fin de estudiar la influencia de esta interfase de interacción en la estabilidad de los dominios de FtsZ, se llevaron a cabo simulaciones de dinámica molecular con un método de solvatación implícita. La utilización de este método nos permitió efectuar simulaciones en condiciones nativas y desnaturantes (alta temperatura) de una proteína FtsZ termófila (*Methanococcus jannaschii*) y una mesófila (*Pseudomonas aeruginosa*) para las cuales existen estructuras cristalinas disponibles.

Las simulaciones mostraron que en condiciones nativas ambas proteínas son estables, sin embargo el dominio de unión a nucleótido de la proteína mesófila muestra una menor estabilidad que el dominio de la proteína termófila. Las diferencias se localizan en las zonas aledañas al lazo T3, involucrado en la unión del nucleótido. La

forma general del desplegamiento de la proteína mesófila y termófila es muy similar, observándose solo diferencias en la rapidez con la cual se pierde estructura secundaria, siendo más estable la proteína termófila. Sin embargo, cálculos de estabilidad de los núcleos hidrófobos de la interfase y los dominios, no mostraron diferencias significativas entre los dominios aislados y las proteínas nativas. Los únicos cambios significativos se logran al comparar las estabilidades del dominio carboxilo con el dominio amino, donde este último muestra una mayor estabilidad. En conjunto, los resultados sugieren una relación entre la estabilización de los dominios por medio de la interfase de interacción y la unión del nucleótido con el mecanismo de cambio conformacional de FtsZ.

## SUMMARY

### (IMPLICIT SOLVENT MOLECULAR DYNAMICS STUDY OF THE INFLUENCE OF THE INTERACTION INTERPHASE BETWEEN FTSZ PROTEIN DOMAINS IN STABILITY AND FOLDING)

FtsZ is a bacterial protein essential for the cellular division process. It localizes in the internal membrane forming a ring in the longitudinal center of the cell, recruiting other cellular division proteins. This “Z ring” is composed of FtsZ polymers which are able to interact laterally to form sheets of multiple filaments. Polymerization takes place after GTP binding to the FtsZ monomer. Inside the filament, FtsZ acquire GTPase activity, which creates GDP which destabilize filaments, causing its disassembly.

The structural changes occurring after GTP hydrolysis can be explained based on the FtsZ structure. This protein is composed by two domains; the amine domain displays Rossmann GTP binding fold and the carboxyl domain an  $\alpha/\beta$  fold. Both interact tightly through an hydrophobic intephase generated between the domain's beta sheets and the central helix H7. This interphase not only has a functional role in the polymerization mechanism, but also plays a stabilizing role between the FtsZ domains. While the isolated domains of *Thermothoga maritima* FtsZ are independently stable, the FtsZ *Escherichia coli* domains only shows secondary structure when stabilizing agents are added to the solution.

The aim of this thesis was to study the influence of the interaction interphase in the FtsZ domain stability. The method used was molecular dynamics simulations with implicit solvent in denaturant and native conditions of a termophilic (*Methanococcus jannaschii*) and a mesophilic (*Pseudomonas aeruginosa*) FtsZ protein, which have X-ray structures available.

The simulations showed that under native conditions both proteins are stable, nevertheless the nucleotide binding domain of the mesophilic protein shows a lower stability than the one of the termophilic protein. The differences are localized around the T3 loop, which is involved in the nucleotide binding. The general form of the unfolding pathway is very similar for both proteins, but the kinetics of the secondary structure unfolding differ, being the velocity of secondary structure loss of the termophilic protein

smaller than the one of the mesophilic protein. However, stability analysis of the hydrophobic nuclei of the interphase and the protein domains did not show differences between the isolated domains and the native proteins. The only meaningful changes were obtained comparing the stability of the amine and carboxyl domains, being the amine domain of higher stability in both proteins. Overall, the results suggest a relationship between interdomain stabilization, ligand binding, and the conformational change mechanism of FtsZ.

## INTRODUCCIÓN

### **División bacteriana.**

Durante su ciclo de vida, una bacteria consume nutrientes de su entorno, transformándolos en componentes de su membrana y pared celulares, maquinaria metabólica y material genético, entre otros. Sin embargo, el crecimiento en biomasa no garantiza la separación de la bacteria en dos células independientes; este proceso, llamado división celular, requiere de la septación de las membranas en la mitad de la célula. Este proceso se inicia con la formación del llamado *divisoma*, una estructura molecular compleja, que está compuesto por proteínas que interactúan en forma coordinada para separar el material genético duplicado.

La septación se produce en la mitad longitudinal de la bacteria para evitar la generación de una progenie asimétrica, gracias a la existencia en otros lugares de la célula de proteínas inhibitoras del divisoma tales como MinC, MinD y MinE (Rothfield y cols. 1999), la cuales generan un gradiente de concentración que disminuye desde los polos bacterianos, evitando de ese modo la división en zonas distales. Estas proteínas actúan inhibiendo a las proteínas Fts (*Filamentous temperature sensitive proteins*), las cuales son las proteínas estructurales del divisoma (Rothfield y Justice 1997). La primera proteína en localizarse es la FtsZ, que se piensa genera un anillo, que se fija a la cara interna de la membrana celular por medio de las proteínas ZipA y FtsA (protoanillo), al cual se unen posteriormente FtsK, FtsQ, FtsB-FtsL, FtsW, FtsI, y FtsN para formar el llamado divisoma, que contrae la membrana y pared bacterianas produciendo la división (Romberg y Levin 2003). Recientemente se postuló un modelo denominado “iterative pinching” donde FtsZ genera la fuerza que contrae la membrana a través de un ciclo de polimerización dependiente de GTP, unión a la membrana, cambio conformacional, depolimerización e intercambio del nucleótido (Li y cols. 2007).

## La proteína FtsZ.

FtsZ es la única proteína de división presente en todas las bacterias y arqueas cuyo genoma ha sido secuenciado, a excepción de los miembros de la familia *Chlamidiaceae* y la especie *Ureaplasma urealyticum* (Vaughan y cols. 2004). Se ha demostrado que el anillo Z está constituido de filamentos de FtsZ cuya formación está mediada por la unión de GTP y  $Mg^{2+}$  a monómeros de FtsZ (Fig. 1E), lo que produce su polimerización (Bramhill y Thompson 1994). Una vez en el filamento FtsZ hidroliza GTP a GDP, que genera un cambio conformacional que desestabiliza los filamentos, curvándolos y produciendo su posterior desensamblaje. El acoplamiento entre hidrólisis y desensamblaje le confiere dinamismo al anillo Z, cuyos monómeros tienen un tiempo de vida media de permanencia en el anillo de aproximadamente 30 segundos (Stricker y cols. 2002). Adicionalmente, FtsZ puede generar sabanas de filamentos *in vitro* cuando se utiliza un pH cercano a 6.0 (Bramhill y Thompson 1994), policondicionamiento tales como DEAE-dextrano (Mukherjee y Lutkenhaus 1994), así como también altas concentraciones de calcio (Lowe y Amos 1999), glutamato (Beuria y cols. 2003) o DAPI (Nova y cols. 2007).

La actividad GTPásica intrínseca de FtsZ, así como su capacidad de formar filamentos tanto *in vivo* como *in vitro* evidenció que FtsZ era homóloga a la proteína eucarionte tubulina, la cual forma filamentos y estructuras macromoleculares como los microtúbulos (Fig. 1F) (Desai y Mitchison 1997). Tras la obtención de un modelo tridimensional de un dímero de  $\alpha\beta$  tubulina por medio de cristalografía electrónica de sabanas inducidas por  $Zn^{2+}$  (Nogales y cols. 1998b) y de un monómero de FtsZ de *Methanococcus jannaschii* a través de cristalografía de rayos X (Lowe 1998), se comprobó que a pesar de tener 7% de identidad de secuencia, es posible superponer las hélices  $\alpha$  o sabanas  $\beta$  de estas proteínas con un RMSD (*Root Mean Square Deviation*) de carbonos  $\alpha$  de 2.4 Å, siendo las zonas de los lazos las que presentan las diferencias más notorias, existiendo largas inserciones en tubulina (Andreu y cols. 2002) (Fig. 1A-D, residuos rojos). FtsZ está compuesta por dos dominios, el dominio amino y el dominio carboxilo, los cuales son equivalentes a los dominio amino e intermedio de  $\alpha$  y  $\beta$  tubulinas. El dominio amino posee un plegamiento tipo Rossmann,



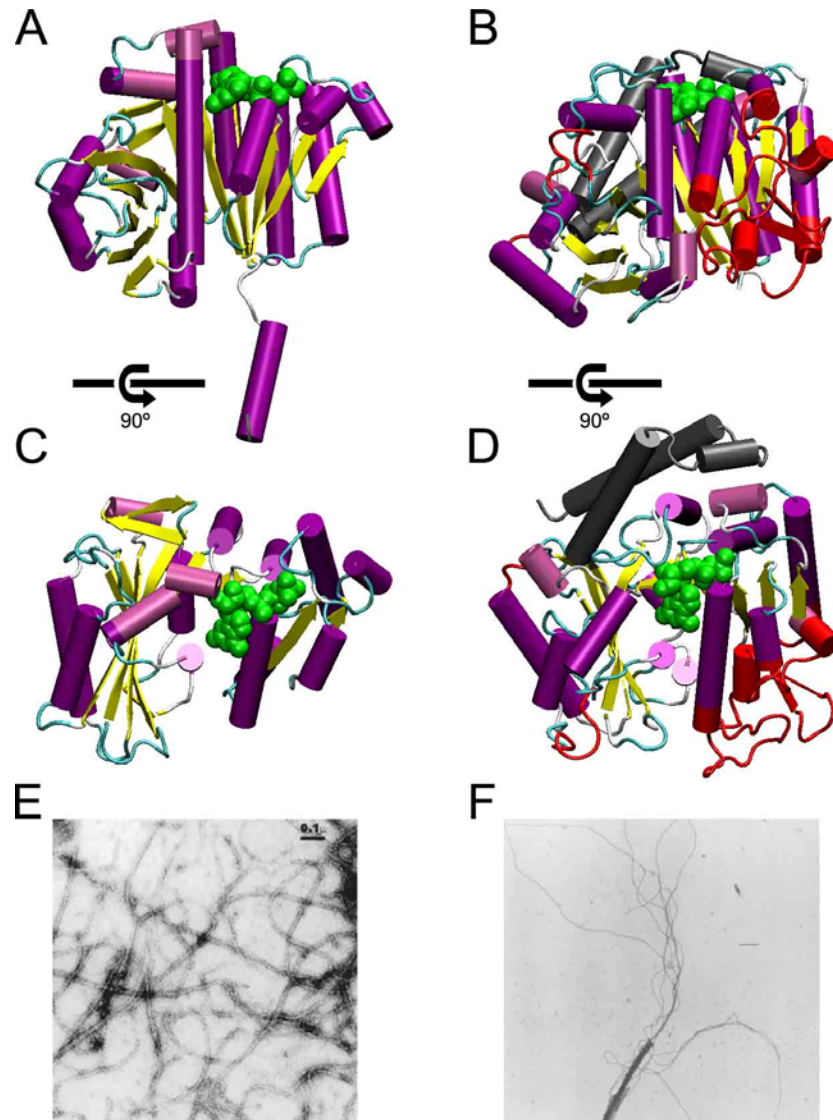
el cual presenta el sitio de unión a nucleótido, compuesto por 6 lazos (T1 a T6) donde se encuentran la mayoría de los residuos conservados. El lazo T4 presenta el llamado motivo de “firma de tubulina”, patrón PROSITE PS00227, [SAG]GGTG[SA]G, el que junto con los otros lazos con contactos con el nucleótido distinguen a FtsZ y tubulina como una familia de GTPasas distinta a las GTPasas clásicas como p21(ras) (Nogales y cols. 1998a). Unido al dominio amino por medio de una interfase hidrofóbica se encuentra el dominio carboxilo, relacionado estructuralmente a la proteína corismato mutasa de *Bacillus subtilis*.

A pesar de la alta similitud estructural entre FtsZ y tubulina, sus vías de plegamiento son distintas. En primer lugar, estudios de desplegamiento con cloruro de guanidinio (GdmCl) seguidos por dicroísmo circular, muestran que tubulina no llega a un estado nativo tras la extracción del agente caotrópico, mientras que bajo las mismas condiciones FtsZs de *Escherichia coli* (EcFtsZ) y *Methanococcus jannaschii* (MjFtsZ) son capaces de replegarse y polimerizar (Andreu y cols. 2002). En segundo lugar, tubulina y EcFtsZ pierden estructura secundaria tras la liberación del nucleótido desde el sitio activo del dominio amino, mostrando un desplegamiento en dos etapas, por el contrario, MjFtsZ se pliega en una sola transición, lo que sugiere que la ausencia de ligando no desestabiliza la estructura como ocurre en el caso de EcFtsZ y tubulina.

Quimeras de MjFtsZ a las cuales se le insertaron lazos de tubulina (zonas rojas, Figura 1) mostraron un perfil de desplegamiento irreversible por la presencia de intermediarios, sugiriendo que estos no se producen por desestabilización del dominio amino por pérdida del nucleótido como en el caso de EcFtsZ, sino más bien a la presencia de los lazos de tubulina que impiden su completo plegamiento en ausencia de chaperonas (Bertrand y cols. 2005).

El plegamiento de MjFtsZ es de dos estados lo que sugiere que el plegamiento de sus dominios ocurre en forma cooperativa. Los dominios de FtsZ de *Thermotoga maritima* se expresaron en forma independiente sin cambios mayores en estructura secundaria o terciaria y se restituyó la actividad GTPasa de la proteína al juntar los dominios *in vitro* (Anfinsen y Haber 1961; Anfinsen y cols. 1961). Por el contrario, los

dominios aislados de EcFtsZ parece que se encuentran mal plegados, pues el contenido de estructura secundaria de estos es diferente a la de la proteína silvestre;



**Figura 1. Comparación estructural de FtsZ y tubulina. Paneles A y C;** estructura molecular de FtsZ (PDBid 1W5B, monómero A). **Paneles B y D;** estructura molecular de beta tubulina (PDBid 1JFF). Los cilindros morados corresponden a hélices  $\alpha$ , cilindros rosados hélices  $\pi$ , flechas amarillas hebras  $\beta$ , tubos celeste vueltas  $\beta$ , tubos blancos lazos. Los ligandos GDP (1W5B) y GTP (1JFF) se muestran en esferas de van der Walls verdes. Los lazos de tubulina que no se encuentran en FtsZ se muestran en color rojo y el dominio carboxilo terminal de tubulina se muestra en color gris. **Paneles E y F;** micrografía electrónica de los filamentos formados por FtsZ en presencia de GTP,  $Mg^{2+}$  y KCl (Mukherjee & Lutkenhaus, 1999) y de un microtúbulo y los protofilamentos que lo componen (Caplow 1994), respectivamente. Figuras A a la D fueron creadas con el programa VMD (Humphrey y cols. 1996).

como se observó por dicroísmo circular (Felipe Montesinos, comunicación personal), indicando que la superficie de interacción de los dominios poseería un papel importante en la estabilidad de los dominios de la proteína. Consistente con estos datos, la desnaturalización al equilibrio de mutantes de triptófano de EcFtsZ (F40W, F135W, F275W y I294W), mostraron curvas de desnaturalización con valores de  $[GdmCl]_{50\%}$  similares, no importando si las mutaciones se encontraban en el dominio amino o carboxilo. Estos resultados muestran que la interfase de interacción entre los dominios de FtsZ juega un rol preponderante en la estabilidad y plegamiento. Sin embargo, a partir de los resultados experimentales de desnaturalización al equilibrio no es posible tener una visión detallada del proceso de plegamiento, pues corresponde a una visión macroscópica del proceso, donde los parámetros termodinámicos dan cuenta del comportamiento de una solución compuesta de muchas moléculas de proteína. Una descripción detallada del proceso requiere de un análisis de estos cambios a nivel microscópico, para lo cual es necesario complementar los estudios empíricos con simulaciones del proceso de desplegamiento.

### **Plegamiento de proteínas**

Uno de los desafíos más importantes de la Biología Estructural actual es la dilucidación del mecanismo de plegamiento de las proteínas. Grandes avances se han hecho desde el experimento fundador de Anfinsen (Shea y Brooks III 2001), donde se demostró por primera vez que una proteína era capaz de plegarse en forma independiente sin necesidad de catalizadores externos. Este descubrimiento suscitó un gran interés y prontamente se empezaron a estudiar pequeñas proteínas entre las cuales se encuentra el inhibidor de quimiotripsina (CI2), que presenta solo una transición durante el plegamiento tanto en experimentos de equilibrio como en experimentos cinéticos (Snow y cols. 2005; Snow y cols. 2004). La interpretación termodinámica de los resultados mostró al plegamiento como la difusión desde un colectivo de estructuras extendidas y fluctuantes (estado desplegado) a un colectivo con menor diversidad estructural (estado nativo), pasando por un estado de transición

(Dinner y Karplus 1999), así, el estudio del desplegamiento daría luces acerca del proceso de plegamiento

### **Determinación del plegamiento proteico por medio de simulación molecular**

Los métodos de simulación molecular han contribuido al entendimiento del proceso de plegamiento, permitiendo estudiar el comportamiento dinámico de moléculas a resolución atómica en la escala de femtosegundos, tanto en condiciones de equilibrio como de no equilibrio (Shea y Brooks III 2001). Para simular condiciones de equilibrio, se efectúan simulaciones múltiples para obtener un muestreo suficiente del espacio conformacional de la proteína y así paisajes de energía libre (*free energy landscapes*), a partir de los cuales es posible calcular datos cinéticos y termodinámicos. Con estos métodos se ha predicho con gran exactitud la estabilidad y cinética de plegamiento de pequeños péptidos (< 20 aminoácidos) a partir de un estado completamente desplegado (Snow y cols. 2005; Snow y cols. 2004). Sin embargo, debido al elevado costo computacional no se han realizado estas simulaciones en proteínas de mayor tamaño. En estos casos, se han utilizado métodos de no equilibrio, donde se perturba el sistema con agentes caotrópicos (usualmente temperatura) con el fin de describir el proceso de desplegamiento de una sola simulación. El proceso de desplegamiento daría luces acerca del proceso de plegamiento, pues simulaciones de modelos reticulares (*lattice models*) indican que el principio de reversibilidad microscópica es aplicable a simulaciones de desplegamiento (Dinner y Karplus 1999).

El grupo de Fersht y Daggett, pionero en el campo de simulaciones de no equilibrio, ha hecho aportes considerables al entendimiento del desplegamiento de la proteína CI2 (Daggett y cols. 1996; Li y Daggett 1994; Li y Daggett 1996), Barnasa (Bond y cols. 1997; Clarke y cols. 2000), Lisozima (Kazmirski y Daggett 1998), así como también al efecto en el plegamiento de agentes caotrópicos como urea (Bennion y Daggett 2003; Bennion y Daggett 2004; Zou y cols. 2002) y condiciones acídicas (Armen y cols. 2004), utilizando el campo de fuerza ENCAD (Levitt y cols. 1995) y solvente explícito. De las proteínas antes mencionadas, CI2 es la que se ha estudiado

en mayor detalle tanto desde un punto de vista experimental como computacional, debido a que representa un modelo de “unidad de plegamiento”, es decir, un dominio pequeño que es capaz de plegarse en forma autónoma en una transición de dos estados (Daggett 2006). Se han hecho más de 100 mutantes de sus 64 residuos (Itzhaki y cols. 1995; Otzen y cols. 1994) y se han calculado los llamados valores  $\Phi$ , los cuales se pueden interpretar como el grado de similitud entre la estructura circundante a un residuo en el estado de transición con respecto al estado nativo (Fersht y Daggett 2002; Fersht y cols. 1992). Simulaciones de dinámica molecular (Apéndice I) han mostrado gran correspondencia con los valores  $\Phi$  experimentales (Daggett y cols. 1996; Li y Daggett 1996). Estas simulaciones han mostrado que el plegamiento de C12 procede por un mecanismo de nucleación-condensación, donde la proteína comienza a plegarse localmente generando una estructura secundaria fluctuante, la cual colapsa dando lugar al núcleo hidrófobo de la proteína, que estabiliza las estructuras secundarias produciendo el estado nativo. Dependiendo del grado de estabilidad de las estructuras secundarias, es posible verificar el mecanismo de difusión-colisión donde las estructuras secundarias preformadas colapsan entre sí, como en el caso de la proteína “engrailed homeodomain”, compuesta de tres hélices que interactúan entre sí en forma paralela generando un núcleo hidrófobo donde participan cadenas laterales de las tres hélices (Daggett y Fersht 2003). También, bajo condiciones de desnaturación ácida, o en presencia de agentes caotrópicos químicos como urea, se observa el estado de “glóbulo fundido” donde la estructura terciaria se forma antes que la secundaria (Kuwajima 1977; Ptitsyn 1995). Independiente de la presencia de estructura secundaria, el mecanismo básico que comparten todos estos modelos es el colapso hidrófobo. Esto se ha verificado en simulaciones Monte Carlo de modelos reticulares de proteínas (*lattice models*), donde se ha visto que el colapso rápido a un estado no nativo semicompacto es suficiente para asegurar que una proteína alcance el estado nativo en la escala de segundos, aun cuando se efectúe una búsqueda de conformaciones completamente aleatoria (Sali y cols. 1994).

En base a los antecedentes anteriormente mencionados, se definieron dos hipótesis:

**HIPOTESIS I**  
**(METODOLÓGICA)**

“Existen métodos de dinámica molecular adecuados para estudiar la influencia de interfases de interacción en la estabilidad de dominios de proteínas”

**HIPÓTESIS II**  
**(FENOMENOLÓGICA)**

“Dado que los dominios de FtsZ del organismo termófilo *Thermotoga marítima* pueden separarse manteniendo su estructura secundaria, lo que no ocurre en dominios de FtsZ de *Escherichia coli*, entonces los dominios de proteínas termófilas son estructuralmente más estables que los dominios de sus contrapartes mesófilas”.

Para probar la hipótesis por métodos de simulación molecular, se define el siguiente objetivo general:

**Objetivo General:**

Simular y analizar el proceso de desplegamiento de la proteína FtsZ de una especie mesófila y una termófila con el fin de obtener una interpretación microscópica de los datos experimentales del desplegamiento de EcFtsZ y MjFtsZ, con énfasis en la influencia de la interfase de interacción sobre la estabilidad de los dominios.

Para lograr el objetivo general, se proponen los siguientes objetivos específicos:

**Objetivos específicos:**

1. Seleccionar de la base de datos PDB (*Protein Data Bank*), una estructura cristalina representativa de proteínas mesófilas y otra de proteínas termófilas para realizar dinámica molecular.
2. Buscar y seleccionar un método de dinámica molecular para estudiar el desplegamiento térmico de las proteínas FtsZ elegidas.
3. Efectuar simulaciones de dinámica molecular para caracterizar el desplegamiento de las proteínas FtsZ elegidas.
4. Desarrollar programas para analizar las diferencias en el plegamiento de los dominios aislados por medio de distintos parámetros estructurales, tales como la integridad de la estructura secundaria y la integridad de los núcleos hidrófobos.

## MATERIALES Y MÉTODOS

### Elección de las estructuras cristalinas usadas en las simulaciones

Para realizar simulaciones de dinámica molecular se requiere estructuras iniciales del sistema a estudiar. Estas pueden ser obtenidas por difracción de rayos X o NMR. Hasta el año 2007, la base de datos PDB (Berman y cols. 2000) tiene depositadas doce estructuras cristalográficas de FtsZ, proveniente de cuatro especies distintas. Ellas difieren en el nivel de resolución, presencia de ligandos y la presencia de zonas no resueltas en la estructura tridimensional (Tabla 1), mientras que no existen estructuras de FtsZ resueltas por NMR. Para realizar comparaciones del comportamiento de una proteína FtsZ mesófila con una termófila, se requiere elegir proteínas de alta similitud estructural y de secuencia, evitando de este modo integrar variables adicionales que puedan explicar las diferencias en el comportamiento dinámico de las proteínas.

Existen cinco estructuras cristalinas de FtsZs mesófilas, tres de *Mycobacterium tuberculosis* (MbtFtsZ) (Leung y cols. 2004a) y una de *Pseudomonas aeruginosa* (PaFtsZ) (Cordell y Lowe 2004). Como primer filtro de selección se consideró solo aquellas estructuras que no tuviesen zonas sin resolver en el cristal. De ellas se eligió PaFtsZ, puesto que posee una mayor identidad de secuencia con EcFtsZ (66% y una similitud de 80%) al alinearlas con el programa de alineamiento local BLAST (Altschul y cols. 1997)

Para la variante termófila, existen siete estructuras, seis de *Methanococcus jannaschii* (MjFtsZ) (Lowe 1998; Oliva y cols. 2004) y una de *Thermotoga marítima* (TmFtsZ) (Oliva y cols. 2004). La estructura de TmFtsZ corresponde a un dímero permutado (*swapped dimer*) donde el dominio amino del monómero A interactúa con el dominio carboxilo del monómero B. Eso deja a MjFtsZ como la única especie termófila con estructuras cristalizadas funcionales. Al igual que con la variante mesófila, se desestimaron las estructuras cristalinas que tuviesen zonas sin resolver. Las cuatro



estructuras restantes tienen GTP y Mg<sup>2+</sup> o GDP unido, y la resolución varía entre los 2,2 y 3 Å.

**Tabla 1. Resumen de las características de las estructuras cristalinas de FtsZ disponibles en la base de datos PDB.** Para cada organismo, se lista cada una de las estructuras cristalinas disponibles. En el caso de existir más de un monómero de FtsZ en la celda unitaria del cristal, se lista cada uno de los monómeros con la nomenclatura del archivo PDB(A,B). El grado de certidumbre en las coordenadas esta dada por la resolución de las estructuras y su factor R libre, mientras menor su valor, mejor el modelo. En negrita se muestran las estructuras que fueron elegidas para llevar a cabo las simulaciones de dinámica molecular.

Organismo	PDBid	Resolución (Å)	Factor R <sub>libre</sub>	Monómero	Residuos Totales	Ligandos	Lista de zonas sin resolver
<i>Methanococcus jannaschii</i>	1fsz	2,8	0,28		333	GDP	No
	1w5a	2,4	0,26	A	333	GTP, Mg <sup>2+</sup>	No
				B	335		
	1w59	2,7	0,3	A	351	SO4	17 – 20
				B	342		No
	<b>1w5b</b>	<b>2,2</b>	<b>0,26</b>	<b>A</b>	<b>333</b>	<b>GTP</b>	<b>No</b>
				<b>B</b>	<b>335</b>		
	1w58	2,5	0,25	1	338	G2P, Mg <sup>2+</sup>	357 -359
				A	331		
				B	332		
C				331			
D				331			
1w5e	3	0,3	E	332	GTP	No	
			F	331			
			G	331			
			H	332			
			I	331			
<i>Thermotoga maritima</i>	1w5f	2	0,23	A	315	G2P, Mg <sup>2+</sup>	No
				B	315		
<i>Pseudomonas auruginosa</i>	<b>1ofu</b>	<b>2,1</b>	<b>0,26</b>	<b>A</b>	<b>306</b>	<b>GDP</b>	<b>No</b>
<i>Mycobacterium tuberculosis</i>	1rlu	2,1	0,22	A	304	GSP	No
				B	306	GOL	59 - 70 169 -174
	1rq2	1,9	0,22	A	304	CIT	63 – 69
				B	305		59 - 70 170 -174
	1rq7	2,6	0,24	A	304	GDP	63 – 69
			B	307	59 - 70 170 -174		

Se eligió el monómero A de la estructura cristalina 1W5B en lugar de 1FSZ para favorecer una mejor resolución sobre la coincidencia en el ligando unido de PaFtsZ (1OFUA), pues se han encontrado diferencias en la estabilidad de estructuras de proteínas obtenidas por cristalografía de rayos X y NMR que se han estudiado por simulaciones de dinámica molecular (Fan y Mark 2003) y se ha observado que una menor resolución produce un mayor grado de incertidumbre en la posición relativa de los átomos, y con ello en las simulaciones.

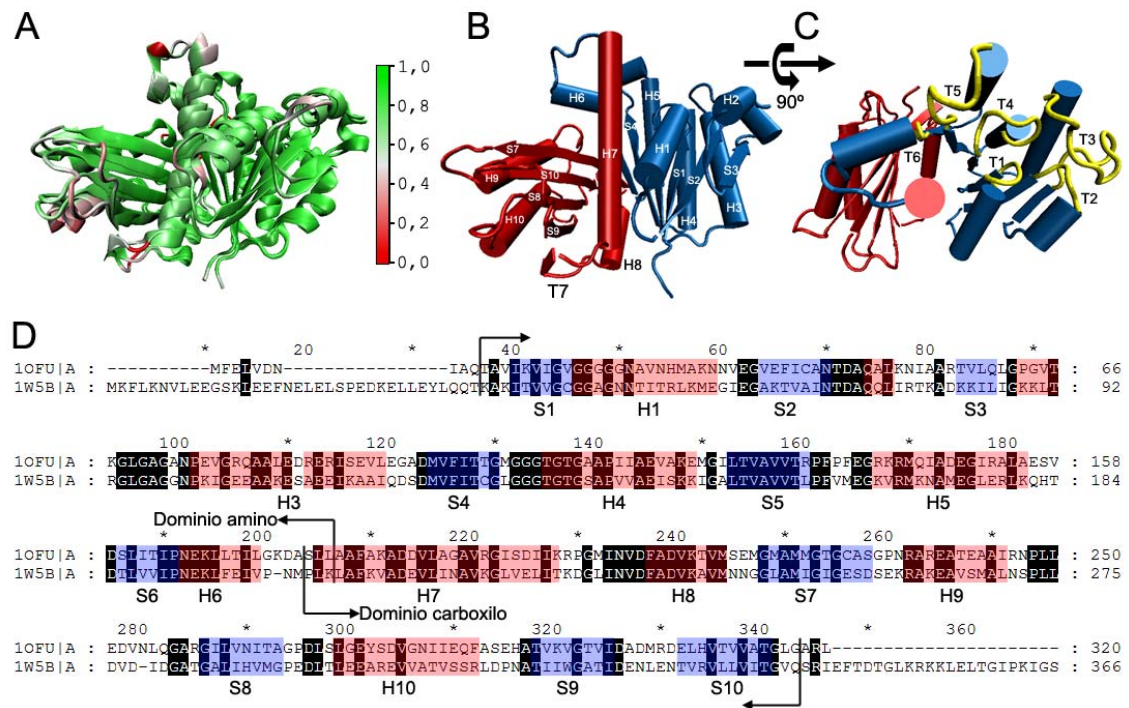
Al superponer estructuralmente PaFtsZ y MjFtsZ, se puede observar que la mejor superposición se logra en los elementos de estructura secundaria como hélices  $\alpha$  y hebras  $\beta$ , mientras que los lazos muestran las mayores diferencias, siendo estas más evidentes en el dominio carboxilo (Fig. 2A), lo que resulta consistente con los mayores factores de temperatura cristalográficos presentados por este dominio (Diaz-Espinoza y cols. 2007). Fuera de estas diferencias locales, ambas proteínas muestran una buena superposición (RMSD de  $C\alpha$  de 3,6 Å)

### **Definición de los dominios de FtsZ**

Los dominios de FtsZ interactúan a través de una interfase hidrofóbica, compartiendo la hélice H7. Estudios con MtFtsZ han mostrado que al dejar la hélice H7 y el lazo T7 en el dominio amino, este es capaz de polimerizar rápidamente e hidrolizar GTP (Borhani y White 2004; Gupta y cols. 2004; Leung y cols. 2004b), puesto que el lazo T7 tiene el residuo de ácido glutámico requerido para la hidrólisis de GTP. Por otro lado, estudios de separación de dominios en TmFtsZ muestran que el dominio amino solo es estable si no se considera a H7 y el lazo T7 como parte de este dominio (Oliva y cols. 2004). Tomando esto en consideración, se definió el dominio carboxilo incluyendo a H7 y T7 (Fig. 2B y 2C).

Para definir la zona exacta de separación de los dominios, las estructuras de MjFtsZ y PaFtsZ se alinearon estructuralmente con el programa STAMP (Russell y Barton 1992) (Fig. 2A y 2D). Se seleccionaron tres residuos en común entre los dominios amino y carboxilo para evitar que la presencia de cargas adicionales

producidas por el corte de los dominios afecte la estructura de H6 y H7 (Fig 2D, flechas internas). Adicionalmente, se eliminaron de la proteína MjFtsZ aquellos residuos que no tuvieran un equivalente en PaFtsZ (Fig. 2D, flechas externas), con el fin de hacer comparaciones entre proteínas con el mismo número de residuos. La Tabla 2 muestra un resumen de los bordes y tamaño de los dominios.



**Figura 2. Comparación estructural de PaFtsZ y MjFtsZ y definición de dominios de FtsZ. Panel A.** Superposición estructural de MjFtsZ (1W5B monómero A) y PaFtsZ (1OFU monómero B). Zonas en verde muestran buena superposición estructural, mientras que zonas en rojo corresponden a zonas con diferencias estructurales. **Panel B.** y **C.** Dominios de MjFtsZ. En azul se muestra el dominio amino y en rojo el dominio carboxilo, los lazos que unen al nucleótido se muestran como tubos amarillos. **Panel D.** Alineamiento de secuencias de MjFtsZ y PaFtsZ. Los residuos conservados se muestran en fondo negro, las hebras  $\beta$  con fondo azul y las hélices  $\alpha$  en fondo rojo. Los límites utilizados para definir los dominios a simular se muestran con flechas negras. El alineamiento se obtuvo a partir de una superposición estructural de ambas secuencias utilizando STAMP. La identidad de secuencia entre ambas proteínas es de 43%, y una similitud de 64% utilizando la matriz de sustitución BLOSUM62.

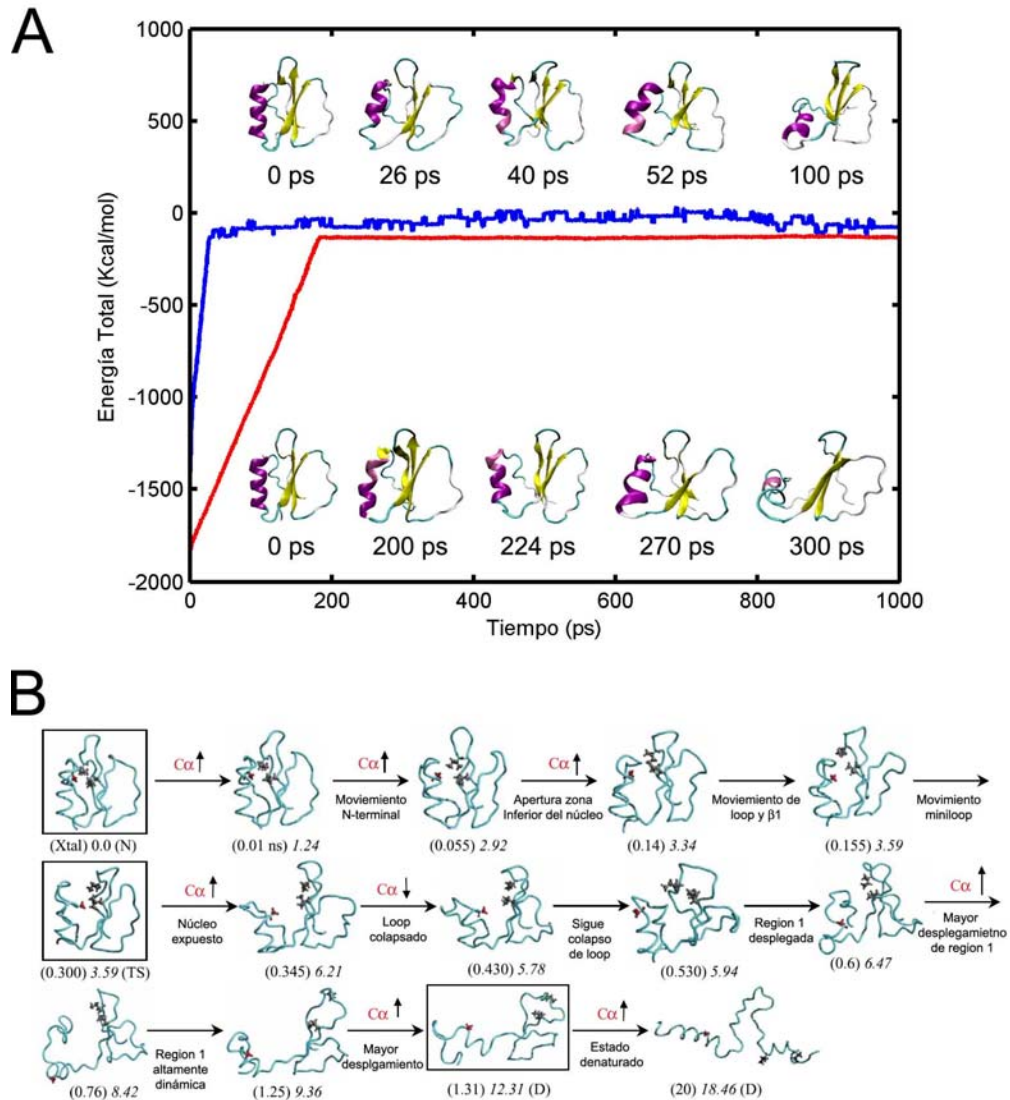
**Tabla 1. Rango de residuos que comprenden los dominios amino y carboxilo de PaFtsZ y MjFtsZ.** La numeración corresponde a las usadas en las estructuras cristalinas 1W5B (MjFtsZ) y 1OFU (PaFtsZ). En paréntesis se muestra el número total de residuos tanto de los dominios amino y carboxilo como de la proteína silvestre.

Proteína	Silvestre	Dominios	
		Amino	Carboxilo
PaFtsZ	10 – 317 (306)	11 – 180 (170)	178 – 317 (139)
MjFtsZ	36 – 341 (306)	36 – 205 (170)	203 – 341 (139)

### Protocolo de simulación con EEF1.

En las simulaciones se utilizó el método de solvatación implícita EEF1 (Lazaridis y Karplus 1999), dentro del programa de simulación CHARMM (Brooks y cols. 1983). Cada simulación involucró dos partes. En la primera se realizó una minimización energética de la estructura utilizando 1000 pasos del algoritmo ABNR (*Adapted Basis Newton-Raphson*). Posteriormente, la estructura minimizada se sometió a una simulación de dinámica molecular a 300, 400, 450 o 500 K. Como primera aproximación, se utilizó el protocolo descrito para realizar el desplegamiento de la proteína CI2 utilizando el solvente implícito EEF1 (Lazaridis y Karplus 1997). Este protocolo usa el algoritmo de integración Leapfrog Verlet, con un tiempo de integración de 2 femtosegundos, el algoritmo SHAKE para mantener fija la distancia de todos los enlaces que involucraran hidrógenos y una actualización de las listas no enlazantes cada 20 femtosegundos (parámetro INBFRQ del programa CHARMM, corresponde a la lista de átomos que se encuentran a menos de 9.5 Å de distancia de cada átomo). El sistema se calentó desde 240 K a la temperatura final de la simulación utilizando incrementos de 10 K cada 1 ps de simulación. El calentamiento se detuvo al alcanzar la temperatura deseada, la cual se mantuvo constante reasignando los vectores de velocidad de los átomos a la distribución de Boltzmann de la temperatura final de la simulación en el caso de producirse desviaciones de más de 10 K. La energía total del sistema no se mantiene constante con este protocolo, como se observa en la Figura

3A, línea azul. Modificando los parámetros de la simulación, se encontró que la única forma de lograr una buena conservación de energía fue actualizar las listas de interacciones no enlazantes en cada paso de la simulación (Fig 3A, línea roja). No obstante, al comparar la forma del desplegamiento logrado por ambos métodos se observó que la falta de conservación de energía no afecta la forma general del proceso de desplegamiento, sino que solo aumenta su velocidad (Fig. 3A, insertos superior e inferior). Adicionalmente, al comparar los resultados de los dos protocolos antes mencionados con las simulaciones de solvente explícito descritas por el grupo de Daggett, se observó una similitud cualitativa (Fig 3B). Los experimentos de Daggett muestran que el desplegamiento de CI2 procede en tres etapas (Li y Daggett 1996) 1) el extremo amino terminal se inestabiliza y hace que la parte inferior del núcleo hidrófobo se exponga al solvente; 2) el empaquetamiento del núcleo hidrófobo disminuye desestabilizando la sábana beta y 3) la hélice se despega del núcleo hidrófobo perdiéndose la estructura de la sábana beta. Las simulaciones realizadas con EEF1 concuerdan con estas observaciones (Fig 3A, 3B). Considerando esto, se decidió utilizar el protocolo no conservativo (Lazaridis y Karplus 1997) con el fin de aumentar el muestreo conformacional durante las simulaciones y con esto mejorar la descripción del proceso de desplegamiento.



**Figura 1. Protocolo no conservativo aumenta la velocidad del desplegamiento de CI2 sin cambiar la vía de desplegamiento. Panel A)** Se comparó la conservación de energía del protocolo de desplegamiento a 500 K publicado para CI2 (INBFRQ = 10, calentamiento rápido, línea azul), con el protocolo que mostró mejor conservación (INBFRQ = 1, calentamiento lento, línea roja). Los cambios estructurales experimentados por la proteína al utilizar un protocolo de desplegamiento no conservativo (estructuras en zona superior) son similares a los producidos por el uso de un protocolo conservativo (estructuras en la zona inferior), **Panel B)** Desplegamiento de CI2 en solvente explícito utilizando ENCAD (Day y cols. 2002). Los cambios estructurales reportados entre el estado nativo (N) a estados posteriores al estado de transición (TS) concuerdan cualitativamente con el desplegamiento de CI2 utilizando EEF1

## Simulaciones a 300 K y a temperaturas de desplegamiento

En la literatura se han descrito simulaciones de desplegamiento de proteínas a altas temperaturas utilizando el protocolo descrito anteriormente, verificándose un desplegamiento casi total a 500 K antes del nanosegundo de simulación. Para explorar la estabilidad de la estructura secundaria y los núcleos hidrófobos, se llevaron simulaciones de 5 ns a 300, 400, 450 y 500 K de las proteínas PaFtsZ, MjFtsZ y sus dominios aislados. En base a los resultados de estas simulaciones, se decidió realizar 20 simulaciones de 1 ns de cada uno de los sistemas, para obtener resultados estadísticamente significativos. En la Tabla 3 se muestra un resumen de las simulaciones realizadas. En total se simularon 600 ns de dinámica molecular.

**Tabla 2. Simulaciones de desplegamiento**

Proteína	Sistema	Número de simulaciones <sup>2</sup>	
		5 ns	1 ns
PaftsZ	Silvestre <sup>1</sup>	4	80
	Dominio amino	4	80
	Dominio carboxilo	4	80
MjFtsZ	Silvestre	4	80
	Dominio amino	4	80
	Dominio carboxilo	4	80
Tiempo total de simulación		120 ns	480 ns

<sup>1</sup> Se refiere a la proteína completa, sin separar los dominios

<sup>2</sup> Para cada sistema, se simularon 4 temperaturas, a saber: 300, 400, 450 y 500 K. De este modo se realizó una simulación de 5 ns y 20 simulaciones de 1 ns para cada temperatura.

## Algoritmo de análisis de la estructura secundaria

Para calcular la estructura secundaria de los residuos durante las simulaciones de desplegamiento, se utilizó el programa STRIDE (*secondary STRuctural IDentification*) (Frishman y Argos 1995), el cual utiliza funciones para cada una de las estructuras secundarias (hélices  $\alpha$ ,  $\pi$  y  $3_{10}$ ; hebras  $\beta$ , vueltas  $\beta$ , giros y lazos) que consideran tanto la formación de patrones de puente de hidrógeno característicos de cada estructura como los ángulos diedros  $\phi$  y  $\psi$  del esqueleto de la proteína. Por

ejemplo, en el caso de las hélices alfa, se considera la existencia de una hélice entre los residuos  $k$  y  $k+4$  si al menos dos residuos consecutivos cumplen con la condición:

$$E_{hb}^{k,k+4} \left( 1 + W_1^\alpha + W_2^\alpha + \frac{P_k^\alpha + P_{k+4}^\alpha}{2} \right) < T_1^\alpha \quad \text{Eq.1}$$

Donde  $E_{hb}^{k,k+4}$  es la energía del puente de hidrógeno entre los residuos  $k$  y  $k+4$ ,  $P_k^\alpha$  y  $P_{k+4}^\alpha$  corresponden a la probabilidad de encontrar en la bases de datos PDB residuos en hélices  $\alpha$  que tengan los ángulos diedros  $\phi$  y  $\psi$  correspondientes a los residuos  $k$  y  $k+4$ , mientras que  $W_1^\alpha$ ,  $W_2^\alpha$  y  $T_1^\alpha$  son constantes empíricas que se determinaron durante el proceso de optimización del algoritmo.

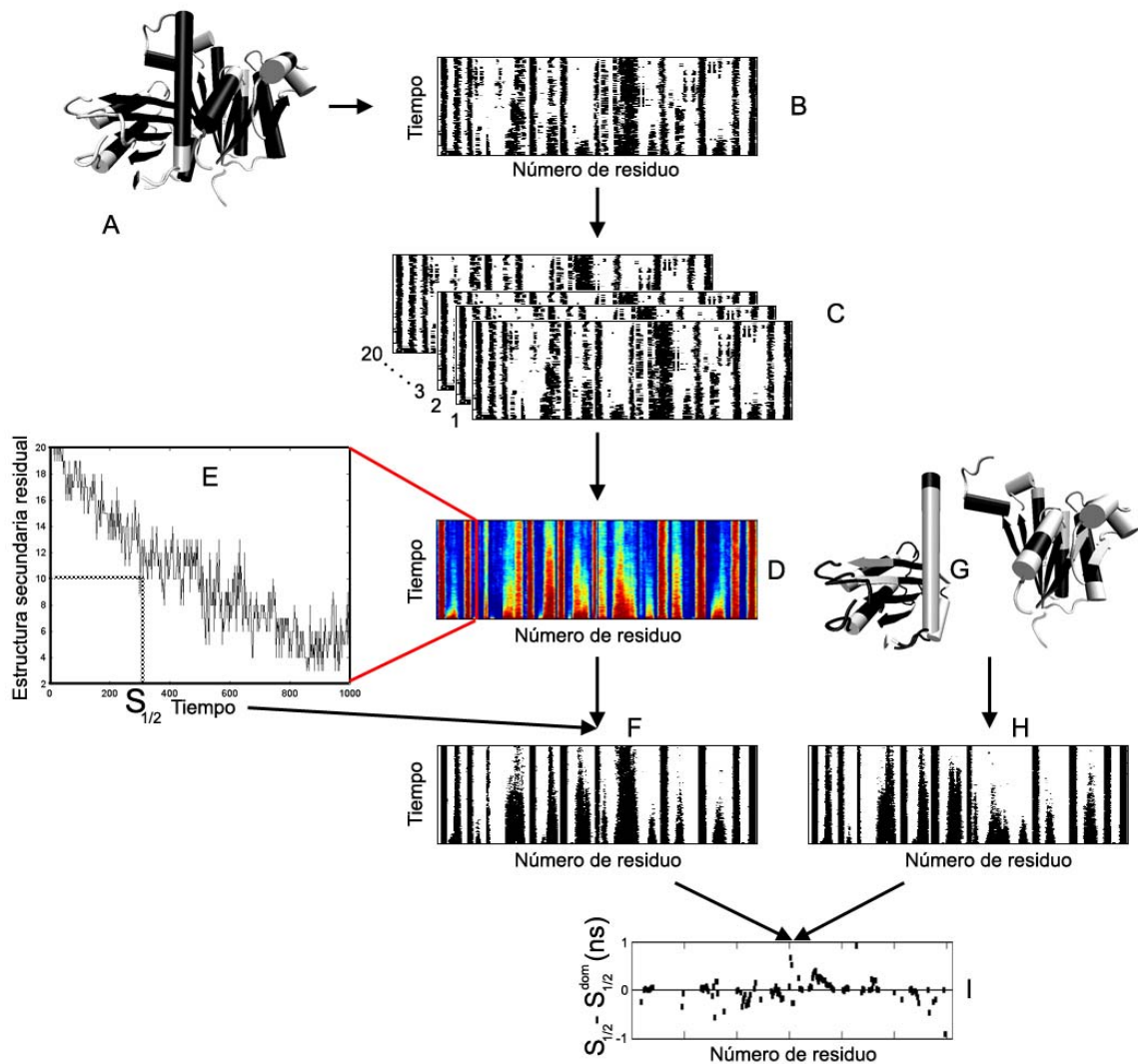
### **Análisis del cambio de estructura secundaria durante el desplegamiento**

Uno de los objetivos de este trabajo es estudiar la influencia de la interfase de interacción entre los dominios sobre su comportamiento durante el desplegamiento. Para comparar la pérdida de estructura secundaria en la proteína completa y los dominios aislados, se ideó un análisis basado en matrices de estructura secundaria.

En toda simulación, es posible observar que algunos residuos tienen estructura de lazo o bien pierden estructura secundaria (Fig. 4A, residuos de color blanco en la estructura) mientras otros residuos mantienen su estructura secundaria (Fig. 4A, residuos de color negro en la estructura). La presencia o ausencia de estructura secundaria en un residuo durante una simulación puede representarse por la matriz  $S$  de tamaño  $n \times m$  donde  $n$  es el número de estructuras que componen la trayectoria en el tiempo y  $m$  el número de residuos de la estructura ( Fig. 4B). Si  $S_{nm} = 0$ , entonces el residuo  $m$  no presenta estructura secundaria en el tiempo  $n$ , lo cual se visualiza como un espacio en blanco en la matriz; en forma análoga si  $S_{nm} = 1$ , entonces el residuo  $m$  presenta estructura secundaria en el tiempo  $n$ , lo cual se visualiza como cuadros



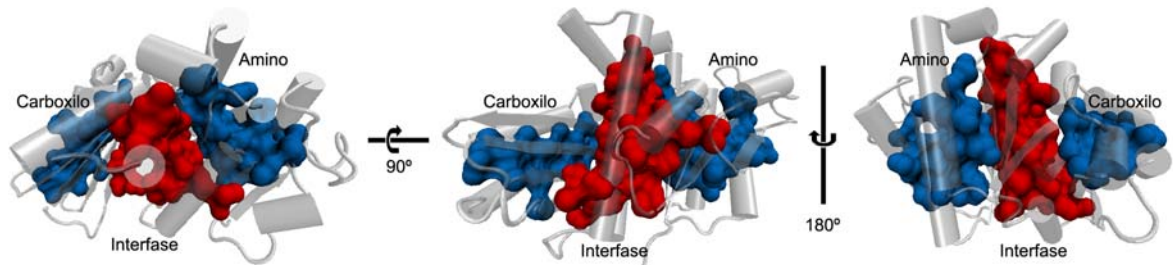
negros en la matriz. Para obtener un promedio representativo del comportamiento de cada residuo, se calculó la matriz  $S$  para las 20 simulaciones de 1 ns (Fig 4C). Luego se construyó la matriz de estructura secundaria residual  $R$  (Fig. 4D), de tamaño  $n \times m$  que corresponde a la suma de las 20 matrices  $S$ . Cada casilla de la matriz  $R$  esta coloreada de acuerdo a un gradiente que va desde el rojo al azul, lo que representa respectivamente que el residuo  $m$  presenta estructura secundaria en todas las simulaciones ( $R_{nm} = 20$ ) o en ninguna ( $R_{nm} = 0$ ). En el caso de que el residuo presente estructura secundaria en un número de simulaciones intermedio entre estos valores extremos ( $0 < R_{nm} < 20$ ), se muestran en graduaciones de amarillo, verde y celeste. Para comparar el comportamiento de la proteína silvestre con los resultados por dominios se definió un valor que depende solo del residuo  $m$  que se llamó  $S_{1/2}^m$  (Fig. 4E). Este corresponde al tiempo  $n$  en que  $R_{nm} = 10$ . Una forma de representar el valor  $S_{1/2}^m$  consiste en mostrar la matriz  $R$  con una graduación de color donde todo elemento de la matriz que presente estructura secundaria en menos de 10 simulaciones se muestra vacío, mientras que si es superior a 10 se muestra en negro (Fig. 4F). Con el valor  $S_{1/2}^m$  calculado para la proteína silvestre, es posible compararlo con las matrices  $R$  (Fig. 4H) de los dominios aislados. De estas matrices se puede derivar un valor  $S_{1/2}^{dom}$ , que al restarse a  $S_{1/2}^m$  nos entrega un valor que da cuenta de la diferencia de estabilidad de la estructura secundaria para un residuo dado en la proteína silvestre y en los dominios aislados.



**Figura 2. Protocolo para comparar el desplegamiento de las proteínas PaFtsZ y sus dominios aislados.** A partir de una simulación de dinámica molecular de una estructura cristalina (A) es posible calcular una matriz  $S$  de estructura secundaria (B). Si se llevan a cabo múltiples repeticiones de esta simulación utilizando semillas aleatorias distintas (C) es posible combinar los resultados en una matriz  $R$  de estructura secundaria (D) la cual muestra un gradiente de color desde el rojo al azul para representar la presencia o ausencia de estructura secundaria en las múltiples simulaciones. Al representar el gradiente de color de la matriz  $R$  en forma numérica para un residuo  $m$  particular (E) es posible definir el valor  $S_{1/2}^m$ , que si es calculado para cada residuo permite generar un perfil de valores  $S_{1/2}^m$  (F), que si es comparado con perfiles de valores  $S_{1/2}^m$  de matrices  $R$  obtenidas a partir de simulaciones de los dominios aislados (G,H) es posible obtener una cuantificación del cambio en estabilidad de la estructura secundaria de los residuos tras la separación de los dominios (I).

### Definición de los núcleos hidrófobos en PaFtsZ y MjFtsZ

Se definió un núcleo hidrófobo como un conjunto de residuos no expuestos al solvente cuyas cadenas laterales generan una red de contactos conexas. Se utilizó el programa NACCESS (Hubbard y Thornton 1993) para calcular la accesibilidad al solvente de los residuos de ambas estructuras. Luego, se alinearon localmente las secuencias de ambas proteínas con el programa BLAST (Altschul y cols. 1997), para establecer pares de residuos equivalentes. Un residuo se consideró “escondido” si su accesibilidad a solvente (o la de su equivalente) era menor o igual a 7%. Esto permitió que todas las comparaciones entre PaFtsZ y MjFtsZ se efectuaran entre núcleos del mismo tamaño. El segundo paso consistió en definir las redes de contactos. Por inspección visual, se pudo definir tres de ellas, a saber; la de la interfase de interacción entre los dominios, la del dominio carboxilo y la del dominio amino (Fig. 5). En el Anexo III se listan los residuos involucrados en cada uno de estos núcleos.



**Figura 3. Núcleos hidrófobos de MjFtsZ.** Se muestra la superficie de las cadenas laterales de los núcleos hidrófobos del dominio amino (azul), carboxilo (azul) e interfase (rojo). El esqueleto polipeptídico se presenta en forma esquemática como hebras y cilindros transparentes.

## Definición de contactos nativos

Una medida que logra diferenciar estados plegados de estados desplegados y semidesplegados es el número de contactos nativos,  $Q$  (Shea y Brooks III 2001). Para una estructura de proteína de  $n$  residuos,  $Q$  esta dado por:

$$Q = \sum_{i=1, j=i+2}^n \delta_{ij} e_{ij} \quad \text{Eq. 2}$$

Donde  $\delta_{ij} = 1$  si los centros de masa de los residuos  $i$  y  $j$  se encuentran a una distancia menor o igual a 6 Å y  $\delta_{ij} = 0$  en cualquier otro caso. Por otra parte,  $e_{ij} = 1$  si los residuos  $i$  y  $j$  están en contacto en la estructura de referencia que consideramos nativa, de lo contrario  $e_{ij} = 0$ . Para construir la matriz  $E$  que contiene los coeficientes  $e_{ij}$ , se procedió a calcular la matriz de frecuencia de contactos  $F$  a partir de las simulaciones a 300 K de PaFtsZ y MjFtsZ. Para cada proteína, se utilizaron estructuras extraídas cada 10 ps, dando un total de 2000 estructuras en las 20 simulaciones de 1 ns. De este modo, los coeficientes de la matriz  $F$ , denotados por  $f_{ij}$ , cumplen con la condición  $0 \leq f_{ij} \leq 2000$ . La matriz  $E$  se obtuvo a partir de la matriz  $F$  a partir de aquellos contactos presentes en al menos en un 70% de las estructuras (Sham y cols. 2002), es decir para los cuales  $f_{ij} \geq 1400$ .

## RMSD (Root Mean Square Deviation)

El valor RMSD corresponde a la distancia euclidiana entre dos vectores, normalizada por el número de dimensiones de estos. Una estructura de proteína puede considerarse un vector  $3N$  dimensional donde  $N$  corresponde al número de átomos de la estructura, de este modo, para dos conformaciones  $a$  y  $b$  de una proteína, su RMSD esta dado por:

$$RMSD(a,b) = \sqrt{\frac{1}{3N} \sum_{i=1}^n (b_{ix} - a_{ix})^2 + (b_{iy} - a_{iy})^2 + (b_{iz} - a_{iz})^2} = \sqrt{\frac{1}{3N}} \cdot D(a,b) \quad \text{Eq. 3}$$

donde  $a_{ix}$  es la coordenada  $x$  del  $i$ -ésimo átomo de la conformación  $a$  y  $b_{ix}$  es la coordenada  $x$  del  $i$ -ésimo átomo de la conformación  $b$ . Los cálculos se realizaron utilizando solo los carbonos  $\alpha$  de los residuos. Los cálculos de RMSD se realizaron con del programa VMD (*Visual Molecular Dynamics*) (Humphrey y cols. 1996).

### **RMSF (Root Mean Square Fluctuation)**

RMSF es una medida de flexibilidad local de una estructura. A partir de una simulación de dinámica molecular, el RMSF del átomo  $i$  da cuenta del grado de fluctuación de las coordenadas de dicho átomo respecto a su promedio temporal:

$$RMSF(i) = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \langle x_i \rangle)^2 + (y_i - \langle y_i \rangle)^2 + (z_i - \langle z_i \rangle)^2} \quad \text{Eq. 4}$$

Donde los términos entre paréntesis oblicuos corresponden a los promedios temporales. Los cálculos fueron realizados utilizando VMD.

### **SASA (Solvent Accesible Surface Area)**

La superficie accesible al solvente de una proteína es una propiedad para la cual no existe una fórmula analítica, debido a que el solvente está compuesto por elementos discretos. Esto mismo hace que la superficie molecular no pueda ser descrita como la suma de las superficies de los átomos del soluto; la rugosidad de esta hace que ciertas cavidades no puedan ser llenadas por el solvente, de este modo la superficie accesible al solvente se espera sea menor a la superficie molecular definida por los radios de Van der Waals de los átomos de la proteína.

La superficie accesible al solvente se calcula numéricamente utilizando un algoritmo que hace rodar una esfera de radio 1,4Å sobre la superficie molecular, el cual corresponde al radio de giro de la molécula de agua. Los cálculos se realizaron dentro del programa VMD.

En el anexo IV se listan los programas generados durante esta memoria para realizar los cálculos anteriormente descritos, y los cuales se encuentran en el CD que acompaña esta memoria.

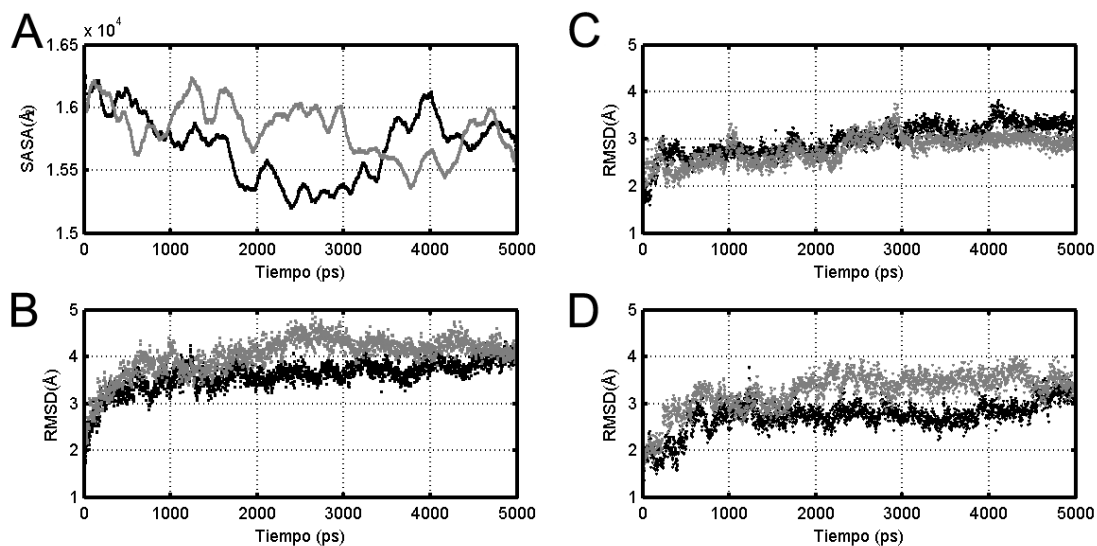
## RESULTADOS

### Estabilidad de las estructuras en simulaciones a 300 K

Para determinar el comportamiento dinámico de PaFtsZ y MjFtsZ en condiciones nativas se hicieron simulaciones de 5 ns a 300 K. Es de esperar, en una simulación de proteína correctamente realizada, que la proteína no se despliegue o presente distorsiones muy pronunciadas, es decir, que mantenga las características de la estructura cristalina. Para estudiar la deriva estructural de las simulaciones a 300 K, se utilizaron descriptores moleculares como RMSD, RMSF, SASA y variación estructura secundaria en el tiempo.

### RMSD y accesibilidad al solvente

Las accesibilidades a solvente de las estructuras cristalinas de PaFtsZ y MjFtsZ son similares ( $16104 \text{ \AA}^2$  y  $16207 \text{ \AA}^2$ , respectivamente) donde las diferentes características son atribuibles a las diferencias de secuencia de ambas proteínas. Durante el primer nanosegundo de simulación, ambas estructuras experimentan una compactación que se hace visible como una disminución en la accesibilidad a solvente. La accesibilidad a solvente de MjFtsZ sigue disminuyendo hasta alcanzar un mínimo a los 2,5 ns, para luego aumentar hasta los 4 ns de simulación, mientras que PaFtsZ incrementa su accesibilidad después del primer nanosegundo, para disminuir después y alcanzar un mínimo a los cuatro nanosegundos de simulación. Después de los cuatro nanosegundos, PaFtsZ y MjFtsZ alcanzan accesibilidades similares, de  $15300 \text{ \AA}^2$  y  $15350 \text{ \AA}^2$ , respectivamente. Estas disminuciones en accesibilidad corresponden a un cambio de 5% de las accesibilidades iniciales. Los pequeños cambios en accesibilidad muestran que las proteínas se mantienen plegadas durante las simulaciones.



**Figura 4. Estabilidad de las simulaciones a 300 K de MjFtsZ (negro) y PaFtsZ (gris). Panel A.** Superficie accesible a solvente, los datos se suavizaron usando un promedio móvil con una ventana de 200 ps. **Panel B.** RMSD con respecto a la estructura de los cristales. **Panel C.** RMSD del dominio amino. **Panel D.** RMSD del dominio carboxilo.

Al analizar el valor RMSD a partir de las simulaciones de ambas proteínas, se observa en la Figura 6B un aumento durante el primer nanosegundo de simulación, seguido por una abrupta disminución de la pendiente en los siguientes 4 ns hasta alcanzar un RMSD de 4Å, que es mayor a lo esperado para una simulación de una proteína globular a 300 K. Para dilucidar los factores que afectan la movilidad de ambas estructuras, se determinó el RMSD de cada uno de los dominios por separado (Fig. 6C y 6D). Se observa que el RMSD de los dominios separados alcanza los 3Å después de 5 ns de simulación, sugiriendo que los cambios en la accesibilidad al solvente y el alto RMSD pueden explicarse en base a la fluctuación de cuerpos rígidos de los dos dominios. Los dominios amino de las proteínas muestran un comportamiento similar hasta los 3 ns, después de lo cual el dominio amino de MjFtsZ muestra un aumento en RMSD de 0,5 Å sobre el dominio amino de PaFtsZ, alcanzando un RMSD a los 5 ns de 3 y 3.5 Å, respectivamente. Las mayores diferencias se muestran en el dominio carboxilo, donde PaFtsZ muestra un desplegamiento más rápido durante el primer nanosegundo de simulación. El RMSD de ambos dominios se estabiliza en 3 Å, y a los 1,7 ns el dominio carboxilo de PaFtsZ muestra un aumento de RMSD alcanzando



valores que fluctúan entre 3,5 y 4 Å. Hacia el final de la simulación, el RMSD de MjFtsZ alcanza al de PaFtsZ, sugiriendo que aunque los dominios carboxilo de ambas proteínas pueden mostrar fluctuaciones de la misma amplitud, el dominio carboxilo de PaFtsZ es más dinámico. Este comportamiento puede deberse o bien a la inestabilidad del dominio carboxilo de PaFtsZ o a efectos de contactos cristalográficos (ver discusión).

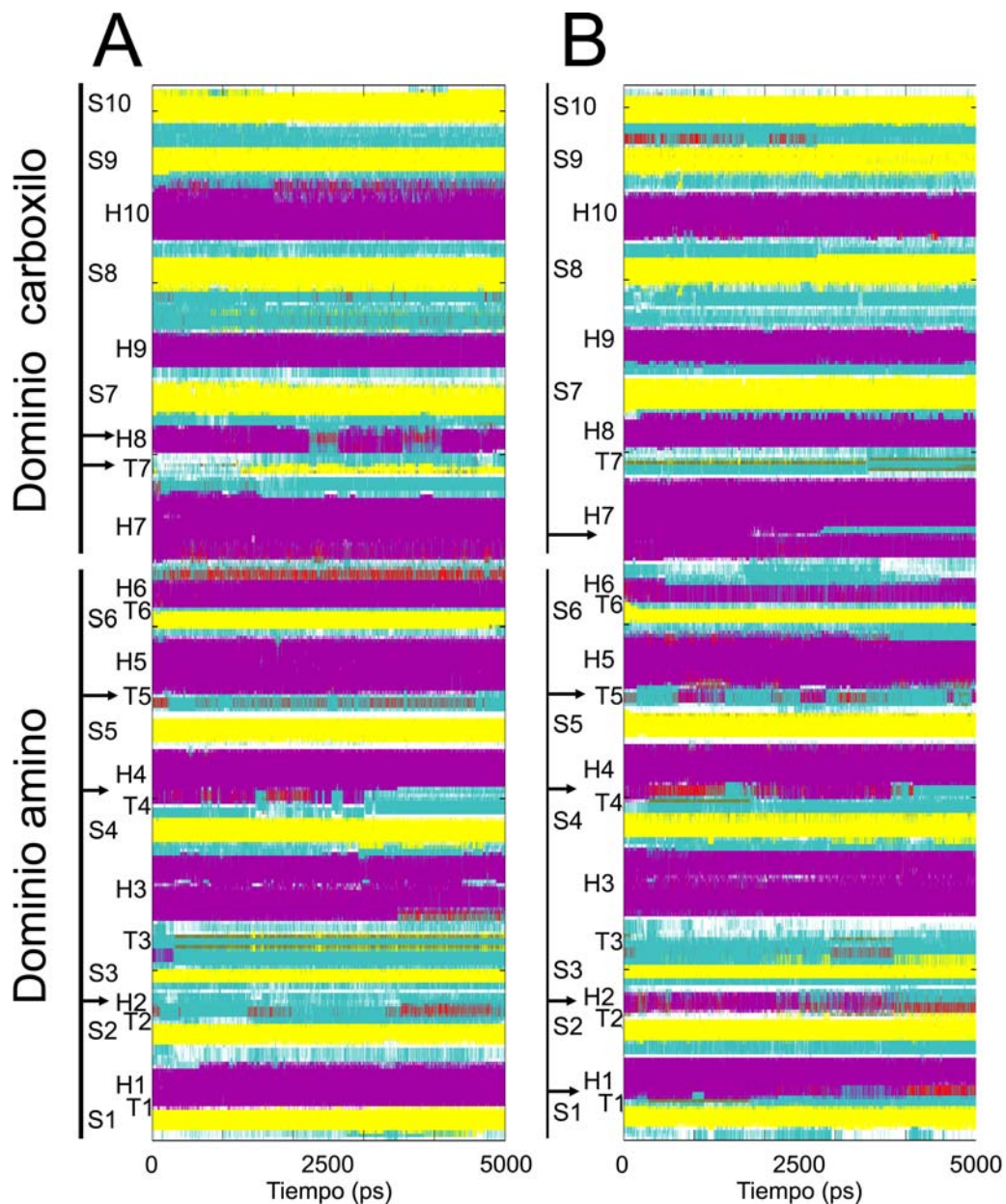
### **Cambio de estructura secundaria durante la simulación**

Los análisis de RMSD y SASA no son capaces de reconocer si los cambios estructurales se deben a cambios en estructura terciaria o bien estructura secundaria. Para evaluar los cambios en la estructura secundaria, se hizo un análisis de las fluctuaciones de las estructuras secundarias en el tiempo.

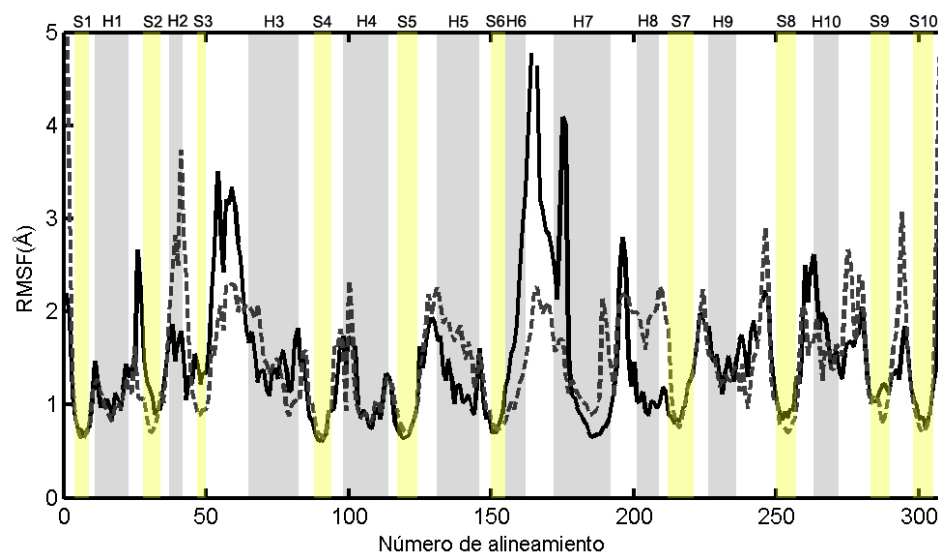
La proteína PaFtsZ (Fig 7A) mantiene la integridad de las sabanas beta y hélices alfa durante la dinámica. Las mayores diferencias se evidencian en los lazos asociados a la unión a ligando (T1 – T6, ver Fig. 2) y las zonas de estructura secundaria aledañas a éstos. La zona carboxilo de la hélice H7, el lazo T7 y la hélice H8 presentan cambios en estructura secundaria (el lazo T7 se asocia al dominio carboxilo tomando estructura beta, mostrada en amarillo, y H8 pierde estructura secundaria momentáneamente). Estos cambios en estructura secundaria no corresponden a un desplegamiento local de estas zonas, sino a un cambio de conformación (datos no mostrados). Fuera de estas diferencias, la estructura secundaria de PaFtsZ se mantiene estable durante la simulación. Las fluctuaciones de de MjFtsZ son similares a las de PaFtsZ (Fig. 7B).

Se observan las mismas fluctuaciones en las zonas asociadas a los lazos del sitio activo (T1 – T6) con pérdida de estructura en la hélice H2, y no se observan cambios en T7, H7 y H8, debido a que en MjFtsZ el movimiento entre los dominios no es tan pronunciado como en el caso de PaFtsZ. Una característica distinta entre ambas proteínas es la pérdida de estructura secundaria en la zona amino terminal de la hélice H7. Al analizar la dinámica es posible evidenciar que esta zona (la cual interactúa con la base nitrogenada del ligando) experimenta un quiebre debido a la fuerte atracción de

residuos de lisina por residuos del sitio activo (imágenes no mostradas), lo cual puede deberse a la falta de solvatación explícita.



**Figura 5. Comportamiento temporal de la estructura secundaria de simulaciones a 300 K. Panel A. PaFtsZ, Panel B. MjFtsZ.** Los colores representan distintos elementos de estructura secundaria, calculados con STRIDE: morado: hélices  $\alpha$ ; amarillo: hebras  $\beta$ ; verde: vueltas  $\beta$ ; rojo: hélices  $\pi$ .



**Figura 6. Fluctuaciones de carbonos  $\alpha$  de simulaciones a 300 K.** Línea continua MjFtsZ; línea segmentada PaFtsZ. El alineamiento de las estructuras se realizó con el algoritmo STAMP en VMD, de este modo zonas que aparecen con espacios sin valor de RMSF son inserciones en el alineamiento estructural. Las zonas achuradas corresponden a las hélices  $\alpha$  (gris) y hebras  $\beta$  (amarillo) de la estructura cristalina de MjFtsZ.

### Cálculo de fluctuaciones de los carbonos $\alpha$

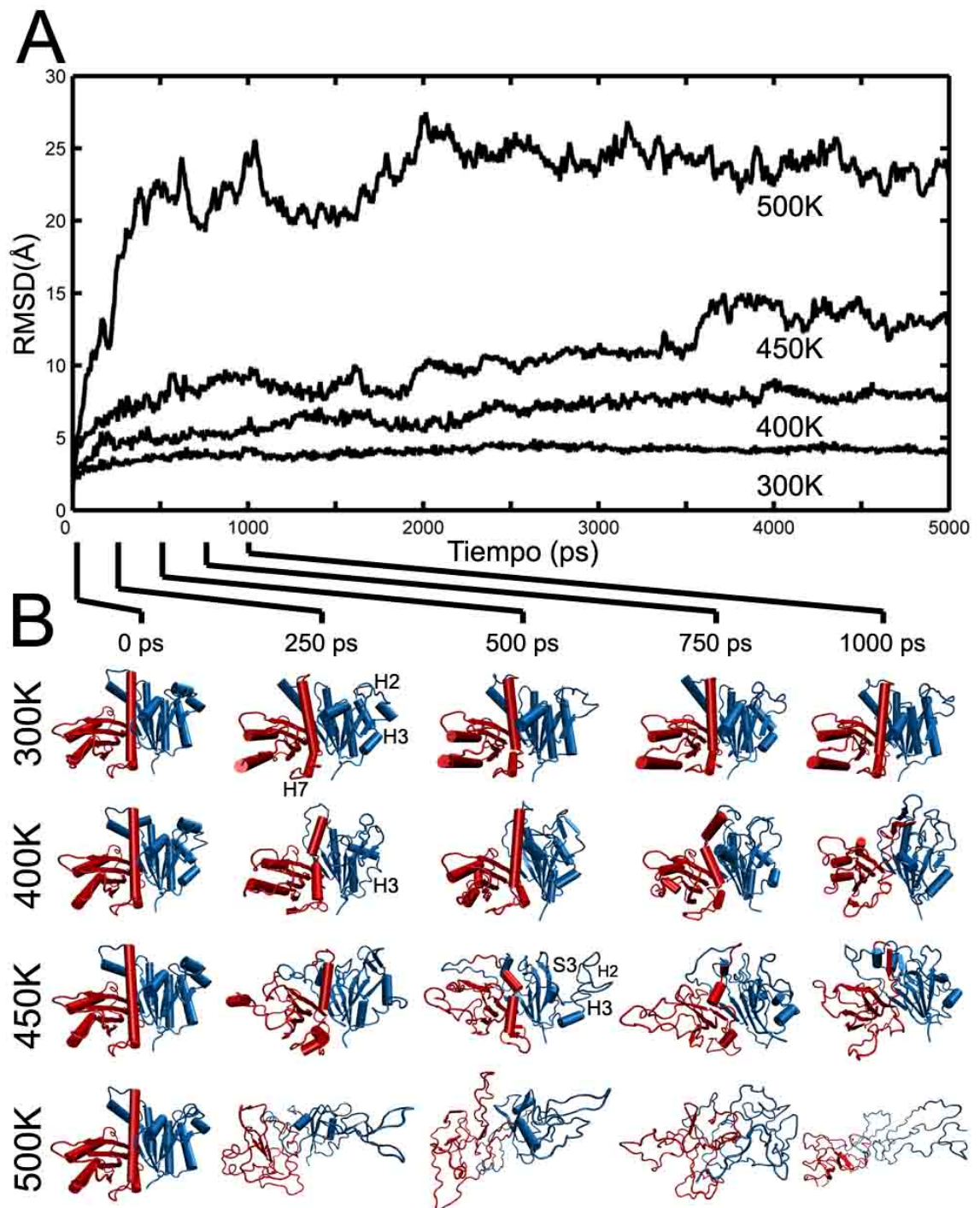
Luego de verificar la mantención del plegamiento de la proteína y la estructura secundaria a 300 K, es conveniente hacer un análisis cuantitativo de la fluctuación estructural a nivel de residuo, para comparar en forma detallada las diferencias entre PaFtsZ y MjFtsZ. Una medida adecuada es el RMSF (eq. 4, ver Materiales y Métodos), que indica la fluctuación de la estructura con respecto a la conformación promedio en la simulación. Al superponer los perfiles de RMSF de ambas proteínas (Fig. 8) las diferencias más notorias se encuentran en las zonas de los lazos, las cuales presentan un RMSF mayor al tener menos restricciones en su movimiento. En primer lugar, las zonas de mayor fluctuación en el dominio amino son las zonas aledañas a los lazos T2 y T3. La hélice H2 presenta un alto RMSF, lo cual es consistente con la desaparición de estructura secundaria en esta zona en ambas proteínas.

### **Simulaciones de desplegamiento térmico: aspectos cualitativos**

Para estudiar el proceso de desplegamiento térmico de las proteínas FtsZ, éste se simuló a tres temperaturas donde se favorece el estado no desplegado: 400, 450 y 500 K. Se hará una descripción cualitativa de los eventos observados en estas simulaciones.

### **Simulaciones del desplegamiento térmico de PaFtsZ**

Las simulaciones a altas temperaturas producen el desplegamiento de PaFtsZ, donde el RMSD aumenta con respecto al observado a 300 K, y este aumento es proporcional a la temperatura, como se observa en la Figura 9A. La curva de desplegamiento a 400 K, a diferencia de la 300 K que alcanza una plataforma, muestra un aumento del RMSD durante toda la simulación, lo cual significa que a esta temperatura se requiere más de 5 ns para llegar a un desplegamiento completo, que puede interpretarse como una desnaturalización lenta de la proteína. Lo mismo ocurre a 450 K, donde la desnaturalización es más rápida y no se alcanza el desplegamiento en los 5 ns de simulación. A 500 K la desnaturalización de la proteína alcanza rápidamente un RMSD máximo para luego oscilar entre 20 y 25 Å, sugiriendo que se ha alcanzado el estado desplegado.



**Figura 7. Efecto de la temperatura sobre la estructura de PaFtsZ. Panel A.** Variación en el tiempo de RMSD de carbonos a obtenidos a partir de simulaciones de dinámica molecular realizadas a 300, 400, 450 y 500 K. **Panel B.** Imágenes de la estructura de la proteína PaFtsZ a intervalos regulares durante el primer nanosegundo de simulación. Se anotan los nombres de las estructuras secundarias cuyo comportamiento es descrito en el texto.

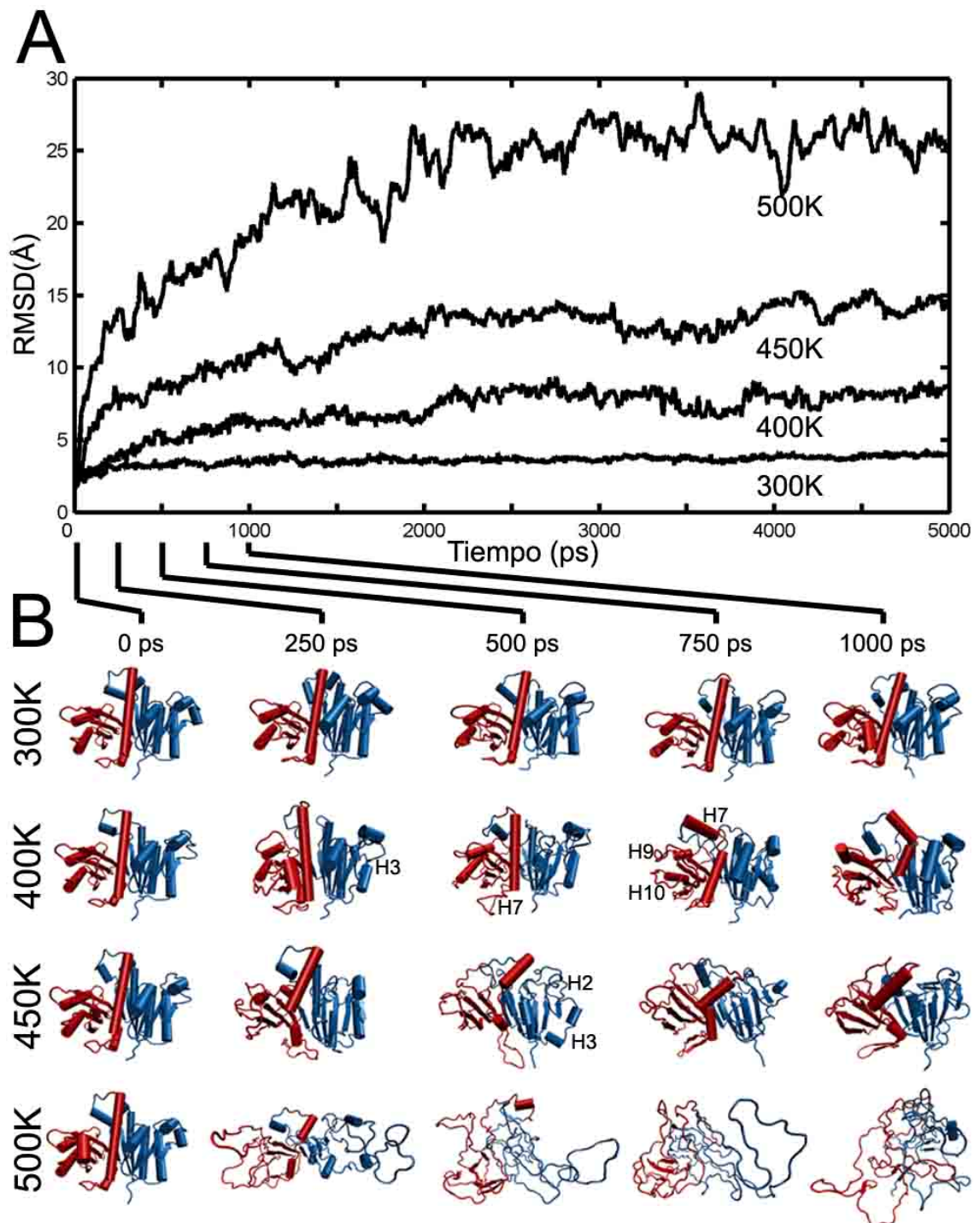
Al analizar las estructuras que se producen en el primer nanosegundo de simulación a 300 K (Fig. 9B), se observan los cambios ya enunciados en el análisis de la sección anterior. En primer lugar, a 250 ps es evidente el desplegamiento de la hélice H2, la cual toma la forma de lazo. Por otra parte, se observa que el extremo carboxilo de la hélice H7 comienza a perder estructura, y que el dominio carboxilo experimenta un movimiento de sus hélices en virtud del cambio de orientación del dominio. Estos movimientos de las hélices H9 y H10 pueden explicar los elevados RMSD descritos en la sección anterior. Adicionalmente, es evidente que la hélice H3, que presenta un quiebre en su mitad, presenta movimientos entre sus dos mitades que hacen perder estructura helicoidal al residuo D85 en PaFtsZ, que se encuentra en la zona del quiebre. A 400 K (Fig. 9B) las zonas que presentan mayor flexibilidad a 300 K son las primeras en desplegarse. La hélice H2 se encuentra completamente desplegada, así como parte de H3 pierde su estructura helicoidal pero se mantiene plegada (250 ps). Asimismo la hélice H7 presenta zonas donde la estructura se pierde, generando quiebres (250 - 750 ps) para finalmente perder toda estructura (1000 ps). Las hélices externas de ambos dominios presentan alto grado de desplegamiento pero aún es posible observar zonas donde esta se mantiene. A 450 K se observa nuevamente el quiebre de la hélice H7 y su posterior pérdida de estructura (250 - 1000 ps). Las hélices de ambos dominios se despliegan (250 ps) y la zona que comprende las estructuras H2, H3 el lazo T3 y S3 se despliegan alejándose del núcleo hidrófobo (500 ps). La proteína se vuelve una gran sabana plisada formada por la unión de los dos dominios (1000 ps) donde solo ciertas hélices del dominio amino muestran estructura secundaria residual. A 500 K la proteína se despliega completamente (250 ps), desapareciendo la estructura secundaria, tomando un aspecto de lazo azaroso.

### **Simulaciones del desplegamiento térmico de MjFtsZ**

Al igual que PaFtsZ, MjFtsZ se despliega parcialmente a 400 y 450 K y completamente a 500 K (Fig. 10A). Las simulaciones donde se alcanza un estado denaturado no llegan a un equilibrio, mientras que la simulación a 500 K alcanza un equilibrio que fluctúa entre 20 y 30 Å después de los 3 ns. Al analizar las estructuras que se muestran en la Figura 10B, se observa que en la simulación a 300K la

estructura experimental pocos cambios, salvo una notoria inclinación de la hélice H7, lo cual puede considerarse como un movimiento de desplazamiento entre ambos dominios, pero no tan pronunciado como en el caso de PaFtsZ. Se observa que la hélice H2 mantiene su estructura durante el primer ns de simulación y que a diferencia de PaFtsZ, la zona del lazo T3 y la hélice H3 no muestra fluctuación. A 400 K el desplegamiento procede en forma similar al de PaFtsZ. La hélice H2 no pierde estructura, pero sí lo hace parcialmente la hélice H3 (250 - 1000ps). La hélice H7 muestra quiebres y pérdida de estructura cerca del lazo T7 (500 - 1000 ps) y las hélices del dominio carboxilo pierden estructura (250 - 1000 ps). En el estado semidesplegado a 450 K la pérdida de estructura helicoidal genera un estado donde solo se mantiene la estructura de sabana  $\beta$ , al igual que en PaFtsZ. A 500 K la proteína se despliega, llegando a un estado desordenado, aunque es evidente la existencia de mayor estructura residual que en el caso PaFtsZ.

Al considerar el proceso de desplegamiento y plegamiento reversible, se infiere de las simulaciones de PaFtsZ y MjFtsZ que la vía de plegamiento de las proteínas se conserva, con diferencias solo locales. Es evidente que el dominio amino de MjFtsZ es más estable que el de PaFtsZ, donde la zona comprendida entre el lazo T3 y la hebra S3 es la que presenta las mayores diferencias en estabilidad. Adicionalmente, los dominios carboxilos de ambas proteínas son menos estables que los dominios amino adyacentes.



**Figura 8. Efecto de la temperatura sobre la estructura de PaFtsZ. Panel A.** Variación en el tiempo de RMSD de carbonos  $\alpha$  obtenidos a partir de simulaciones de dinámica molecular realizadas a 300, 400, 450 y 500 K. **Panel B.** Imágenes de la estructura de la proteína PaFtsZ obtenidos a diferentes tiempos durante el primer nanosegundo de simulación. Se anotan los nombres de las estructuras secundarias cuyo comportamiento es descrito en el texto.



## **Simulaciones de desplegamiento: aspectos cuantitativos**

En la sección anterior se analizó el proceso de desplegamiento térmico de PaFtsZ y MjFtsZ a distintas temperaturas en forma cualitativa. Esta sección muestra los resultados de 20 simulaciones de 1 ns de PaFtsZ, MjFtsZ y sus dominios aislados con el fin de cuantificar de manera estadística la ocurrencia de las observaciones antes descritas.

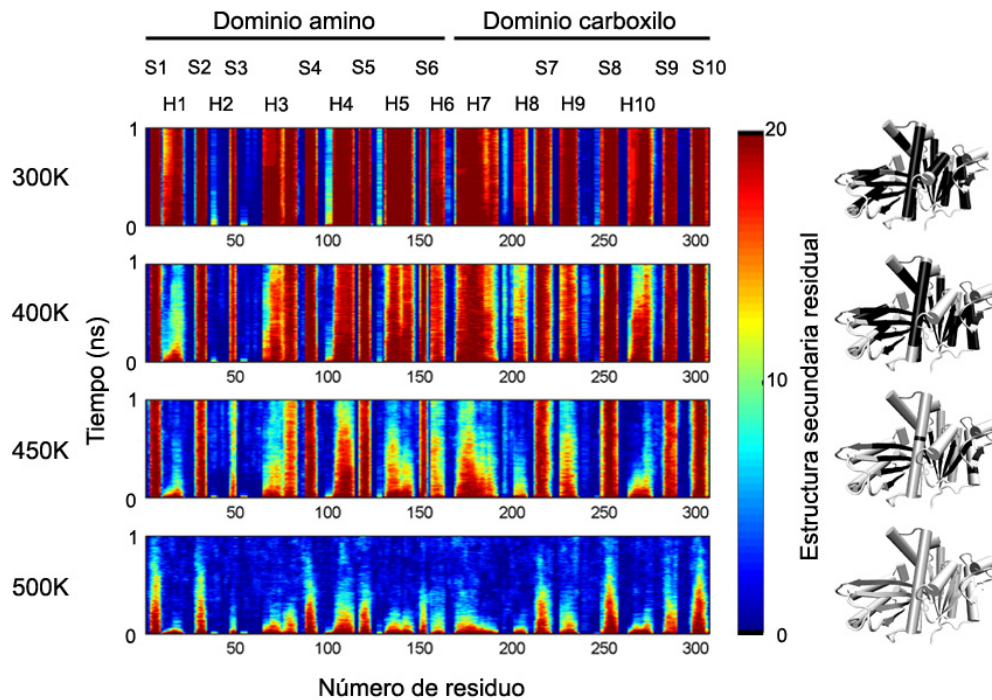
El desplegamiento con un agente denaturante de la estructura secundaria de EcFtsZ y MjFtsZ seguido por dicroísmo circular a 220 nm muestra una típica conducta cooperativa (Andreu y cols. 2002). De estos resultados experimentales se obtiene un promedio de la estructura de las moléculas en solución, así esta espectroscopia no muestra que ocurre en la estructura de una molécula individual durante la desnaturación. Para determinar los cambios estructurales de una molécula promedio durante el desplegamiento, se hizo un análisis cuantitativo detallado con las matrices  $R$  de estructura secundaria (ver Materiales y Métodos).

### **Comportamiento de los residuos de aminoácidos durante el desplegamiento de las estructuras secundarias: análisis de matrices $R$ .**

Las matrices  $R$  son matrices de frecuencia de estructura secundaria, que resumen los resultados de muchas simulaciones. Para PaFtsZ (Fig. 11) la matriz  $R$  a 300 K muestra que las zonas cercanas a los lazos T1 – T6, especialmente en las hélices H1, H3, H4 y H5 pierden estructura secundaria; la hélice H2, presente al inicio de la simulación, se despliega rápidamente, la zona carboxilo de la hélice H7 se inestabiliza, y el lazo T7 adquiere estructura secundaria (hebra beta) formando puentes de hidrógeno con la hebra S9 del dominio carboxilo. A 400 y 450 K es posible comparar la estabilidad relativa de las hélices, la cual muestra la siguiente jerarquía:

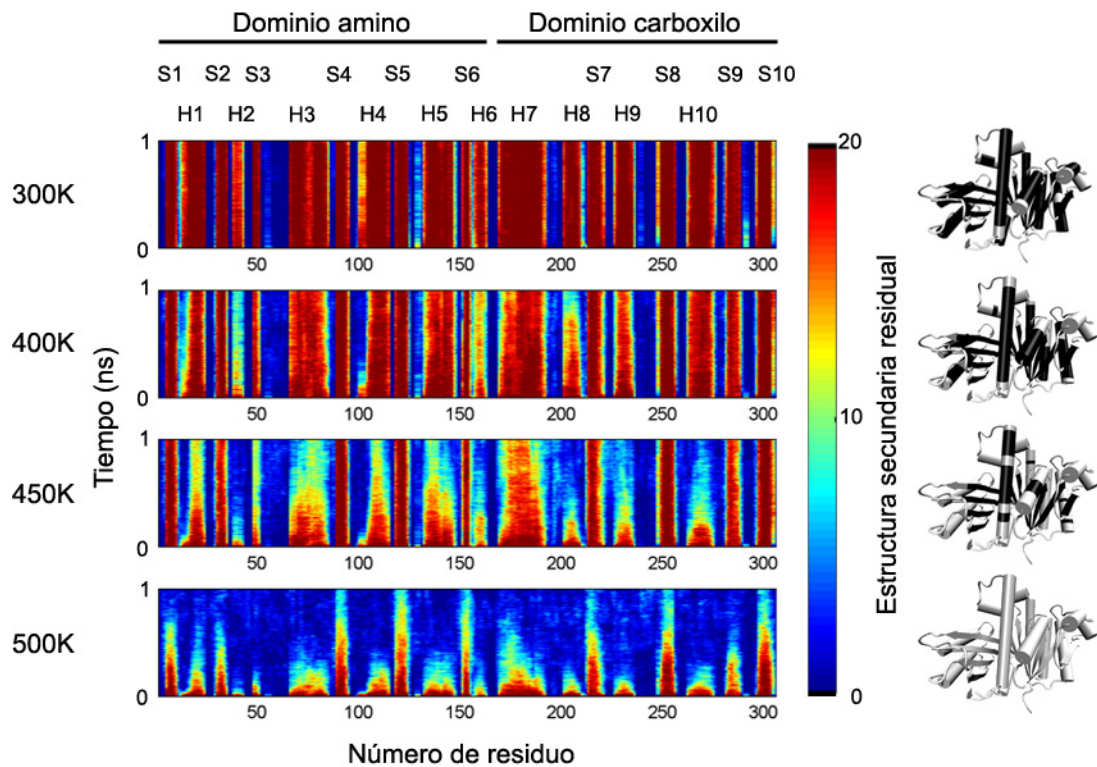
$$H2 \gg H1 > H10 \geq H8 \gg H3(T3) \sim H5 \sim H6 \sim H7 \sim H9 > H3 > H4$$

H3 representa un caso especial, pues la primera sección es muy inestable, posiblemente por encontrarse relacionada con el lazo T3, uno de los más móviles del dominio amino según los análisis de RMSF (Fig. 8). Es notorio que la hebra S3 es la única que se despliega a una temperatura menor a 500 K. De este modo el fragmento H2-S3-T3-H3 es altamente inestable y presenta el desplegamiento más rápido. A 500 K el desplegamiento total ocurre antes de los 500 ps. Las hélices de toda la estructura muestran el desplegamiento más rápido, desapareciendo antes de los 200 ps, mientras que las sabanas beta demoran entre 100 y 500 ps en desplegarse. Aunque en estas simulaciones no se observó la presencia de estructuras donde las sabanas beta de ambos dominios se unen como en el caso de la simulación a 450 K, la diferencia en



**Figura 11. Matrices  $R$  de estructura secundaria para las simulaciones del desplegamiento de PaFtsZ a 300, 400, 450 y 500 K.** La presencia de estructura secundaria en las 20 simulaciones de 1 ns realizadas a cada temperatura se muestra con un gradiente de color que va de cero (azul, ausencia de estructura en todas las simulaciones) a rojo (presencia de estructura secundaria en todas las simulaciones). A la derecha de cada uno de los gráficos se muestran en negro los residuos que mantienen su estructura secundaria en más de 10 simulaciones.

estabilidad de los elementos de estructura secundaria sugiere que el desplegamiento mantiene la forma observada a temperaturas inferiores, donde primero ocurre la desestabilización de las hélices alfa y luego la de sus sabanas beta.



**Figura 12. Matrices  $R$  de la estructura secundaria para simulaciones del desplegamiento de MjFtsZ a 300, 400, 450 y 500 K.**

Al comparar la matriz  $R$  de MjFtsZ (Fig. 12) con la de PaFtsZ se observa que las zonas cercanas a los lazos T1 – T3 presentan una pérdida de estructura secundaria y que la zona H2-S3-T3-H3 es más estable. Aunque S3 es la hebra beta que primero se despliega, la hélice H3 muestra un comportamiento homogéneo: se despliega completamente sin diferencias evidentes entre sus dos secciones, contrario a lo que ocurre en PaFtsZ. Como se observó en el análisis de RMSF (Fig.8) el lazo T3 en MjFtsZ tiene la mayor fluctuación en la estructura del dominio amino, y a su vez la fluctuación de su hélice H3 es menor que la de PaFtsZ. Esta diferencia puede

explicarse en términos de un confinamiento de la fluctuación del lazo T3 en MjFtsZ, mientras que la inestabilidad del lazo T3 parece distribuirse hacia H3 y S3 en PaFtsZ. Adicionalmente la hélice H1, la cual es la segunda hélice más inestable en PaFtsZ, es más estable en MjFtsZ. Como esta hélice interactúa con el fragmento H2-S3-T3-H3, su mayor estabilidad también puede influenciar la mayor estabilidad de éste. La mayor estabilidad de esta hélice puede deberse a su estabilidad intrínseca o bien a la estabilización por su interacción con la hélice H7, que muestra una mayor estabilidad que en el caso de PaFtsZ (Figura 11 y 12). Finalmente, la única hélice que tiene una menor estabilidad que la proteína mesófila es la hélice H6, posiblemente debido al movimiento entre los dos dominios. En conjunto, la estabilidad de las hélices de MjFtsZ tiene la siguiente jerarquía:

$$H2 \gg H6 \sim H8 > H9 \sim H10 \gg H5 \sim H3 \sim H4 \gg H1 > H7$$

Es evidente que las hélices del dominio carboxilo son más inestables que las del dominio amino, y que la mayor estabilidad de H7 parece estabilizar el fragmento H1-S2-H2-S3-T3-H3 con respecto a PaFtsZ. A 500 K, se observa que las sábanas beta poseen una estabilidad mayor que las hélices alfa, a diferencia de PaFtsZ, donde ambos elementos de estructura secundaria se despliegan en el mismo intervalo de tiempo. En conclusión, los resultados cuantitativos de las matrices son concordantes con los resultados cualitativos de la sección anterior, mostrando que los cambios observados corresponden a tendencias consistentes en las simulaciones.

### **Influencia de la interfase en la estabilidad de los dominios de FtsZ: valores $S_{1/2}^m$**

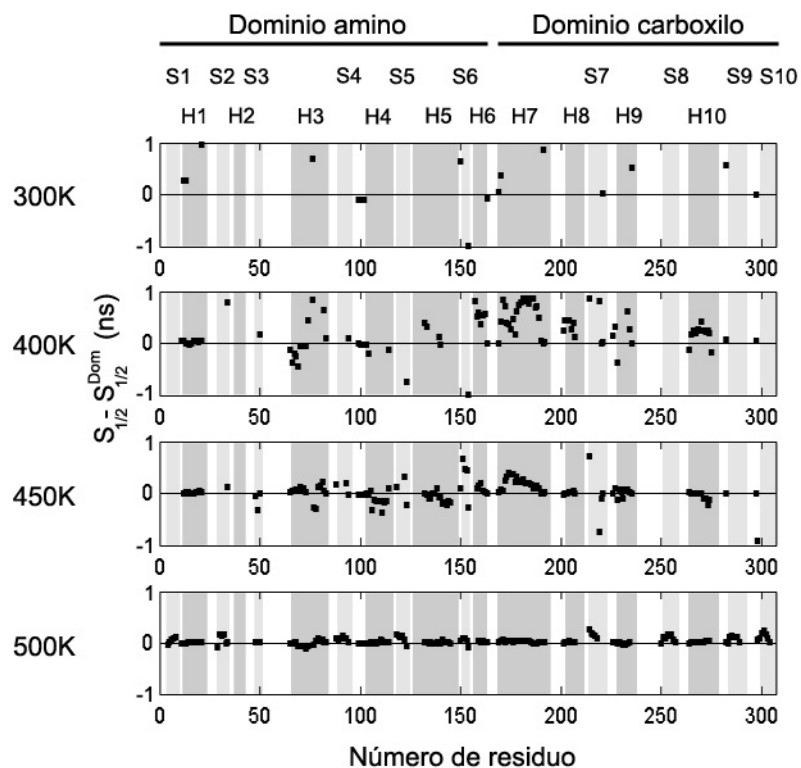
En la sección anterior se analizó la estabilidad relativa de las estructuras secundarias de PaFtsZ y MjFtsZ por medio de matrices  $R$ . Para evaluar la influencia de la interfase de interacción en la estabilidad de los dominios, se construyeron matrices  $R$  a partir de simulaciones de los dominios aislados. El análisis de estas matrices muestra cambios en la estabilidad de los dominios, sin embargo, es difícil establecer diferencias cuantitativas que facilite el análisis. Para superar este inconveniente, se definieron los

valores  $S_{1/2}^m$  y  $S_{1/2}^{dom}$  para cada residuo (Ver Materiales y Métodos y Fig. 4). En las siguientes secciones, se calcula la diferencia  $S_{1/2}^m - S_{1/2}^{dom}$  para las matrices  $R$ .

### **Estabilidad de la estructura secundaria de los dominios aislados de PaFtsZ**

En las simulaciones en condiciones nativas (300 K), se observa que la estabilidad de la estructura secundaria se conserva (Fig. 13). La mayoría de los residuos que presentan cambios se desestabilizan, principalmente en la cercanía de lazos, los cuales presentan una mayor movilidad. El único residuo que se desestabiliza dentro de un elemento de estructura secundaria es el residuo central de la hélice H3, sugiriendo que esta zona aumenta su fluctuación por sobre la fluctuación presentada en la proteína nativa. Las simulaciones a temperaturas desnaturantes ( $T \geq 400$  K) muestran que la zona más afectada por el desplegamiento es la hélice H7, la cual ya a 400 K se encuentra completamente desplegada. Las hélices del dominio carboxilo se ven más afectadas que las del dominio amino, siendo H3 la única excepción, mostrando zonas de desestabilización y zonas de estabilización. Esta elevada movilidad de H3 es consistente con su rol en la unión del ligando, donde se piensa puede actuar como un “switch” conformacional. El aumento de temperatura hace que un mayor número de residuos muestren diferencias en estabilidad (aumento en el número de puntos) sin embargo también hacen que las diferencias en estabilidad sean más pequeñas (puntos cercanos a cero), pues a altas temperaturas tanto la proteína nativa como los dominios aislados son más estables en el estado plegado que en el desplegado.

A 450 K todas las hélices se ven afectadas, siendo nuevamente la hélice H7 la que presenta los mayores cambios. Las diferencias en estabilidad, sin embargo, son más pequeñas que a 400 K, mostrándose más casos donde las hélices muestran una estabilización en ciertas zonas, como es el caso de H3, H4 y H5, H9 y H10. A 500 K todas las estructuras se ven afectadas, con valores de  $S_{1/2}^m - S_{1/2}^{dom}$  de menos de 100 ps.



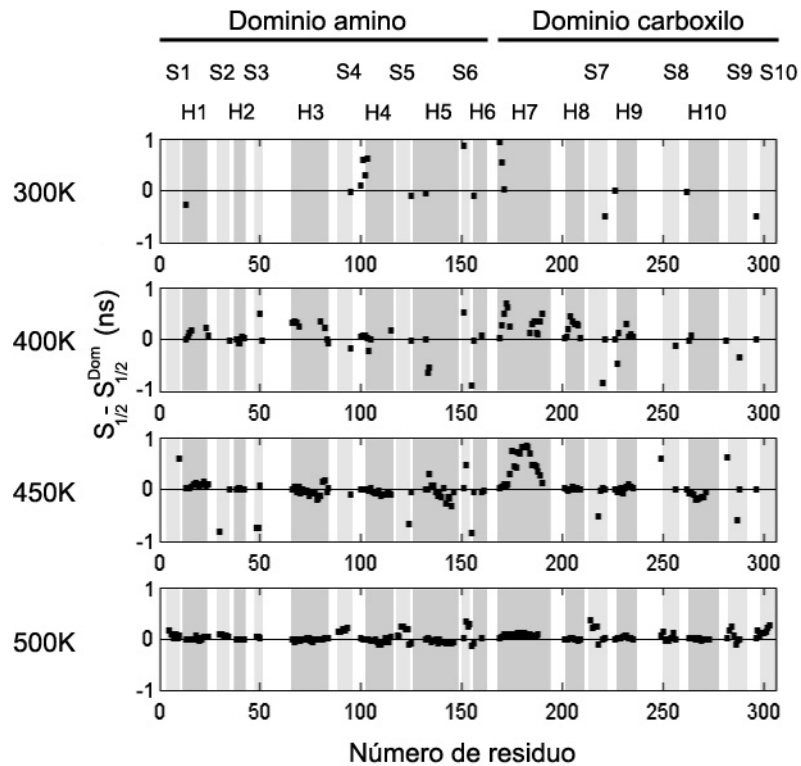
**Figura 13. Comparación del valor de estructura secundaria medio para PaFtsZ silvestre ( $S_{1/2}^m$ ) y el de sus dominios aislados ( $S_{1/2}^{dom}$ ). Un valor de  $S_{1/2}^m - S_{1/2}^{dom}$  positivo indica que la separación de los dominios produce una desestabilización de la estructura secundaria del residuo. Las zonas achuradas en color gris oscuro corresponden a alfa hélices, mientras que las achuradas en gris claro corresponden a hebras beta. Solo se muestran los valores para los residuos que mostraron diferencias de estabilidad.**

### **Estabilidad de la estructura secundaria de los dominios aislados de MjFtsZ**

Al igual que PaFtsZ, a 300 K (Fig. 14) solo algunos residuos cercanos a los lazos muestran cambios en estabilidad. A 400 K un mayor número de residuos se desestabiliza, y el grado de desestabilización es menor que en el caso de PaFtsZ. H7 sigue siendo la hélice más inestable, con desestabilizaciones en sus dos extremos. De las demás hélices del dominio carboxilo, solo H8 presenta una desestabilización similar a la de la hélice H8 de PaFtsZ, mientras que la separación de los dominios afecta menos a H9 y no afecta a H10. En el dominio amino, las hélices H3 y H1 muestran una notoria desestabilización en sus extremos, siendo más notoria para el caso de H3. La

hélice H2 muestra una desestabilización, sin embargo las diferencias son muy pequeñas. Las hélices restantes muestran cambios pequeños en algunos residuos, generalmente en los extremos, siendo la única excepción la estabilización de algunos residuos en la mitad de la hélice H5. A 450 K la hélice más afectada es la H7, mostrando un perfil inverso al observado a 400 K: los extremos de la hélice son los más estables, disminuyendo la estabilidad linealmente hacia el centro, lo cual muestra que la separación de los dominios inestabiliza fuertemente esta hélice. En efecto, al comparar las matrices  $R$  de ambas simulaciones es evidente que mientras en la proteína completa H7 mantiene un grado elevado de estructura residual hacia el nanosegundo de simulación, el dominio amino aislado pierde toda estructura secundaria ya a los 200 ps. Las demás hélices, tanto en el dominio amino como en el carboxilo, presentan una pequeña estabilidad en la mayoría de sus residuos, siendo excepciones la hélice H1, H2 y H8. A 500 K los cambios son pequeños, pero es posible observar que la estabilidad de las hélices sigue el mismo patrón observado a 450 K. En el caso de las hebras beta se ven las primeras diferencias, la mayoría de ellas muestra una desestabilización, siendo excepcionales algunas zonas de las hebras S5, S6, S7 y S10.

En resumen, se puede concluir que los dominios aislados de PaFtsZ son más inestables que los de MjFtsZ donde el más afectado es el dominio carboxilo y principalmente la hélice H7.



**Figura 14. Comparación del valor de estructura secundaria medio para MjFtsZ silvestre ( $S_{1/2}^m$ ) y el de los dominios aislados ( $S_{1/2}^{dom}$ ).** Un valor de  $S_{1/2}^m - S_{1/2}^{dom}$  positivo para el residuo  $m$  indica que la separación de los dominios produce una desestabilización de la estructura secundaria en el residuo, mientras que valores negativos muestran una estabilización. Las zonas en color gris oscuro corresponden a hélices alfa, mientras que las en gris claro corresponden a sabanas beta. Solo se muestran los valores para los residuos que mostraron diferencias de estabilidad.

### Contactos nativos de los núcleos hidrófobos

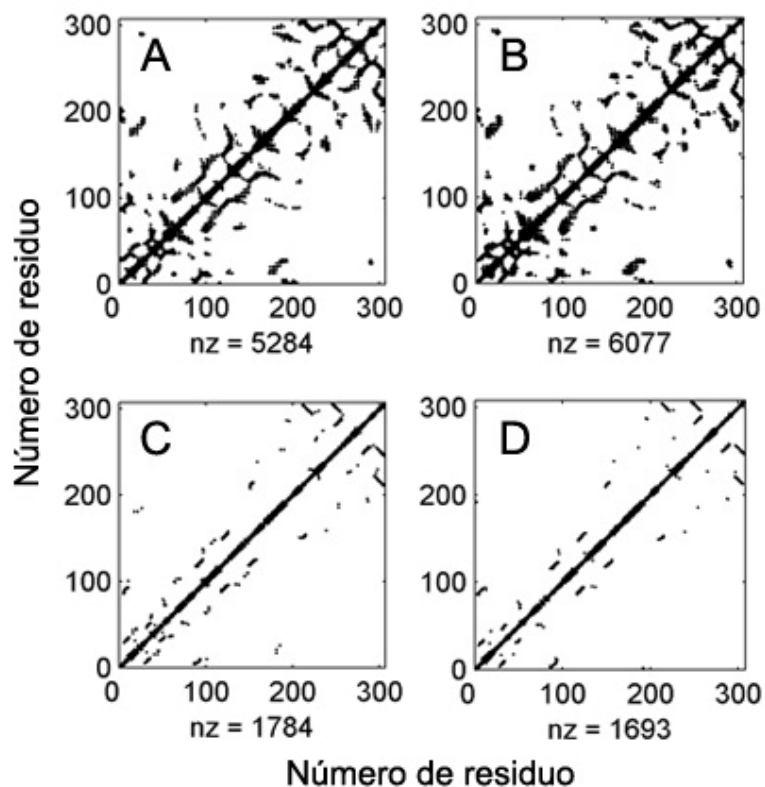
En las próximas secciones, se estudiará el desplegamiento analizando la estructura terciaria de las proteínas, basándonos en el comportamiento de los núcleos hidrófobos de FtsZ (Fig. 5).

En la Figura 15 se muestran las matrices  $F$  y  $E$  construidas a partir de las simulaciones a 300 K (ver Materiales y Métodos). Las matrices  $F$  de PaFtsZ y MjFtsZ tienen un perfil similar (Fig. 15A y 15B), sin embargo, el número de contactos (nz) de PaFtsZ (Fig. 15B) es mayor que los de MjFtsZ (6077 vs 5284, Fig. 15A). Por otra parte, la matriz  $E$  de MjFtsZ tiene un mayor número de contactos nativos (1784 vs



1693), lo que sugiere que las diferencias en el número de contactos para las matrices de frecuencia se deben principalmente a contactos transitorios. Esto sugiere que MjFtsZ tiene núcleos con mayor número de contactos y que PaFtsZ presenta mayor movilidad que MjFtsZ, pues genera un mayor número de contactos transitorios.

Para caracterizar el comportamiento de los núcleos hidrófobos en forma separada, se extrajeron a partir de la matriz *E* los contactos entre residuos de los tres núcleos hidrófobos, definidos como conjuntos de residuos no accesibles al solvente que forman contactos entre ellos (Ver Materiales y Métodos). La Tabla 4 muestra el número de contactos nativos para cada núcleo hidrófobo. Es posible observar que la simulación hace desaparecer entre 34 y 50% de los contactos de la estructura cristalina, lo cual está cercano a lo reportado para simulaciones de la proteína dihidrofolato reductasa utilizando el campo de fuerza EEF1 (Sham y cols. 2002). Tanto en la estructura cristalina como en las simulaciones, los núcleos hidrófobos de la interfase (I) y del dominio amino (A) tienen un número comparable de contactos nativos, que es aproximadamente tres veces mayor al número de contactos nativos del dominio carboxilo (C). Al comparar el número de contactos nativos de los núcleos hidrófobos de PaFtsZ y MjFtsZ, se observa que los núcleos de MjFtsZ tienen un mayor número de contactos nativos que la proteína mesófila, y que el porcentaje de contactos nativos es mayor a los de la estructura cristalina.



**Figura15. Contactos en la estructura nativa de MjFtsZ y PaFtsZ.** Paneles A. y B. Matrices de frecuencias de contactos  $F$  para las simulaciones a 300 K de MjFtsZ y PaFtsZ, respectivamente. Cada punto representa un contacto que presente en las simulaciones a 300 K, independiente de su frecuencia. **Paneles C. y D.** Matrices de contactos nativos ( $E$ ) de MjFtsZ y PaFtsZ. Cada cuadro negro es un contacto que esta presente en más del 70% de las estructuras.  $nz$  corresponde al número de coeficientes de la matriz que son distintas de cero.

Para comparar el número de contactos nativos de los dominios aislados respecto a los de la proteína nativa, se calcularon matrices de frecuencia de contactos a partir de las simulaciones a 300 K de los dominios aislados de PaFtsZ y MjFtsZ (Tabla 4). Los resultados muestran que la separación de los dominios hace disminuir el número de contactos nativos de PaFtsZ pero no los de MjFtsZ.

**Tabla 3. Número de contactos nativos de los núcleos hidrófobos.** Un contacto fue considerado nativo si se presenta en más de 70% de las estructuras de las simulaciones a 300 K. Los números entre paréntesis corresponden al porcentaje del número de contactos de los núcleos en el cristal.

proteína	Cristal <sup>1</sup>			Silvestre <sup>2</sup>			Dominios <sup>2</sup>	
	I <sup>3</sup>	A <sup>4</sup>	C <sup>5</sup>	I	A	C	A	C
PaFtsZ	89	84	35	51(57)	53(63)	19(54)	55(65)	18(51)
MjFtsZ	86	90	37	56(65)	60(66)	22(59)	59(65)	22(59)

<sup>1</sup>Contactos calculados a partir de la estructura cristalina.

<sup>2</sup>Contactos calculados a partir de matrices de frecuencia de simulaciones de proteínas nativas.

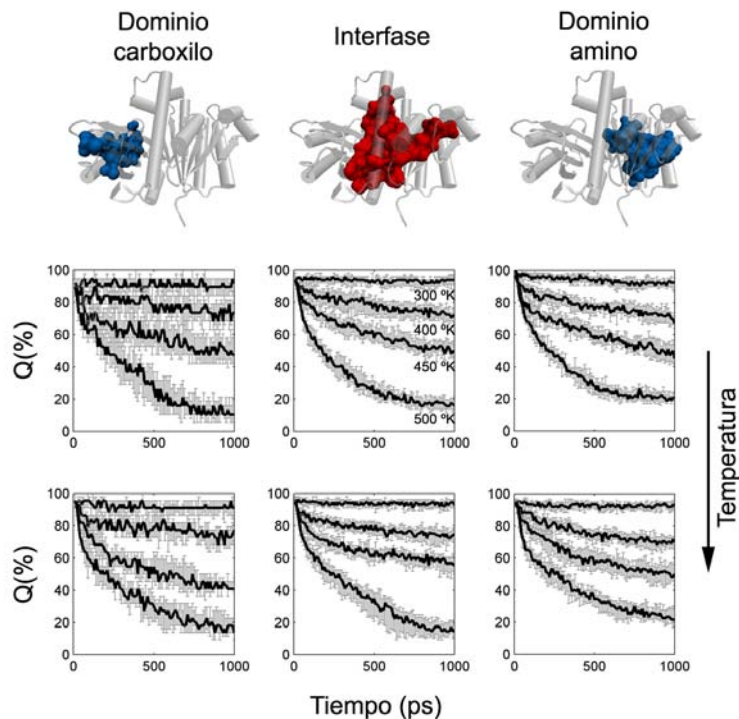
<sup>3</sup>Interfase

<sup>4</sup>Dominio amino

<sup>5</sup>Dominio carboxilo

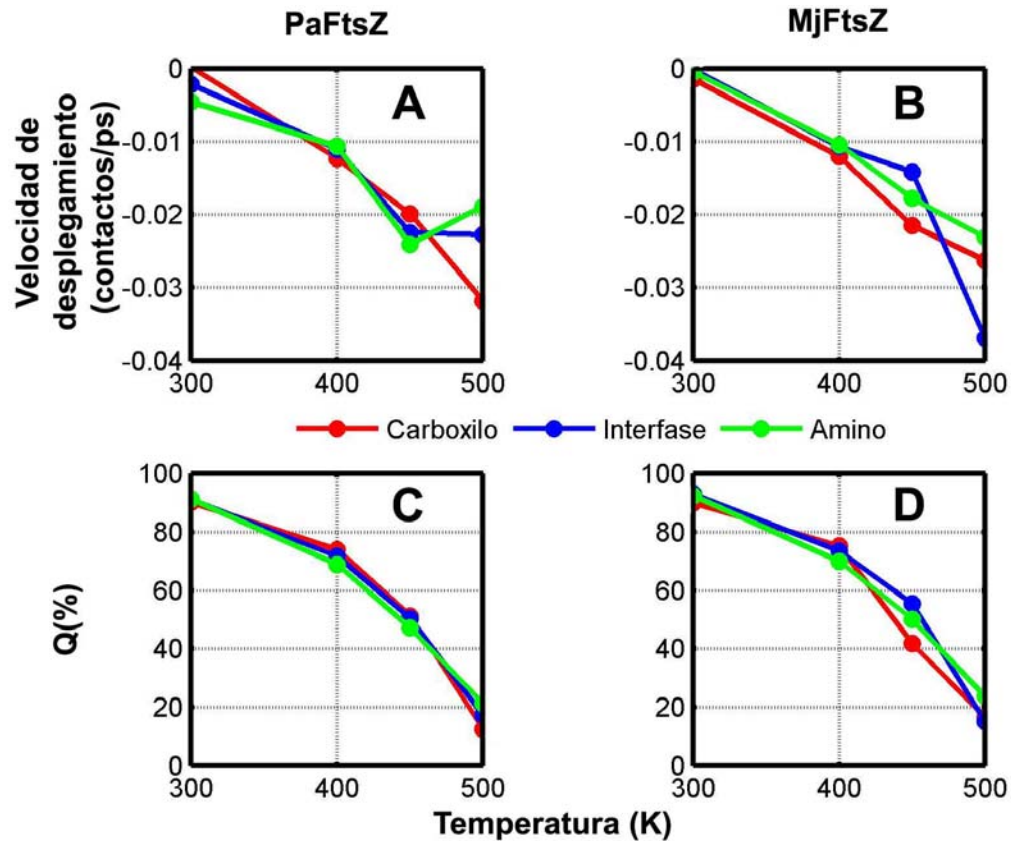
### Estabilidad relativa de los núcleos hidrófobos

Los resultados anteriores muestran que hay diferencias en el número de contactos de los distintos núcleos hidrófobos, lo cual es consistente con los resultados de secciones anteriores que muestran que la estabilidad de la estructura secundaria del dominio carboxilo es menor que la del dominio amino. ¿Implica esto que el núcleo hidrófobo del dominio carboxilo es menos estable que el núcleo hidrófobo de la interfase y del dominio amino? Para responder esta pregunta se calculó el cambio en el tiempo en el número de contactos nativos (ver Materiales y Métodos) de PaFtsZ y MjFtsZ, para cada una de las simulaciones de 1 ns realizadas a 300, 400, 450 y 500 K (Fig 16). Para poder comparar el comportamiento del núcleo del dominio carboxilo con los de la interfase y el del dominio amino los datos se expresan como porcentaje de contactos respecto a la estructura nativa (Q(%)),



**Figura 16. Conducta cinética del porcentaje de contactos nativos<sup>6</sup> Q de los núcleos hidrófobos de PaFtsZ y MjFtsZ. A.** Cadenas laterales de los residuos de los núcleos hidrófobos de MjFtsZ en superficie molecular, **B y C.** Porcentaje de contactos nativos (Q) para los núcleos de PaFtsZ y MjFtsZ, respectivamente. Cada uno de los gráficos muestra resultados para simulaciones a 300, 400, 450 y 500 K. Cada curva representa el promedio (líneas negras) y desviación estándar (barras de error grises) de 20 simulaciones.

En la figura 16 se observa que el Q(%) de los núcleos disminuye asintóticamente en las simulaciones a alta temperatura, no llegando a un estado estable tras 1 ns de simulación. Puede observarse también que la dispersión de los datos de Q(%) (barras de error grises) son mayores en el dominio carboxilo, lo cual se debe en parte al reducido número de contactos nativos de este núcleo. Fuera de estas diferencias, los núcleos de PaFtsZ y MjFtsZ muestran un comportamiento similar a diferentes temperaturas, lo que sugiere que las diferencias de estabilidad entre estos dominios son pequeñas.



**Figura 17. Efecto de la temperatura en el desplegamiento de los núcleos hidrófobos de PaFtsZ y MjFtsZ. A y B,** cambio en la velocidad de desplegamiento del dominio carboxilo (rojo), la interfase (verde) y el dominio amino (azul). Esta se calculó a partir de la pendiente de los últimos 600 ps de las curvas de desplegamiento mostradas en la Fig. 16. **C y D,** número promedio de contactos nativos al final de las simulaciones de 1 ns.

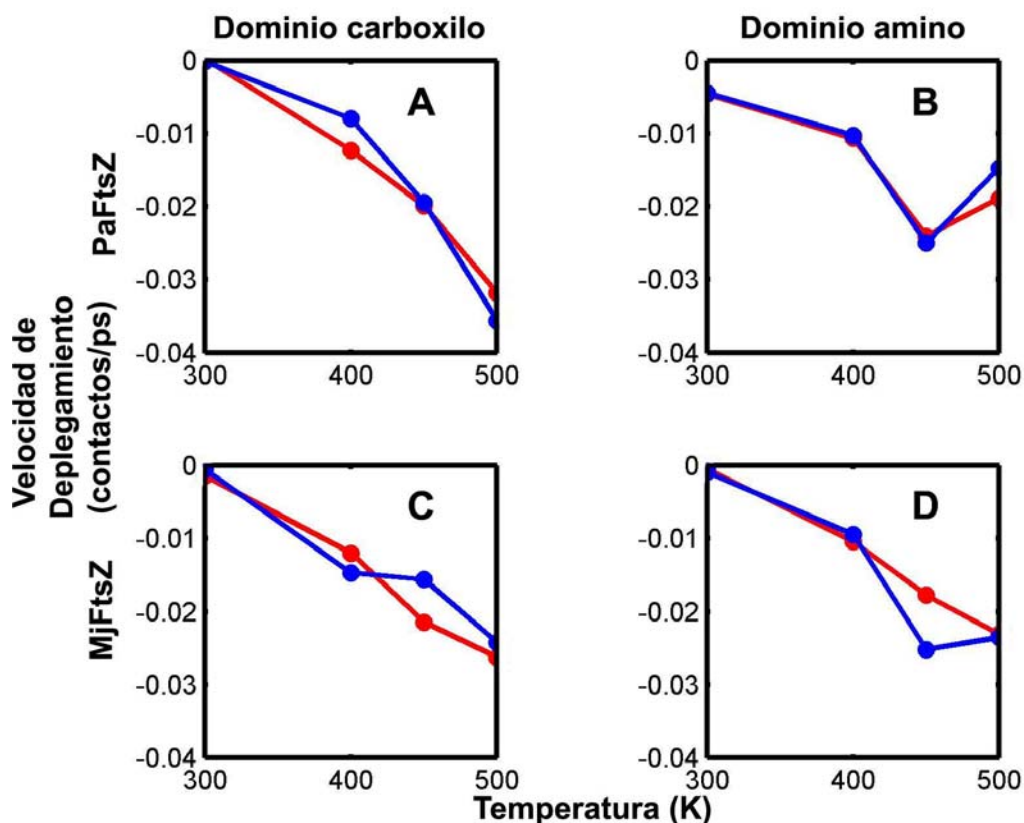
En los primeros 100 ps de simulación, cuando se calienta lentamente el sistema, el número de contactos nativos decae exponencialmente para luego disminuir en forma lineal. Al comparar las velocidades de desplegamiento obtenidas a partir de un ajuste lineal de los últimos 600 ps de simulación, se observa que para ambas proteínas la estabilidad de los tres núcleos es similar (Fig. 17A y B). En el caso de PaFtsZ (Fig. 17A) las velocidades de desplegamiento de los tres núcleos hidrófobos es similar hasta los 450 K, observándose diferencias solo a 500 K, donde el dominio amino se despliega con una cinética más lenta que la de los núcleos de la interfase y del dominio carboxilo. En el caso de MjFtsZ (Fig. 17B), aunque las velocidades de

desplegamiento son similares, el desplegamiento del dominio carboxilo es más rápido que el del dominio amino y la interfase. A 500 K, el núcleo de la interfase se vuelve menos estable que los núcleos de los dominios. Al usar el número promedio del porcentaje de contactos nativos en los últimos 200 ps de simulación como medida de la estabilidad relativa de los dominios, las diferencias cinéticas antes mencionadas no son evidentes (Fig. 17C y 17D)

A modo de resumen, los resultados generales de esta sección muestran que los núcleos hidrófobos de MjFtsZ y PaFtsZ muestran una velocidad de desplegamiento similar, observándose una disminución de la estabilidad del núcleo carboxilo y la interfase con respecto al núcleo del dominio amino a 450 y 500 K.

### **Influencia de la interfase en la estabilidad de los núcleos hidrófobos**

Para comparar las estabildades de los núcleos de los dominios amino y carboxilo en los dominios aislados y la proteína silvestre, se calcularon la velocidad de desplegamiento y el porcentaje de contactos nativos al final de las simulaciones. Este último análisis no muestra cambios significativos (datos no mostrados) y se muestran solo los cambios en velocidad de plegamiento (Fig. 18). Ninguno de los dominios aislados muestra una clara tendencia al aumento de la velocidad de desplegamiento, en efecto, se observa una disminución de la velocidad de desplegamiento a algunas temperaturas (Fig. 18A, 400 K; 18B, 500 K; 18C 400 y 450 K; 18D, 450 K ). El dominio amino de PaFtsZ destaca por disminuir la pendiente de desplegamiento a 500 K, lo que se debe a que en este caso la pendiente de la componente exponencial es mayor que en otros casos, disminuyendo la pendiente de la segunda componente lineal.



**Figura 18. Efecto de la interfase de interacción en la estabilidad de los núcleos hidrófobos.** La velocidad de desplegamiento se obtuvo ajustando una función lineal a los últimos 600 ps de las curvas de cambio en el porcentaje de contactos nativos. Los dominios de las proteínas silvestres se muestran en color rojo, y las de los dominios aislados en color azul. **A.** Dominio carboxilo de PaFtsZ, **B.** Dominio amino de PaFtsZ. **C.** Dominio carboxilo de MjFtsZ, **D.** Dominio amino de MjFtsZ.

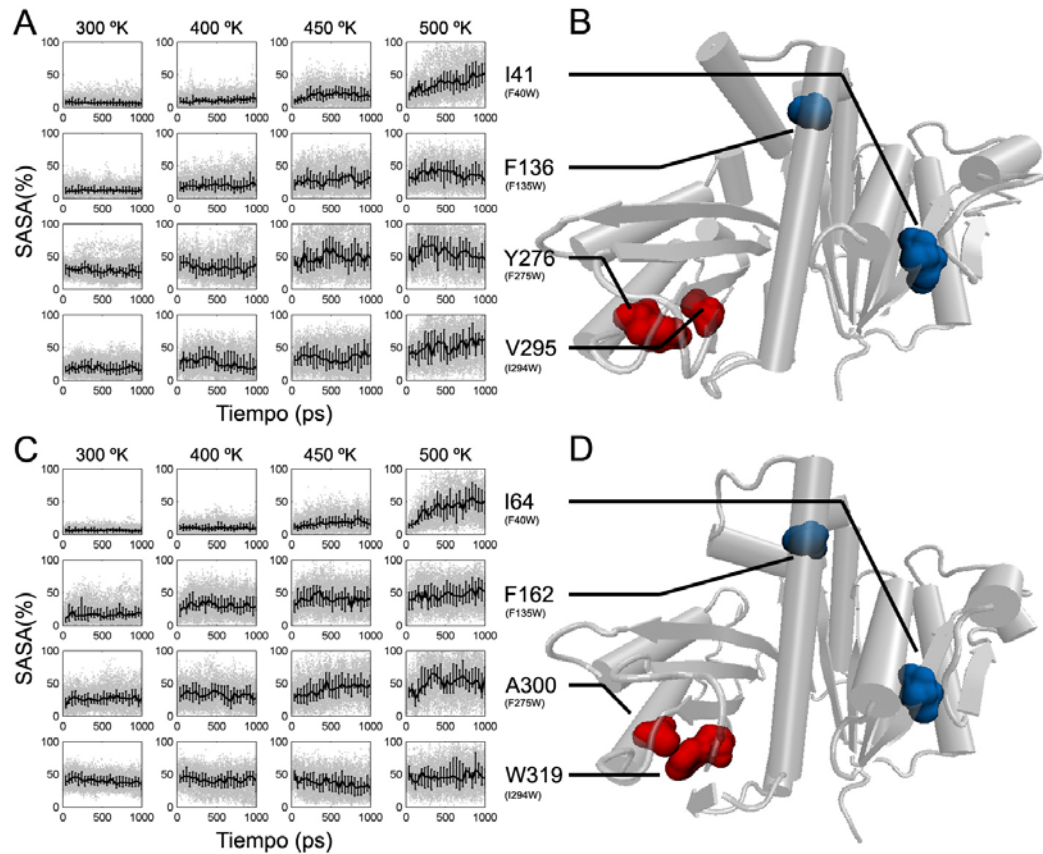
#### **Análisis de accesibilidad al solvente de los residuos cambiados en las mutantes de triptófano de EcFtsZ: F40W, F135W, F275W e I294W**

Uno de los métodos más sensibles para seguir el desplegamiento de proteínas es la fluorescencia intrínseca de triptófanos. La FtsZ, de *E. coli* se ha usado para caracterizar su plegamiento, pues carece de triptófanos y se puede introducir este aminoácido como sonda intrínseca en lugares determinados para seguir su plegamiento y desplegamiento. Con este propósito, en trabajos anteriores se construyeron mutantes con un triptófano en FtsZ: F40W, F135W, F275W e I294W,

donde F40W y F135W actúan como sondas intrínsecas para el plegamiento del dominio amino y F135W y F275W para el carboxilo (Fig. 19 B y D) (Diaz-Espinoza y cols. 2007). Cada una de ellas mostró un desplegamiento reversible, sugiriendo que el plegamiento del dominio carboxilo y el dominio amino proceden en forma simultánea. Adicionalmente se observó que las mutantes F40W y F275A presentan intermediarios estables entre 0,5 y 1M de GdmCl, mientras que F135W y I294W presentan un plegamiento más cooperativo con la ausencia de intermediarios, indicativo de un proceso de dos estados.

Para entender los cambios estructurales responsables de los cambios de la fluorescencia intrínseca del triptófano de las mutantes, se calculó en las simulaciones de PaFtsZ y MjFtsZ el cambio en la accesibilidad al solvente de los residuos que se mutaron por triptófano. Como los experimentos de fluorescencia dan cuenta del comportamiento de un colectivo de moléculas, se calculó la accesibilidad de los residuos cambiados en las 20 simulaciones a cada temperatura, y se muestra el promedio temporal junto con la dispersión de los datos (Fig. 19A). Para poder comparar los resultados experimentales usando desnaturación química con los resultados de simulación utilizando desnaturación térmica, es necesario definir como se va a establecer la presencia o ausencia de intermediarios al analizar los cambios de accesibilidad a solvente. Por esto, se consideró que un residuo presenta un intermediario si durante la desnaturación térmica, un cambio en la temperatura no afecta la accesibilidad a solvente en forma significativa.





**Figura19. Cambio en el porcentaje de accesibilidad al solvente de los residuos de PaFtsZ y MjFtsZ que se reemplazaron en las mutantes de EcFtsZ. Panel A. PaFtsZ Panel C. MjFtsZ. Las columnas corresponden a simulaciones a distintas temperaturas, mientras que las filas corresponden a los residuos equivalentes a las mutantes de triptófano (derecha, entre paréntesis). Cada gráfico muestra la dispersión de los datos de 20 simulaciones en puntos grises y el promedio temporal junto con la desviación estándar en líneas negras. Paneles B. y D. Dibujos esquemático de la posición de los residuos equivalentes a EcFtsZ en PaFtsZ y MjFtsZ.**

En PaFtsZ, se observa que el residuo V295 (Fig 19A, fila 4) no presenta un intermediario puesto que la accesibilidad a solvente incrementa con la temperatura. Por el contrario, en el residuo I41 (Fig 19A, fila 1) se observa que el paso de 300 a 400 K no genera un cambio en la accesibilidad a solvente aumentando levemente a 450 K. Solo a 500 K es posible observar una transición completa. Este comportamiento indica que el residuo I41 podría ser parte de un intermediario en el desplegamiento con GdmCl. Por otra parte, F136 e Y276 no serían parte de intermediarios. F136 (Fig 19A, fila 2) presenta un perfil similar a V295 hasta 450 K, sugiriendo que es parte de un intermediario. Sin embargo la accesibilidad a solvente a 500 K no es notablemente

superior a la observada a 450 K, lo que sugiere la posibilidad de haber llegado o bien al máximo de accesibilidad de solvente de este residuo en el contexto en el cual se encuentra, o a un intermediario. En efecto, al verificar los aminoácidos cercanos en el esqueleto de la proteína, es posible observar que esta fenilalanina se encuentra haciendo un acoplamiento aromático con otra fenilalanina que esta separada de la primera por un residuo de prolina. Las restricciones conformacionales de la prolina, junto con el elevado volumen de la fenilalanina, pueden explicar que este residuo no llegue a una accesibilidad de 100% en el estado desplegado. Tomando esto en consideración, el residuo F136 presenta un comportamiento más cercano a la ausencia de intermediarios. Por otra parte, el residuo Y276 muestra un comportamiento similar a F136, pero con una mayor dispersión, lo que da cuenta de la mayor flexibilidad estructural del dominio carboxilo.

El comportamiento de MjFtsZ es similar al de PaFtsZ (Fig. 19C). El residuo I64 tiene un comportamiento casi idéntico al mostrado por I41, sugiriendo la presencia de un intermediario; F162 tiene un perfil similar a F136, sin embargo es posible observar que el comportamiento a 500 K se aleja en una forma más notoria del mostrado a 400 K, sugiriendo una transición más cooperativa. Lo mismo ocurre con A300, que se parece más al comportamiento temperatura-dependiente de F162. La gran diferencia la muestra W319, que pareciera tener un intermediario de accesibilidad a solvente reducida a 450 K. Adicionalmente a este intermediario, es posible ver que la accesibilidad del residuo W319 es casi insensible a la temperatura, lo que puede deberse al elevado volumen del residuo de triptófano, encontrándose semi-accesible en la mayoría de las simulaciones.

Un resumen de los resultados del análisis anterior se muestra en la Tabla 6. Las simulaciones de PaFtsZ y MjFtsZ son consistentes con la observación experimental, pues al seguir el plegamiento de la mutante F40W de EcFtsZ con fluorescencia se observa la presencia de intermediario.

**Tabla 4. Resumen de los resultados del análisis de accesibilidad al solvente de los residuos equivalentes a las mutantes de triptófano F40W, F135W, F275W y I294W de EcFtsZ.**

EcFtsZ		PaFtsZ		MjFtsZ	
Mutante	I <sub>E</sub> <sup>1</sup>	Residuo <sup>2</sup>	I <sub>S</sub> <sup>3</sup>	Residuo <sup>2</sup>	I <sub>S</sub>
F40W	si	I41	si	V67	si
F135W		F136		F162	
F275A	si	Y276		A300	
I294W		V295		W319	

<sup>1</sup> Presencia de intermediario en estudios de desnaturación con GdmCl

<sup>2</sup> Nombre y número de residuo equivalente en MjFtsZ o PaFtsZ.

<sup>3</sup> Presencia de intermediario en simulaciones de desplegamiento. L.

## DISCUSIÓN

En el presente estudio hemos caracterizado, por medio de dinámica molecular con solvente implícito, las vías de desplegamiento y las estabilidades de elementos de estructura secundaria y núcleos hidrófobos de una proteína FtsZ mesófila y una termófila, y hemos utilizado esta información para evaluar el efecto de la interfase de interacción en la estabilidad de los dominios amino y carboxilo. Para establecer estas diferencias usamos métodos clásicos para establecer la estabilidad de simulaciones nativas, tales como RMSF, RMSD y cambio de estructura secundaria en el tiempo, así como también introdujimos un nuevo método para analizar el cambio de estructura secundaria en simulaciones de desplegamiento, el cual llamamos matrices  $R$  (de matrices de estructura secundaria Residual), así como también los perfiles de valores  $S_{1/2}^m$ , que pueden derivarse de dichas matrices. Adicionalmente seguimos el desplegamiento de cada uno de los núcleos hidrófobos por separado para comparar su estabilidad relativa. Los resultados confirman nuestra hipótesis de trabajo, y adicionalmente muestran datos inesperados que pueden ser utilizados para explicar datos experimentales para los cuales no se tenía evidencia directa. A continuación se discuten las limitaciones del método de estudio sobre los resultados, así como también las observaciones hechas sobre el desplegamiento de FtsZ y sus dominios.

### **Evaluación del método de solvente implícito**

La mayoría de los campos de fuerza y métodos de simulación disponibles en la literatura han sido derivados buscando reproducir las propiedades energéticas y geométricas a temperatura y presión normales, usando pequeños compuestos modelo que tienen características similares a los grupos químicos presentes en las proteínas. De este modo, los campos de fuerza están optimizados para reproducir la dinámica nativa, siendo los estudios a altas temperaturas una extrapolación del comportamiento en dichas condiciones. De este modo, las simulaciones de desplegamiento han sido, antes que nada, una especulación.

De los métodos de simulación de desplegamiento descritos en la literatura, ENCAD ha mostrado una de las mejores correlaciones con datos experimentales de valores  $\Phi$ . Sin embargo, al momento de comenzar este estudio, estos métodos no estaban disponibles para su uso por parte de la comunidad científica. Por otra parte, aunque se ha descrito que usando 498 K es posible desplegar CI2 en cerca de 500 ps, estos resultados no son reproducibles con otros campos de fuerza utilizados comúnmente para estudiar proteínas en estado nativo (Valerie Daggett, comunicación personal). En efecto, simulaciones de solvente explícito utilizando campos de fuerza como AMBER (Ponder y Case 2003b) y GROMOS96 (Schuler y cols. 2001) son incapaces de desplegar completamente proteínas a altas temperaturas aún después de 10 ns, siendo evidentes sólo las etapas iniciales del desplegamiento. Tres posibles factores pueden explicar este comportamiento. Primero, ENCAD puede ser un campo de fuerza intrínsecamente inestable; segundo, los métodos de Ewald para tratar interacciones electrostáticas en simulaciones con condiciones de borde periódicas utilizados con campos de fuerza distintos a ENCAD sobre-estabilizan los sistemas (Weber y cols. 2000), mientras que el grupo de Daggett usa métodos de truncamiento de interacciones no enlazantes esféricos del tipo corrimiento de fuerza (*forceshift*) que no presentan este tipo de comportamiento y tercero, los métodos de truncamiento esféricos para interacciones van der Waals y electrostáticas usado por Daggett y colaboradores usan distancias menores a las comúnmente usadas por otros campos de fuerza (8Å frente a 12Å). En efecto, simulaciones llevadas a cabo con el campo de fuerza CHARMM19 con truncación esférica de 7,5Å y solvente explícito pueden desplegar a la proteína rubedoxina en alto grado en la escala de picosegundos (Lazaridis y cols. 1997). Tomando eso en cuenta, se buscaron métodos de simulación molecular que mostraran las propiedades mencionadas para ENCAD y que no utilizaran solvente explícito.

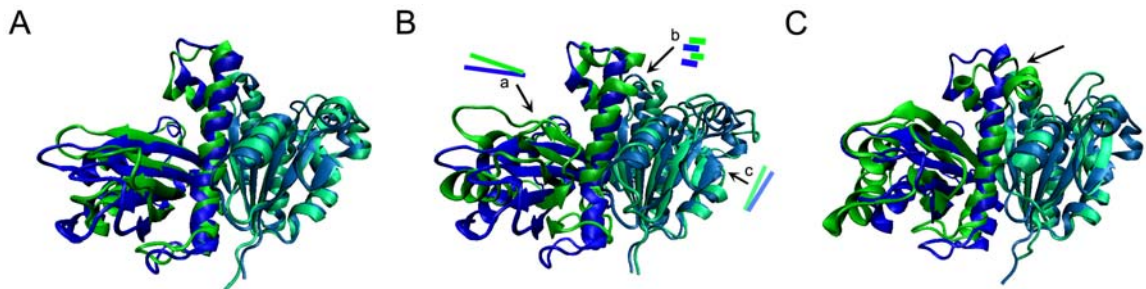
En la literatura (Lazaridis y Karplus 1999) solo un formalismo de solvente implícito se ha probado en el desplegamiento de CI2, mostrando buena correlación con los datos experimentales y simulaciones de solvente explícito, llamado EEF1 (*Effective Energy Function 1*). Adicionalmente, este campo de fuerza ha sido utilizado para estudiar el proceso de plegamiento de otras proteínas mostrando buena correlación de

los datos experimentales. Por ello, se decidió utilizarlo para simular el desplegamiento de las proteínas FtsZ.

Las simulaciones de FtsZ mostraron ser estables a 300 K, lo cual sirvió como un control positivo para la realización de simulaciones a temperaturas de desplegamiento. Sin embargo, en las simulaciones se evidenciaron ciertos comportamientos anómalos que resulta necesario discutir para establecer la validez de las observaciones de esta tesis. En primer lugar, en las simulaciones a 300 K de MjFtsZ, se observó la pérdida de estructura secundaria en la parte amino de la hélice H7 (Fig. 7). Al analizar la estructura de la proteína, pudo verificarse la formación de un quiebre de la hélice en los residuos D212 y D213, produciendo que los residuos K205 y K209 se dirigieran al sitio activo. Es probable que estos cambios se produzcan debido a la falta de solvatación explícita, lo cual puede aumentar la energía de interacción de residuos cargados. Para evitar el efecto de estos cambios en los análisis, se decidió llevarlos a cabo durante el primer nanosegundo de simulación a distintas temperaturas, favoreciendo de este modo que las diferencias observadas entre las simulaciones se debieran a las diferencias en estabilidad entre las proteínas y no a posibles artificios generados por el método. En segundo lugar, durante las simulaciones de desplegamiento fue posible observar la presencia de intermediarios donde los dos dominios se unían generando una gran sabana plisada (Figs. 9B, 400 K y 10B, 450K ), lo cual sugeriría que el desplegamiento ocurriría con intermediarios que solo tienen estructura  $\beta$ . Es posible que la elevada estabilidad de las estructuras  $\beta$  sea resultado de la carencia de solvente explícito, el cual podría desestabilizar las hebras  $\beta$  al competir sus grupos dadores y aceptores de puentes de hidrógeno con los de la proteína. Fuera de estas observaciones, que posiblemente se deben a artificios producidos por el método de estudio, se observaron diferencias que dan luces acerca de las propiedades de las proteínas en cuestión, las cuales se discuten a continuación.

## Cambios conformacionales a 300 K

Una de las primeras observaciones de este estudio dice relación con la movilidad de los dominios de FtsZ. El elevado RMSD presentado por las simulaciones sugería una alta movilidad entre los dominios amino y carboxilo de PaFtsZ y MjFtsZ. Al analizar visualmente las dinámicas fue posible verificar que el movimiento mencionado corresponde al descrito para el cambio conformacional entre un estado recto favorecido en la presencia de GTP, a un estado curvo favorecido en presencia de GDP, que correspondería a la ausencia de nucleótido en las simulaciones. Estudios comparativos realizados por otros investigadores durante el transcurso de esta tesis mostraron que el cristal 1W5B (MjFtsZ) presenta una conformación tipo curva, (Fig. 20A, verde) mientras que 1OFU (PaFtsZ) presenta una conformación recta (Fig. 20A, azul) (Buey y cols. 2006; Nogales y Wang 2006).



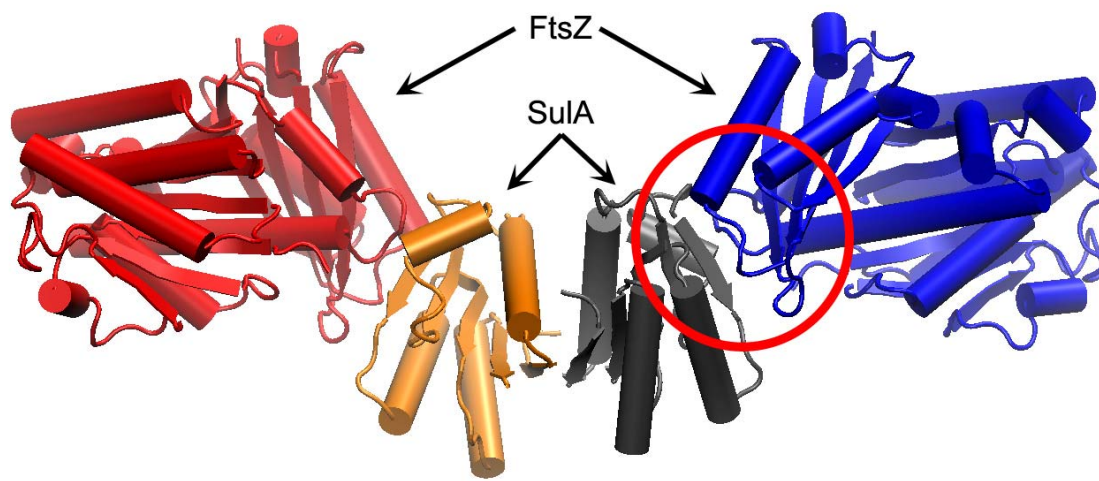
**Figura 20. Superposición estructural de las estructuras iniciales de PaFtsZ y MjFtsZ con las obtenidas después de 5ns de simulación, por medio de ajuste del dominio amino. Panel A.** estructura cristalina de PaFtsZ (azul, en estado recto ) y de MjFtsZ (verde, en estado curvo). **Panel B.** Estructura inicial de PaFtsZ tras minimización energética (azul) y después de 5 ns de dinámica (verde). a) muestra el cambio de orientación del dominio carboxilo, b) el ascenso de la hélice H7 y c) el corrimiento experimentado por las sabanas del dominio amino. **Panel C.** Estructura inicial de MjFtsZ tras minimización energética (azul) y después de 5 ns de dinámica (verde). Con una flecha se muestra el quiebre en la hélice H7 producido por interacciones fuertes entre residuos cargados.

Los mayores cambios producidos durante el supuesto “switch” conformacional son el movimiento ascendente del dominio carboxilo y el ascenso de la hélice H7 con respecto al dominio amino. El caso de PaFtsZ es diferente a las demás estructuras cristalizadas de FtsZ, pues junto a las de *Mycobacterium tuberculosis* son las únicas que presentan una conformación tipo “recta”. Aún por haber sido cristalizada unida a

una molécula de GDP, la cual favorece el estado curvo en los monómeros, PaFtsZ se mantiene en estado recto debido a la unión del inhibidor Sula. Durante las simulaciones del presente estudio tanto el ligando como el inhibidor Sula fueron removidos; estas interacciones que podrían haber congelado el monómero en un estado recto, al desaparecer pudieron permitir el cambio conformacional anteriormente descrito, afectando también la estabilidad de este dominio (Fig. 21).

Se ha propuesto que este cambio conformacional es suficiente para promover los cambios en estado de polimerización de tubulina, lo cual podría extenderse a FtsZ. Al comparar la estructura inicial y final de la simulación de PaFtsZ es posible observar cambios consistentes con una transición recto-curvo que ha sido propuesta para estas proteínas (Fig. 20B a y b). Adicionalmente, es posible observar un pequeño cambio de conformación en el dominio amino de PaFtsZ, el cual se observa en la superposición de las hebras beta de los dos dominios (Fig. 20B c). Este corrimiento puede ayudar a explicar, junto con el desordenamiento de los loops de la estructura del dominio amino, el elevado RMSD de la estructura de PaFtsZ durante la dinámica a 300 K. Buscando una interpretación para este corrimiento, encontramos en la literatura que tras la cristalización de  $\gamma$ -tubulina, a diferencia de tubulina y FtsZ, el monómero en presencia de GTP se encuentra en un estado curvo (Aldaz y cols. 2005). Al superponer el estado curvo de  $\gamma$ -tubulina con uno recto de  $\beta$ -tubulina fue posible observar un reordenamiento en el dominio amino del primero, similar al experimentado por PaFtsZ en la dinámica (Fig. 20B c). De este modo, estos cambios pueden considerarse una posibilidad estructural compartida entre esta familia de GTPasas.





**Figura 21. Contactos cristalográficos entre el dominio carboxilo de PaFtsZ y el inhibidor SulA.**

Al analizar los cambios experimentados por MjFtsZ, es posible observar que la proteína experimenta un movimiento similar al observado en PaFtsZ (Fig. 20C), aun cuando el estado inicial corresponde a un estado de tipo curvo. Una posible explicación para este comportamiento es que la proteína MjFtsZ en estado curvo se cristalizó con ligando, y la remoción de éste le confiere al monómero una mayor flexibilidad que le permite explorar estados de mayor curvatura. Sin embargo, esto no es consistente con los datos que muestran que la estructura de *Mycobacterium tuberculosis* con citrato en su sitio activo tiene una estructura tipo recta, ni con los datos que muestran que MjFtsZ es capaz de polimerizar sin ligando (Huecas y Andreu 2004). En efecto, nuevas estructuras cristalinas publicadas durante el desarrollo de esta memoria (FtsZ de *Methanococcus jannashii* a 1.7 Å de resolución, FtsZ de *Pseudomonas aeruginosa* sin la presencia del inhibidor SulA, y estructuras de FtsZ de *Bacillus subtilis* y *Aquifex aeolicus*) muestran que los cambios de orientación entre los dominios amino y carboxilo son mayores al comparar estructuras de distintas especies que al comparar estructuras de la misma especie pero en presencia o ausencia de ligando (Oliva y cols. 2007), sugiriendo que los cambios conformacionales producidos por la unión de ligando, de existir, serían sutiles y estarían asociados a la flexibilidad de la zona de unión a ligando. Sin embargo, también se mostró que el dominio carboxilo podía presentar modificaciones estructurales en base al tipo de empacamiento molecular del

cristal, lo cual sugiere que el dominio carboxilo posee una flexibilidad intrínseca. En efecto, estudios de mutagénesis sitio dirigida del dominio carboxilo están de acuerdo con esta observación (Diaz-Espinoza y cols. 2007).

Otra alternativa es que este rápido cambio entre conformaciones se deba al efecto del solvente implícito sobre la estructura de las proteínas. Se ha demostrado que el solvente implícito GB/SA induce un desplazamiento del estado de transición de plegamiento de la miniproteína BBA5 a estados más “nativos”, lo cual implica que estos solventes favorecen estados de mayor compactación, al menos en conformaciones semi-nativas (Rhee y cols. 2004). El colapso de los lazos del dominio carboxilo sobre la hélice H7 de ambas proteínas esta de acuerdo con esta interpretación (Fig. 20 B y C). No obstante la presencia de errores en las simulaciones al utilizar formalismos de solvente implícito, las simulaciones muestran que existe una mayor flexibilidad en el dominio carboxilo, lo cual esta de acuerdo con los datos experimentales. Sin embargo, el rol que juega esta flexibilidad en la dinámica de polimerización requiere de nuevos evidencias tanto computacionales como experimentales.

### **Comparación de la estabilidad de MjFtsZ con la de PaFtsZ**

La diferencia en estabilidad de proteínas homólogas depende de las diferencias existentes en su secuencia de aminoácidos. Al comparar la composición aminoacídica de proteínas provenientes de especies mesófilas y termófilas del género *Methanococcus* se observaron cambios estadísticamente significativos (Haney y cols. 1999). Proteínas termófilas tendían a presentar un aumento en la hidrofobicidad, volumen y carga de sus residuos, disminuyendo a la vez la proporción de residuos hidrofílicos. Estudios estructurales han mostrado que estas tendencias se traducen en un aumento de la proporción de residuos en alfa hélices, la estabilización de su dipolo por residuos cargados que actúan en forma de capucha (*cap*), así como también en un aumento de puentes salinos superficiales (Chakravarty y Varadarajan 2002; Kumar y cols. 2000). Se ha sugerido que el efecto de estas modificaciones es la rigidización de las proteínas termófilas, haciendo que su movilidad a altas temperaturas sea similar a la movilidad de proteínas mesófilas a bajas temperaturas.

El comportamiento dinámico de PaFtsZ y MjFtsZ concuerdan con estas predicciones. En primer lugar, tanto la flexibilidad local, evidente en la fluctuación de los lazos (Fig. 8) como la flexibilidad global, evidente en los mayores cambios durante el posible *switch* conformacional (Fig. 6 y Fig 20), así como el número de contactos transitorios para éstas a 300 K (Fig. 15) demuestran la mayor flexibilidad de PaFtsZ. Adicionalmente, la flexibilidad del dominio carboxilo es mayor que la del dominio amino, lo cual permite explicar que la mutante I294W de EcFtsZ (Fig. 19 B y D) sea capaz de polimerizar *in vitro* sin la necesidad de GTP (Díaz-Espinoza y cols. 2007). Esta mutación produce una interacción aromática entre el triptófano 294 y la fenilalanina 275, lo cual disminuiría la flexibilidad del dominio carboxilo. Suponiendo que la polimerización depende del estado conformacional del monómero, entonces el efecto de esta mutante sería modificar la probabilidad de transiciones estructurales asociadas a la polimerización.

En segundo lugar, PaFtsZ pierde estructura secundaria antes que MjFtsZ a todas las temperaturas (Fig. 11 y 12), siendo más afectado el dominio carboxilo. Al analizar las diferencias de estabilidad entre estas proteínas, se observa que una de las mayores diferencias del desplegamiento es la aumentada estabilidad de la hélice H7 de MjFtsZ, sugiriendo que la interfase juega un rol preponderante en el mecanismo de estabilización de MjFtsZ. En efecto, análisis de puentes salinos muestran una mayor red de interacciones electrostáticas entre la hélice H7 y el dominio amino (Mobarec 2005) sugiriendo que las zonas donde se producen estabilizaciones de puentes salinos es también un tema relevante en la estabilización, siendo más efectivos los cambios que afectan a zonas claves en el proceso de plegamiento (en este caso la interfase). Estas observaciones son consistentes con la pérdida de contactos nativos de los núcleos hidrófobos, donde el dominio carboxilo de PaFtsZ pierde estructura más rápido que el dominio amino, y el núcleo de la interfase de MjFtsZ presenta una estabilidad aumentada a 450 K (Fig. 17).

Adicionalmente a los factores antes mencionados, es posible derivar de las curvas de RMSD a distintas temperaturas donde MjFtsZ presenta menores

fluctuaciones en sus curvas de desplegamiento (Fig. 9A y 10A) y los elementos de estructura secundaria de MjFtsZ muestran un desplegamiento más cooperativo (Figs. 11 y 12), una visible similitud en el comportamiento del desplegamiento de los elementos de estructura secundaria dentro de cada dominio.

### **Influencia de la interfase de interacción en la estabilidad de los dominios**

Luego de estudiar las diferencias del proceso de desplegamiento de MjFtsZ y PaFtsZ, pudimos utilizar las matrices  $R$  de estas simulaciones para compararlas con matrices  $R$  obtenidas a partir de simulaciones de los dominios aislados. Los resultados indican que a 300 K, los dominios aislados no muestran diferencias significativas en su estabilidad. Esto no debe interpretarse, sin embargo, como que los dominios de ambas proteínas tienen la misma estabilidad después de la disociación, puesto que las diferencias podrían ser más evidentes con mayores tiempos de simulación. El aumento de la temperatura debería afectar en forma distinta a proteínas con diferencias en estabilidad, lo cual es evidente al analizar los perfiles de valores  $S_{1/2}^m - S_{1/2}^{dom}$  (Fig. 13 y 14), de los cuales se puede concluir que la separación de los dominios desestabiliza la estructura secundaria, donde el dominio carboxilo es el más afectado. Tomando en cuenta los resultados de la sección anterior, esta diferencia es esperable en base a las diferencias de estabilidad intrínseca de los dominios que son evidentes en las simulaciones de las proteínas completas. Sin embargo, esta desestabilización no es tan evidente al analizar la estabilidad de los núcleos hidrófobos (Fig. 18). Esto muestra la importancia de la interfase de interacción en la estabilidad de los dominios: aunque los contactos que se pierden en la estructura se encuentran en la zona de la interfase, en la cara opuesta de donde se encuentran las hélices  $\alpha$ , su estructura es la más afectada. Esto tiene completo sentido en la medida que se analiza el formalismo para simular las proteínas: donde se la representa como una red densa de interacciones no lineales, por lo tanto, aun cuando la acción de las modificaciones no afecte directamente alguna estructura, esta podrá ser afectada por la existencia de correlaciones débiles dentro de esta (Schneidman y cols. 2006). Esto también podría explicar porqué el aumento de temperatura produce una estabilización de ciertas

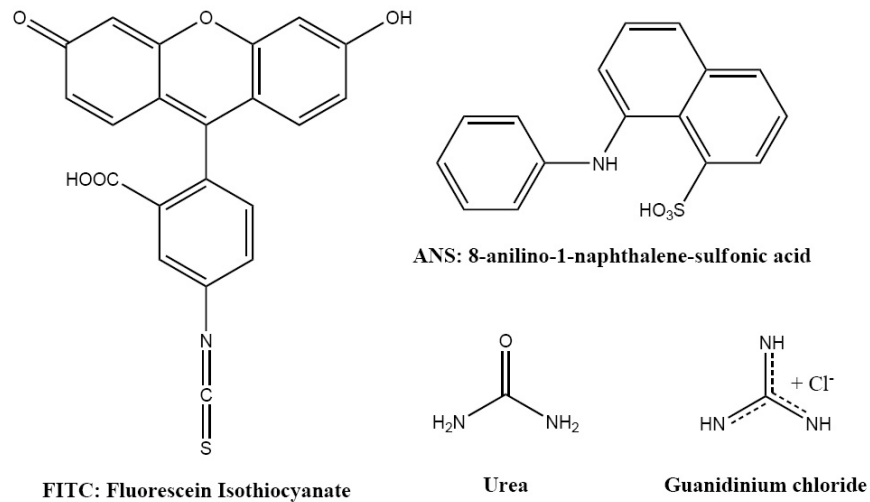
estructuras secundarias. La estructura de los dominios se reacomoda en respuesta a la pérdida de interacciones entre los dominios. Zonas que antes podían tener mayor movilidad, podrían estabilizarse en respuesta a la redistribución de las interacciones dentro de la estructura. Al analizar esta sobre estabilización, es posible ver que es más notoria en el dominio amino, esto puede deberse a que su estructura posee más formas de redistribuir las interacciones perdidas, en este caso un sitio de unión a nucleótido.

### **Probabilidad de la existencia de intermediarios en el plegamiento.**

Por medio de estudios de desplegamiento de MjFtsZ y EcFtsZ con GdmCl y dicroísmo circular, pudo verificarse que la vía de desplegamiento de MjFtsZ presenta una transición de dos estados, mientras que la de EcFtsZ presenta dos transiciones, estando la primera asociada a la salida del nucleótido (Andreu y cols. 2002). A diferencia del comportamiento observado con GdmCl, estudios de desplegamiento de EcFtsZ con urea muestran solo una transición cuando son seguidos por dicroísmo circular, mientras que sondas fluorescentes como ANS y FITC muestran la presencia de intermediarios, los cuales pueden o no corresponder a los observados tras el desplegamiento con GdmCl (Santra y Panda 2003).

Es común que los resultados de análisis de plegamiento utilizando agentes caotrópicos químicos muestren distintos comportamientos. Los resultados de estos experimentos pueden explicarse tomando en cuenta la naturaleza de los agentes utilizados (Fig. 22). Por ejemplo, en el plegamiento de NADPH-ferrodoxin reductasa, las curvas de desplegamiento por urea utilizando dicroísmo circular, intensidad de fluorescencia y cambio en el máximo de emisión de triptófano son completamente superponibles, mientras que las curvas de desnaturación con GdmCl cambian con cada uno de los métodos utilizados (Shukla y cols. 2005). La gran diferencia entre GdmCl y urea es que la primera es una sal, mientras que la segunda es una molécula neutra. Urea favorece el desplegamiento en forma indirecta modificando las propiedades del agua, favoreciendo la solvatación de grupos hidrófobos y directamente uniéndose a grupos polares y estabilizando el estado desplegado (Bennion y Daggett

2003; Zou y cols. 2002). Urea interactúa con la mayoría de las zonas hidrofílicas de la proteína, mientras que GdmCl podría tener interacciones favorables con residuos ácidos principalmente, lo cual muestra que urea se une en forma más general a la proteína, mientras que GdmCl lo haría en forma específica, lo mismo que ocurriría con FITC y ANS, las cuales se unen a zonas apolares. Tomando esto en consideración es posible que GdmCl establezca algún intermediario basado en cambios estructurales locales.

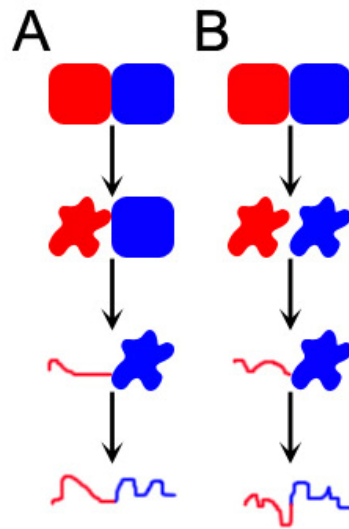


**Figura 22. Agentes caotrópicos y sondas fluorescentes utilizados en estudios de desplegamiento de FtsZ.**

Las simulaciones de PaFtsZ y MjFtsZ tanto a 300 K como a las temperaturas de desplegamiento mostraron que existe una marcada diferencia en el comportamiento del fragmento comprendido entre la hélice H1 y H3. En PaFtsZ, esta zona es una de las primeras en desplegarse, llegando a alejarse del centro de la estructura a 400 y 450 K. Aunque esta zona es altamente móvil en MjFtsZ, H2 es más estable, y H3 no se despliega en su primera porción como ocurre en el caso PaFtsZ. Experimentalmente se ha demostrado que es posible purificar MjFtsZ en ausencia de ligando, y polimerizar en estas condiciones (Huecas y Andreu 2004). Por el contrario, EcFtsZ, la cual consideramos similar a PaFtsZ, requiere de la presencia de ligando para su purificación, lo que sugiere que la proteína posiblemente se denatura en ausencia de ligando. En virtud de estos datos, sugerimos que gran parte de la inestabilidad de

EcFtsZ en solución se debería a la inestabilidad del fragmento H1-H3 en la ausencia de ligando.

¿Es esta zona la misma que genera intermediarios en desplegamiento con urea? posiblemente no. En efecto, al seguir el desplegamiento con urea por dicroísmo circular no se observan intermediarios, sugiriendo que ANS y FITC estabilizan otras conformaciones. Sin embargo datos de microcalorimetría diferencial de barrido muestra al menos tres transiciones, evidenciando tres intermediarios, lo que se espera para una proteína tan grande como FtsZ. En efecto, se ha demostrado que estudios de equilibrio por si solos no son capaces de evidenciar la presencia de intermediarios, siendo necesarios estudios cinéticos para descartar tales posibilidades. Recientemente se ha descrito que interfases estrechas entre dominios pueden enmascarar posibles intermediarios en estudios de equilibrio (Batey y Clarke 2006). Tomando esto en consideración todos los datos disponibles, proponemos un modelo general para el desplegamiento de FtsZ (Fig. 23). El desplegamiento cooperativo de FtsZ ocurre con presencia de intermediarios, que solo son vistos por microcalorimetría diferencial y eventualmente por posibles experimentos cinéticos. En este modelo la estabilidad de las estructuras estaría dada por Carboxilo < Interfase < Amino, siendo el dominio amino el primero en plegarse. Ahora bien, al tomar en cuenta el nivel de cooperatividad de la interfase adicionalmente con la estabilidad similar de los dominios hace que el modelo de la Figura 23A sea un extremo; no se espera encontrar al dominio carboxilo desplegado mientras que el amino se encuentra plegado, sino que el plegamiento del dominio carboxilo seguiría cercanamente al del dominio amino. Esto explicaría la aparente cooperatividad de FtsZ. MjFtsZ y otras variantes termófilas podrían ajustarse más a este modelo. Por otra parte, el desplegamiento de PaFtsZ y otras variantes mésófilas seguiría el mismo patrón general, pero con la diferencia que el desplegamiento en presencia de GdmCl mostraría un intermediario asociado a la inestabilidad del fragmento H1-H3 en el dominio amino. Por otra parte, otros intermediarios se podrían observar entre estas etapas, como lo visto por fluorescencia intrínseca de triptófanos que muestra que el desplegamiento del dominio amino es distinto al usar sondas fluorescentes en distintas partes del núcleo hidrófobo.



**Figura 23. Modelo de desplegamiento de FtsZ basado en simulaciones y datos experimentales. Panel A.** Desplegamiento sugerido para proteínas termófilas, **Panel B.** Desplegamiento sugerido para proteínas mesófilas. En rojo dominio carboxilo y en azul dominio amino.



## **CONCLUSIONES**

En esta tesis fuimos capaces de utilizar un protocolo de dinámica molecular con solvente implícito (EEF1) para simular el proceso de desplegamiento térmico de proteínas mesófilas y termófilas de FtsZ. Nuestras pruebas de simulación con la proteína Cl2 muestran que el método utilizado es consistente con los datos en la literatura, con lo cual la primera hipótesis de esta memoria de título fue confirmada. Se logró cumplir con los objetivos específicos eligiendo un representante termofilo y otro mesófilo de FtsZ para llevar a cabo simulaciones de desplegamiento térmico, así como también se crearon programas escritos en los lenguajes Perl y Tcl para analizar estas simulaciones. Resultados de estos análisis mostraron que las zonas más afectadas por la separación de los dominios es la hélice H7 y la zona comprendida entre las hélices H2 y H3. No obstante, también se observó que zonas distantes a la interfase de interacción (hélices) son las más sensibles a la pérdida de interfase de interacción, siendo la estructura secundaria más sensible que los núcleos hidrofóbicos. Adicionalmente, fue posible observar transiciones estructurales de orientación relativa de los dominios de FtsZ lo cual está de acuerdo con datos experimentales que sugieren una mayor flexibilidad del dominio carboxilo.

Finalmente, hemos dado evidencias consistentes por medio de métodos de dinámica molecular con solvente implícito, que los dominios de proteínas FtsZ mesófilas y termófilas se desestabilizan al eliminar la interfase de interacción entre ellas. Del mismo modo, hemos mostrado que la interfase no afecta de la misma manera a ambos dominios, observándose una mayor desestabilización del dominio carboxilo. Finalmente, esta desestabilización es más evidente en proteínas mesófilas que en proteínas termófilas.

## **PERSPECTIVAS FUTURAS**

Este trabajo entrega nuevos datos para estudios experimentales y de simulación; en primer lugar, la flexibilidad del dominio carboxilo, sugiere que utilizando simulaciones de mayor duración con solvente explícito, sería posible dilucidar por medios dinámicos si existen cambios estables en la orientación de los dominios, al

menos en PaFtsZ. En segundo lugar, la inestabilidad local de PaFtsZ en el fragmento H1-H3 sugiere que si esta zona genera el intermediario de observado en curvas de desnaturación con GdmCl de EcFtsZ, entonces la estabilización de esta zona por mutagénesis sitio-dirigida debería hacer desaparecer la primera transición en estas curvas.

## REFERENCIAS

- Aldaz H, Rice LM, Stearns T, Agard DA. 2005. Insights into microtubule nucleation from the crystal structure of human gamma-tubulin. *Nature* 435(7041):523-7.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25(17):3389-402.
- Andreu JM, Oliva MA, Monasterio O. 2002. Reversible unfolding of FtsZ cell division proteins from archaea and bacteria. Comparison with eukaryotic tubulin folding and assembly. *J Biol Chem* 277(45):43262-70.
- Anfinsen CB, Haber E. 1961. Studies on the reduction and re-formation of protein disulfide bonds. *J Biol Chem* 236:1361-3.
- Anfinsen CB, Haber E, Sela M, White FH, Jr. 1961. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci U S A* 47:1309-14.
- Armen RS, Alonso DO, Daggett V. 2004. Anatomy of an amyloidogenic intermediate: conversion of beta-sheet to alpha-sheet structure in transthyretin at acidic pH. *Structure* 12(10):1847-63.
- Batey S, Clarke J. 2006. Apparent cooperativity in the folding of multidomain proteins depends on the relative rates of folding of the constituent domains. *Proc Natl Acad Sci U S A* 103(48):18113-8.
- Beck DA, Daggett V. 2004. Methods for molecular dynamics simulations of protein folding/unfolding in solution. *Methods* 34(1):112-20.
- Bennion BJ, Daggett V. 2003. The molecular basis for the chemical denaturation of proteins by urea. *Proc Natl Acad Sci U S A* 100(9):5142-7.
- Bennion BJ, Daggett V. 2004. Counteraction of urea-induced protein denaturation by trimethylamine N-oxide: a chemical chaperone at atomic resolution. *Proc Natl Acad Sci U S A* 101(17):6433-8.
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000. The Protein Data Bank. *Nucleic Acids Res* 28(1):235-42.
- Bertrand S, Barthelemy I, Oliva MA, Carrascosa JL, Andreu JM, Valpuesta JM. 2005. Folding, stability and polymerization properties of FtsZ chimeras with inserted tubulin loops involved in the interaction with the cytosolic chaperonin CCT and in microtubule formation. *J Mol Biol* 346(1):319-30.
- Beuria TK, Krishnakumar SS, Sahar S, Singh N, Gupta K, Meshram M, Panda D. 2003. Glutamate-induced assembly of bacterial cell division protein FtsZ. *J Biol Chem* 278(6):3735-41.
- Bond CJ, Wong KB, Clarke J, Fersht AR, Daggett V. 1997. Characterization of residual structure in the thermally denatured state of barnase by simulation and experiment: description of the folding pathway. *Proc Natl Acad Sci U S A* 94(25):13409-13.
- Borhani DW, White EL. 2004. Polymerization of C-terminally truncated Mycobacterium tuberculosis FtsZ is unlikely to be physiologically relevant. *Microbiology* 150(Pt 12):3903-6.
- Bramhill D, Thompson CM. 1994. GTP-dependent polymerization of Escherichia coli FtsZ protein to form tubules. *Proc Natl Acad Sci U S A* 91(13):5813-7.

- Brooks B, Bruccoleri R, Olafson B, States D, Swaminathan S, Karplus M. 1983. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comp Chem* 4:187–217.
- Buey RM, Diaz JF, Andreu JM. 2006. The nucleotide switch of tubulin and microtubule assembly: a polymerization-driven structural change. *Biochemistry* 45(19):5933-8.
- Chakravarty S, Varadarajan R. 2002. Elucidation of factors responsible for enhanced thermal stability of proteins: a structural genomics based study. *Biochemistry* 41(25):8152-61.
- Clarke J, Hounslow AM, Bond CJ, Fersht AR, Daggett V. 2000. The effects of disulfide bonds on the denatured state of barnase. *Protein Sci* 9(12):2394-404.
- Cordell SC, Lowe J. 2004. Structural insights into FtsZ. *Nat Struct Mol Biol* 11(12):1243-50.
- Daggett V. 2006. Protein folding-simulation. *Chem Rev* 106(5):1898-916.
- Daggett V, Fersht AR. 2003. Is there a unifying mechanism for protein folding? *Trends Biochem Sci* 28(1):18-25.
- Daggett V, Li A, Itzhaki LS, Otzen DE, Fersht AR. 1996. Structure of the transition state for folding of a protein derived from experiment and simulation. *J Mol Biol* 257(2):430-40.
- Davis H, Snider A. 1992. *Análisis Vectorial*. Mexico: McGraw-Hill Interamericana.
- Day R, Bennion BJ, Ham S, Daggett V. 2002. Increasing temperature accelerates protein unfolding without changing the pathway of unfolding. *J Mol Biol* 322(1):189-203.
- Desai A, Mitchison TJ. 1997. Microtubule polymerization dynamics. *Annu Rev Cell Dev Biol* 13:83-117.
- Diaz-Espinoza R, Garces AP, Arbildua JJ, Montecinos F, Brunet JE, Lagos R, Monasterio O. 2007. Domain folding and flexibility of *Escherichia coli* FtsZ determined by tryptophan site-directed mutagenesis. *Protein Sci* 16(8):1543-56.
- Dinner AR, Karplus M. 1999. Is protein unfolding the reverse of protein folding? A lattice simulation analysis. *J Mol Biol* 292(2):403-19.
- Fan H, Mark AE. 2003. Relative stability of protein structures determined by X-ray crystallography or NMR spectroscopy: a molecular dynamics simulation study. *Proteins* 53(1):111-20.
- Fersht AR, Daggett V. 2002. Protein folding and unfolding at atomic resolution. *Cell* 108(4):573-82.
- Fersht AR, Matouschek A, Serrano L. 1992. The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J Mol Biol* 224(3):771-82.
- Frishman D, Argos P. 1995. Knowledge-based protein secondary structure assignment. *Proteins* 23(4):566-79.
- Gupta P, Anand SP, Srinivasan R, Rajeswari H, Indi S, Ajitkumar P. 2004. The C-terminally truncated MtFtsZ-DeltaC169 mutant of *Mycobacterium tuberculosis* FtsZ shows GTPase and GTP-induced, GTP-specific polymerization activities in vitro. *Microbiology* 150(Pt 12):3906-8.
- Haney PJ, Badger JH, Buldak GL, Reich CI, Woese CR, Olsen GJ. 1999. Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic *Methanococcus* species. *Proc Natl Acad Sci U S A* 96(7):3578-83.

- Hanna M. 1985. *Mecánica cuántica para Químicos*. Tlalpan: Fondo Educativo Interamericano.
- Hubbard SJ, Thornton JM. 1993. NACCESS: Department of Biochemistry and Molecular Biology, University College London.
- Huecas S, Andreu JM. 2004. Polymerization of nucleotide-free, GDP- and GTP-bound cell division protein FtsZ: GDP makes the difference. *FEBS Lett* 569(1-3):43-8.
- Humphrey W, Dalke A, Schulten K. 1996. VMD: visual molecular dynamics. *J Mol Graph* 14(1):33-8, 27-8.
- Itzhaki LS, Otzen DE, Fersht AR. 1995. The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *J Mol Biol* 254(2):260-88.
- Jorgensen WL, Tirado-Rives J. 1988. The OPLS Force Field for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* 110:1657-1666.
- Kazmirski SL, Daggett V. 1998. Non-native interactions in protein folding intermediates: molecular dynamics simulations of hen lysozyme. *J Mol Biol* 284(3):793-806.
- Kumar S, Tsai CJ, Nussinov R. 2000. Factors enhancing protein thermostability. *Protein Eng* 13(3):179-91.
- Kuwajima K. 1977. A folding model of alpha-lactalbumin deduced from the three-state denaturation mechanism. *J Mol Biol* 114(2):241-58.
- Lazaridis T, Karplus M. 1997. "New view" of protein folding reconciled with the old through multiple unfolding simulations. *Science* 278(5345):1928-31.
- Lazaridis T, Karplus M. 1999. Effective energy function for proteins in solution. *Proteins* 35(2):133-52.
- Lazaridis T, Lee I, Karplus M. 1997. Dynamics and unfolding pathways of a hyperthermophilic and a mesophilic rubredoxin. *Protein Sci* 6(12):2589-605.
- Leach AR. 2001. *Molecular modelling : principles and applications*. Harlow, England ; New York: Prentice Hall. xxiv, 744 , [16] of plates p.
- Leung A, Cordell SC, Lowe J. 2004a. Structural insights into FtsZ protofilament formation. *J Mol Biol* 11(3):953-70.
- Leung AK, Lucile White E, Ross LJ, Reynolds RC, DeVito JA, Borhani DW. 2004b. Structure of Mycobacterium tuberculosis FtsZ reveals unexpected, G protein-like conformational switches. *J Mol Biol* 342(3):953-70.
- Levitt M, Hirschberg M, Sharon R, Daggett V. 1995. Potential energy function and parameters for simulations of the molecular dynamics of proteins and nucleic acids in solution. *Computer Physics Communications* 91:215–231.
- Li A, Daggett V. 1994. Characterization of the transition state of protein unfolding by use of molecular dynamics: chymotrypsin inhibitor 2. *Proc Natl Acad Sci U S A* 91(22):10430-4.
- Li A, Daggett V. 1996. Identification and characterization of the unfolding transition state of chymotrypsin inhibitor 2 by molecular dynamics simulations. *J Mol Biol* 257(2):412-29.
- Li Z, Trimble MJ, Brun YV, Jensen GJ. 2007. The structure of FtsZ filaments in vivo suggests a force-generating role in cell division. *Embo J* 26(22):4694-708.
- Lowe J. 1998. Crystal structure determination of FtsZ from *Methanococcus jannaschii*. *J Struct Biol* 124(2-3):235-43.

- Lowe J, Amos LA. 1999. Tubulin-like protofilaments in Ca<sup>2+</sup>-induced FtsZ sheets. *Embo J* 18(9):2364-71.
- MackKerell J, A. D. , Brooks B, Brooks III CL, Nilsson L, Roux B, Won Y, Karplus M. 1998. CHARMM: The Energy Function and Its Parameterization with an Overview of the Program. Schleyer PvR, editor. Chichester: John Wiley & Sons.
- Mobarec J. 2005. Diseño, construcción y caracterización estructural y funcional de las mutantes de FtsZ de *Escherichia coli* Q194K y N328R: Residuos que forman redes de puentes salinos en FtsZ de *Methanococcus jannaschii*. [Tesis de Ingeniero en biotecnología molecular]. Santiago: Universidad de Chile.
- Mukherjee A, Lutkenhaus J. 1994. Guanine nucleotide-dependent assembly of FtsZ into filaments. *J Bacteriol* 176(9):2754-8.
- Nogales E, Downing KH, Amos LA, Lowe J. 1998a. Tubulin and FtsZ form a distinct family of GTPases. *Nat Struct Biol* 5(6):451-8.
- Nogales E, Wang HW. 2006. Structural mechanisms underlying nucleotide-dependent self-assembly of tubulin and its relatives. *Curr Opin Struct Biol* 16(2):221-9.
- Nogales E, Wolf SG, Downing KH. 1998b. Structure of the alpha beta tubulin dimer by electron crystallography. *Nature* 391(6663):199-203.
- Oliva MA, Cordell SC, Lowe J. 2004. Structural insights into FtsZ protofilament formation. *Nat Struct Mol Biol* 11(12):1243-50.
- Oliva MA, Trambaiolo D, Lowe J. 2007. Structural insights into the conformational variability of FtsZ. *J Mol Biol* 373(5):1229-42.
- Otzen DE, Itzhaki LS, elMasry NF, Jackson SE, Fersht AR. 1994. Structure of the transition state for the folding/unfolding of the barley chymotrypsin inhibitor 2 and its implications for mechanisms of protein folding. *Proc Natl Acad Sci U S A* 91(22):10422-5.
- Ponder JW, Case DA. 2003a. Force fields for protein simulations. *Adv. Prot. Chem.* 66:27-85.
- Ponder W, Case DA. 2003b. Force fields for protein simulations. *Adv Prot Chem* 66:27-85.
- Ptitsyn OB. 1995. Molten globule and protein folding. *Adv Protein Chem* 47:83-229.
- Rhee YM, Sorin EJ, Jayachandran G, Lindahl E, Pande VS. 2004. Simulations of the role of water in the protein-folding mechanism. *Proc Natl Acad Sci U S A* 101(17):6456-61.
- Romberg L, Levin PA. 2003. Assembly dynamics of the bacterial cell division protein FtsZ: poised at the edge of stability. *Annu Rev Microbiol* 57:125-54.
- Rothfield LI, Justice SS. 1997. Bacterial cell division: the cycle of the ring. *Cell* 88(5):581-4.
- Rothfield LI, Justice SS, J. G-L. 1999. Bacterial cell division. *Annu Rev Genet* 33:423-448.
- Russell RB, Barton GJ. 1992. Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. *Proteins* 14(2):309-23.
- Sali A, Shakhnovich E, Karplus M. 1994. How does a protein fold? *Nature* 369(6477):248-51.
- Santra MK, Panda D. 2003. Detection of an intermediate during unfolding of bacterial cell division protein FtsZ: loss of functional properties precedes the global unfolding of FtsZ. *J Biol Chem* 278(24):21336-43.

- Schneidman E, Berry II M, Segev R, Bialek W. 2006. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* 440(7087):1007-1012.
- Schuler LD, Daura X, van Gunsteren WF. 2001. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *Journal of Computational Chemistry* 22(11):1205-1218.
- Sham YY, Ma B, Tsai CJ, Nussinov R. 2002. Thermal unfolding molecular dynamics simulation of Escherichia coli dihydrofolate reductase: thermal stability of protein domains and unfolding pathway. *Proteins* 46(3):308-20.
- Shea J, Brooks III C. 2001. From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. *Annu Rev Phys Chem*:499-535.
- Shukla N, Bhatt AN, Aliverti A, Zanetti G, Bhakuni V. 2005. Guanidinium chloride- and urea-induced unfolding of FprA, a mycobacterium NADPH-ferredoxin reductase: stabilization of an apo-protein by GdmCl. *Febs J* 272(9):2216-24.
- Snow C, Sorin E, Rhee Y, Pande V. 2005. How well can simulation predict protein folding kinetics and thermodynamics. *Annu Rev Biophys Biomol Struct* 34:43-69.
- Snow CD, Qiu L, Du D, Gai F, Hagen SJ, Pande VS. 2004. Trp zipper folding kinetics by molecular dynamics and temperature-jump spectroscopy. *Proc Natl Acad Sci U S A* 101(12):4077-82.
- Stricker J, Maddox P, Salmon ED, Erickson HP. 2002. Rapid assembly dynamics of the Escherichia coli FtsZ-ring demonstrated by fluorescence recovery after photobleaching. *Proc Natl Acad Sci U S A* 99(5):3171-5.
- Vaughan S, Wickstead B, Gull K, Addinall SG. 2004. Molecular evolution of FtsZ protein sequences encoded within the genomes of archaea, bacteria, and eukaryota. *J Mol Evol* 58(1):19-29.
- Weber W, Hünenberger PH, McCammon JA. 2000. Molecular dynamics simulations of a polyalanine octapeptide under ewald boundary conditions: influence of artificial periodicity on peptide conformation. *J. Phys. Chem. B* 104:3668-3675.
- Zou Q, Bennion BJ, Daggett V, Murphy KP. 2002. The molecular mechanism of stabilization of proteins by TMAO and its ability to counteract the effects of urea. *J Am Chem Soc* 124(7):1192-202.

## ANEXO I

### Elementos de Dinámica Molecular

A través de métodos experimentales es posible estudiar muchas propiedades fisicoquímicas de las macromoléculas. Por un método como geles de electroforesis en condiciones desnaturalantes es posible separar proteínas según su peso molecular; si logramos sobre expresarla y purificarla, es posible obtener datos como el radio hidrodinámico utilizando ultracentrifugación; estudiar la vía de desplegamiento de equilibrio utilizando agentes caotrópicos como cloruro de guanidinio o urea y la vía cinética usando métodos de mezcla rápida; calcular propiedades termodinámicas tales como la capacidad calorífica y entalpía utilizando microcalorimetría diferencial, y hasta obtener una idea de cuales son los hidrógenos que son lábiles al solvente utilizando intercambio isotópico con agua deuterada y NMR. Todos los observables experimentales que utilizaríamos para medir estos fenómenos, generalmente por métodos espectrofotométricos, lo que nos entregan es una medida estadística, la suma de muchas contribuciones individuales provenientes de muchas moléculas de proteína, que sumadas, dan sentido al observable *macroscópico*. Pero ¿Qué puedo hacer si quisiera entender porqué una mutación particular en una proteína de mi interés disminuye la capacidad para unir ligando, y las propiedades que puedo medir en el colectivo de moléculas no tienen cambios? ¿o si quisiera conocer los cambios estructurales que ocurren en el proceso de desplegamiento de una proteína que no son obvios a partir de experimentos con muchas moléculas?. Muchas preguntas como éstas, que implican el conocimiento de detalles *microscópicos* de la molécula en cuestión, pueden ser abordadas mediante dinámica molecular. Es cierto que es posible obtener datos microscópicos de las proteínas; inclusive obtener modelos de la disposición espacial de sus átomos por medio de cristalografía de rayos X y NMR, pero ellas son visiones más bien estáticas, no muestran el comportamiento de estas moléculas en el tiempo. Esta visión es la que podemos obtener utilizando dinámica molecular.

Este escrito tiene como objetivo introducir al lector lego a la terminología y conceptos básicos de las metodologías de dinámica molecular utilizadas en esta tesis. Por esta razón, ciertos temas, como el tratamiento de solvatación explícita, no son tratados puesto que esta tesis no las requiere. La presentación es concisa: se sugiere al lector interesado leer el libro *Molecular Modelling: Principles and Applications* (Leach 2001).

#### I. El campo de fuerza

Como habíamos visto anteriormente, es posible obtener datos microscópicos estáticos de las proteínas. Estos datos experimentales son de vital importancia para las simulaciones de dinámica molecular, corresponden a las *coordenadas iniciales* de nuestra molécula, la cual experimentara cambios durante la simulación. Pero lo primero que debemos responder es ¿cómo a partir de esa molécula vamos a simular movimientos? ¿cómo calcularemos la trayectoria de los átomos?. Tenemos que para



una partícula con posición  $r$ , es posible conocer su trayectoria si se conoce la dependencia de su velocidad  $v$  y aceleración  $a$  respecto del tiempo:

$$r(t) = r_0 + v(t) \cdot t + \frac{a(t) \cdot t^2}{2}$$

Para el caso de una proteína conocemos  $r_0$  (una estructura cristalográfica, las coordenadas iniciales) y la velocidad inicial puede asignarse en base a ciertas suposiciones (eso lo veremos a continuación). Lo único que necesitamos es conocer las fuerzas que actúan sobre el sistema para deducir la aceleración de las partículas. Lo que requerimos encontrar es entonces el *campo vectorial*  $F$  de las fuerzas del sistema, una función vectorial que entregue las componentes de fuerza aplicada sobre todas las partículas. El principal teorema del cálculo vectorial dice que para algunos casos es posible derivar este campo vectorial a partir de un *campo escalar*  $E$ , cumpliéndose que

$$F(r) = -\frac{\partial E(r)}{\partial r} = -\text{grad}(E(r))$$

Si esto es posible, se dice que  $F(r)$  es un campo vectorial *conservativo*, y  $E(r)$  es una *función potencial* (Davis y Snider 1992). En el caso de la dinámica molecular, esta función potencial se denomina *campo de fuerza*. Para la dinámica de las partículas, esta corresponde a una función de *energía potencial*, es decir una función que relaciona la configuración del sistema de partículas con su energía total.

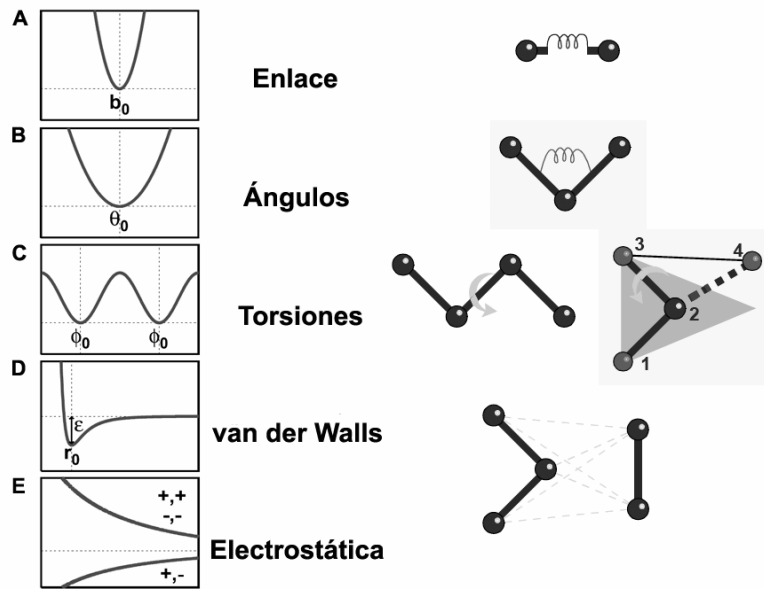
Una descripción detallada de la energía de una molécula requiere métodos de mecánica cuántica. Bajo este formalismo, la energía de una molécula depende tanto de la configuración espacial de los núcleos como de la densidad de probabilidad electrónica. Aunque se dispone de un nivel de teoría elevado para describir las interacciones moleculares, este no es capaz de resolver la energía de la molécula tan simple como  $\text{H}_2^+$ , puesto que este sistema tiene tres cuerpos, y se ha demostrado que un sistema dinámico de tres cuerpos es sensible a la representación numérica, haciendo imposible la predicción de su comportamiento a tiempos largos (Hanna 1985). Para solucionar este problema, y resolver las ecuaciones de mecánica cuántica, se utiliza la aproximación de *Born-Oppenheimer*, según la cual es posible calcular las propiedades electrónicas de las moléculas dejando los núcleos fijos, pues el movimiento de los electrones es mucho más rápido que el de los núcleos, pudiendo ajustarse casi instantáneamente a cambios de conformación. Este desacoplamiento entre núcleo y electrón, nos permite estudiar la dinámica de los átomos suponiendo que sus propiedades electrónicas no cambian.

En ausencia de electrones, es necesario recurrir a funciones empíricas que describan la energía del sistema en base a las posiciones de los átomos. Estas funciones deben reproducir lo más fielmente posible el comportamiento cuántico, siendo a la vez suficientemente simples para no requerir un esfuerzo computacional elevado. Los campos de fuerza para simulación proteínas están compuestos por dos

componentes, una que modela las *interacciones enlazantes*, funciones que permiten mantener la geometría local de las moléculas, y las *interacciones no enlazantes*, las cuales modelan la interacción entre átomos que se encuentran separados por más de dos átomos.

La primera de las interacciones enlazantes modela el comportamiento vibracional de los enlaces. Se considera al enlace atómico como un resorte con cierta constante de fuerza  $K_0$  y distancia de referencia  $b_0$  (Fig. 1A). Con esta componente, es posible simular vibraciones de *tensión*, como las encontradas en moléculas diatómicas ( $N_2$ ) o poliatómicas lineales ( $CO_2$ ). La segunda modela vibraciones de *flexión* entre tres átomos por medio de un potencial armónico asociada con el ángulo descrito por los enlaces unidos al átomo central (Fig. 1B). Con esta interacción, es posible modelar la geometría  $sp^3$  ( $CH_4$ ). Estas dos componentes, se consideran grados de libertad *duros*, pues se requiere de energías elevadas para generar desviaciones significativas de las posiciones o ángulos de referencia. La tercera componente enlazante modela las barreras energéticas existentes al realizar *torsiones* a lo largo de un enlace (Fig. 1C). Se usa una función cosenoidal para representar el perfil de rotación entre dos grupos  $sp^3$ , lo cual da mínimos energéticos en conformaciones *gauche* ( $60^\circ$ ,  $-60^\circ$ ) y *trans* ( $180^\circ$ ) o bien el perfil de rotación de un enlace  $sp^2$ , que tiene mínimos en conformaciones *cis* ( $0^\circ$ ) y *trans* ( $180^\circ$ ). Con las componentes hasta ahora descritas es posible modelar el comportamiento de una molécula de etano ( $CH_3CH_3$ ) o acetamida ( $CH_3CONH$ ).

Las interacciones entre átomos que se encuentran a más de dos átomos de distancia depende principalmente de sus propiedades electrónicas. Como dijimos anteriormente, en dinámica molecular no consideramos la presencia de electrones en forma explícita, sino que consideramos que las propiedades electrónicas de los átomos permanecen fijas. De este modo, una de las propiedades no consideradas en los campos de fuerza es la fluctuación de las nubes electrónicas debido cambios instantáneos en la distribución de electrones. Estas fluctuaciones generan dipolos inducidos en los átomos generando fuerzas de atracción inespecíficas, llamadas *fuerzas de dispersión* o *fuerzas de London*. Cuantitativamente, se ha observado que estas interacciones disminuyen con el inverso de la sexta potencia de la distancia. Una función de energía que modela este comportamiento es el potencial de *Lennard-Jones* (Fig. 1D), la cual adicionalmente modela la repulsión de los núcleos atómicos a cortas distancias por medio de un potencial que aumenta con el inverso de la doceava potencia de la distancia. La unión de estos dos comportamientos da lugar a lo que se denota como *fuerzas de van der Waals*. Por otra parte, al unir dos átomos con electronegatividades diferentes se produce una redistribución de la nube electrónica, lo que genera un dipolo eléctrico. La forma en que el campo de fuerza modela esta tendencia es asignando *cargas parciales* a cada uno de los átomos, y calculando su interacción en base a la ley de Coulomb (Fig. 1E).



F

$$f = - \left( \frac{\partial V}{\partial x} \right)$$

$$V = \sum_i^{\text{enlaces}} K_{b,i} (b_i - b_{0,i})^2 + \sum_i^{\text{ángulos}} K_{\theta,i} (\theta_i - \theta_{0,i})^2$$

$$+ \sum_i^{\text{torsiones}} K_{\phi,i} \{1 - \cos[n_i(\phi_i - \phi_{0,i})]\} + U_{\text{nb}}$$

$$U_{\text{nb}} = \sum_{\text{pares } ij} \left[ \epsilon_{ij} \left( \frac{r_{0,ij}}{r_{ij}} \right)^{12} - 2\epsilon_{ij} \left( \frac{r_{0,ij}}{r_{ij}} \right)^6 \right] + 332 \sum_{\text{pares } ij} \left( \frac{q_i q_j}{r_{ij}} \right).$$

**Figura 1:** Ejemplo de función de energía potencial utilizada para la simulación de proteínas. **Paneles A a E.** gráficos de energía y dibujos esquemáticos de las componentes de un campo de fuerza, **Paneles A y B.** vibraciones de tensión y flexión modelados como un oscilador armónico simple, **Panel C.** Las barreras energéticas para el eclipsamiento entre átomos en posición 1 y 4 se modelan con un potencial cosenoidal, **Panel D.** La interacción dispersiva de London y la repulsión de los átomos se modela con un potencial de Lennard-Jones, **Panel E.** La distribución asimétrica de electrones sobre átomos con diferencias de electronegatividad se modela asignando cargas parciales a los átomos, y calculando su interacción según la ley de coulomb, **Panel F.** Forma funcional del campo de fuerza ENCAD para la simulación de proteínas. Se separan las componentes enlazantes de las no enlazantes. Formulas y dibujos adaptados de (Beck y Daggett 2004)

Los potenciales torsionales, junto con las interacciones de van der Walls y electrostáticas constituyen los grados de libertad blandos, aquellos que se encuentran más relacionados a las propiedades conformacionales de las proteínas. Una *forma funcional* de un campo de fuerza para proteínas que contempla las interacciones antes

descritas se muestra en la figura 1F. Un campo de fuerza, además de la forma funcional, requiere un conjunto de *parámetros*, como las constantes de fuerza de los osciladores, las distancias y ángulos de referencia y las cargas parciales de los átomos. Para no generar un número enorme de ellos, se definen *tipos atómicos*, que corresponden a generalizaciones de átomos en base a sus propiedades locales de interacción. Por ejemplo, los parámetros para átomos carbono involucrados en grupos metileno (-CH<sub>2</sub>-) son los mismos para un residuo de valina o isoleucina.

Dependiendo de la forma funcional del campo de fuerza y sus parámetros, se tienen distintos campos de fuerza. Los más utilizados para la simulación de proteínas, lípidos y ácidos nucleicos son AMBER (Ponder y Case 2003a), CHARMM (MacKerell y cols. 1998), OPLS (Jorgensen y Tirado-Rives 1988) y GROMOS (Schuler y cols. 2001).

La forma funcional de estos campos de fuerza es la más usada para simulación de proteínas debido a que su relación eficiencia-exactitud es suficientemente buena para llevar a cabo simulaciones de sistemas con muchas partículas. Al hacer más detallado el modelo, es posible reproducir mayor número de propiedades de las moléculas. Por ejemplo, el uso de *términos cruzados*, potenciales enlazantes que correlacionan en forma explícita vibraciones y torsiones, permiten modelar mucho mejor las propiedades vibracionales de moléculas pequeñas; las vibraciones enlazantes pueden ser modeladas mejor con un potencial anharmónico como el potencial de *Morse*, o bien utilizando una expansión en serie de potencias; las interacciones de van der Waals son mejor modeladas con un potencial donde la componente repulsiva sea exponencial, como lo que se usa en el potencial de *Buckingham*; así como también es posible representar de mejor manera el comportamiento electrostático de las moléculas si generamos varios átomos fantasma con carga en vez de limitarnos a cargas parciales atómicas.

Si bien los campos de fuerza actuales son suficientemente detallados para reproducir propiedades energéticas y conformacionales de proteínas, es posible que el tamaño del sistema a modelar haga su estudio computacionalmente costoso. En estos casos, es posible simplificar aún más la descripción de nuestro sistema no simulando hidrógenos alifáticos, y considerando de este modo los grupos metileno y metilo como grandes átomos. Estos formalismos de *átomos unidos (united-atom)* permiten modelar sistemas más grandes con el mismo costo computacional. En estos casos, la geometría se mantiene utilizando un mayor número de potenciales torsionales. Para simular sistemas aún más grandes, el nivel de resolución se puede disminuir aun más, produciendo los llamados *modelos gruesos (coarse-grained models)* donde una cadena lateral de aminoácido se modela con un solo átomo, hasta llegar a casos donde los "átomos" representan solo la forma de la proteína y no pueden asociarse a grupos de átomos.

## II. Minimización

En la sección anterior describimos los elementos básicos de un campo de fuerza. Este campo de fuerza, como un campo escalar, implica una superficie de energía potencial que no es posible visualizar debido al gran número de grados de

libertad involucrados. Sin embargo, los elementos básicos que encontramos en una superficie en tres dimensiones se pueden aplicar a dimensiones superiores: encontraremos zonas de baja energía o mínimos, así como también máximos y *puntos de bifurcación o de silla (saddle points)*. Los máximos corresponden a estructuras de *transición* entre dos o más mínimos y un mínimo corresponde a un estado estable del sistema (en ausencia de otras fuerzas externas). En la búsqueda de conformaciones de baja energía son necesarios algoritmos de optimización que permitan encontrar estos mínimos.

Sin embargo existen razones de tipo práctico que hacen necesarios los algoritmos de minimización. Recordemos que las coordenadas iniciales de nuestro sistema se derivan de una estructura cristalográfica, que corresponde a un modelo obtenido a partir de la difracción de el promedio de las moléculas de un cristal. De este modo, las posiciones de los átomos en el archivo PDB no corresponden a una sola molécula, sino a un ajuste sobre una densidad electrónica. Tomando eso en cuenta, y que el campo de fuerza es una aproximación al comportamiento cuántico, es esperable que la estructura del cristal no corresponda a una estructura de baja energía. Tomando en cuenta el tipo de interacciones involucradas en el campo de fuerza, esto implica que habrá grandes fuerzas internas, que pueden hacer que el sistema se desestabilice en una simulación de dinámica molecular.

Los métodos de minimización de energía más usados para dinámica molecular son aquellos que utilizan información de la forma de la superficie de energía potencial, en base a las primeras derivadas y segundas derivadas del campo de fuerza. El algoritmo más simple que usa la primera derivada es el método del descenso rápido (*steepest descent*). En cada iteración  $k$  del algoritmo, se calcula la dirección  $s$  opuesta al gradiente  $\mathbf{g}$ :

$$\mathbf{s}_k = -\mathbf{g}_k / |\mathbf{g}_k|.$$

Moviéndose a lo largo de la línea definida por el vector  $s$ , se busca el punto de mínima energía, y en este punto se repite el procedimiento. Al moverse en la dirección opuesta del gradiente, la energía disminuye rápidamente, por ello, este algoritmo es eficiente para optimizar configuraciones de alta energía. Sin embargo, al calcular la nueva dirección siempre se obtiene una perpendicular a la anterior, no importando la forma del potencial local. Este comportamiento hace que las soluciones oscilen alrededor del mínimo si este es muy amplio. Para evitar este comportamiento de oscilación cerca del mínimo, se puede utilizar un algoritmo de *gradiente conjugado*. En este método, el primer paso es igual al del método *steepest descent*, sin embargo la nueva dirección no se hace perpendicular a la anterior, sino que se sopesa con la dirección anterior:

$$\mathbf{s}_k = -\mathbf{g}_k + \frac{\mathbf{g}_k \cdot \mathbf{g}_k}{\mathbf{g}_{k-1} \cdot \mathbf{g}_{k-1}} \mathbf{s}_{k-1}$$

mientras mayor sea el nuevo gradiente, mayor valor tendrá el peso de la dirección anterior, de este modo se evita perder lo ganado en pasos anteriores. Esta fórmula

permite que los gradientes en cada paso sean ortogonales a todos los anteriores, lo que permite alcanzar el mínimo de una función de  $N$  variables en  $N$  pasos. El tercer método de minimización usado es el de *Newton-Raphson*. En este se utiliza información de la segunda derivada (curvatura) de la función para llegar al mínimo. En este caso los pasos de la minimización están dados por:

$$r_{k+1} = r_k - E'(r)E''(r)^{-1}$$

donde

$$E'(r) = \begin{pmatrix} \frac{\partial E(r)}{\partial x_1} \\ \frac{\partial E(r)}{\partial x_2} \\ \dots \\ \frac{\partial E(r)}{\partial x_n} \end{pmatrix}; E''(r) = \begin{pmatrix} \frac{\partial^2 E(r)}{\partial x_1 \partial x_1} & \dots & \frac{\partial^2 E(r)}{\partial x_1 \partial x_n} \\ \dots & \dots & \dots \\ \frac{\partial^2 E(r)}{\partial x_n \partial x_1} & \dots & \frac{\partial^2 E(r)}{\partial x_n \partial x_n} \end{pmatrix}$$

$E''(r)$  corresponde a la matriz de fuerza o *Hessiano* del sistema. Como para cada paso del algoritmo es necesario invertir la matriz *Hessiana*, este algoritmo es muy costoso computacionalmente.

Los tres algoritmos antes mencionados son complementarios. Por ejemplo, por su simplicidad y eficiencia, steepest descent es utilizado en etapas iniciales, puesto que rápidamente mejora estructuras altamente tensionadas. Como este algoritmo es poco eficiente cerca del mínimo de energía, es útil utilizar un protocolo de gradiente conjugado o *Newton-Raphson*. Si es requerido hacer una minimización muy prolija, entonces la mejor opción es utilizar *Newton-Raphson*. Sin embargo sus diferencias, los tres tipos de protocolos pueden ser utilizados para minimizar estructuras. La diferencia entre ellas será que algoritmos serán más o menos eficientes en ciertas etapas de la minimización.

### III. Dinámica Molecular.

Habiendo minimizado el sistema con alguno de los algoritmos anteriormente mencionados, es posible realizar una dinámica molecular. Para ello, requerimos computar el campo vectorial de fuerzas, para poder calcular aceleraciones y con ello la trayectoria de las partículas. Sin embargo hay que tener algo más en consideración. Dentro del campo de fuerza, un átomo interactúa con múltiples átomos circundantes. Al igual como en el caso de los núcleos y los electrones en mecánica cuántica, la presencia de interacciones de muchos cuerpos impide encontrar una solución analítica para la trayectorias. En estos casos, se utiliza el método de diferencias finitas, que expande las posiciones de los átomos en una serie de Taylor:

$$r(t + \Delta t) = r(t) + v(t)\Delta t + \frac{1}{2}\Delta t^2 a(t) + \frac{1}{6}\Delta t^3 b(t) \dots$$

El detalle de la simulación esta dado por el tiempo de integración (*timestep*) que en dinámica molecular es de 1 o 2 fs. El método más conocido para calcular las trayectorias es el algoritmo de *Verlet*. En este algoritmo se calcula la nueva posición atómica en base a la posición actual y la posición anterior:

$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + \Delta t^2 a(t)$$

Este método tiene como inconvenientes que a) es requerido conocer dos posiciones para calcular la próxima, por lo cual es necesario generar otra conformación, posiblemente por minimización y b) las velocidades no aparecen explícitamente, de modo que es necesario estimar las velocidades en base al cambio de posición después de cada paso. Un algoritmo que soluciona estos inconvenientes es el llamado algoritmo de rana saltarina (*leap-frog*), que tiene la siguiente forma funcional:

$$r(t + \Delta t) = r(t) + \Delta t v(t + \frac{1}{2} \Delta t)$$

$$v(t + \frac{1}{2} \Delta t) = v(t - \frac{1}{2} \Delta t) + \Delta t a(t)$$

En este caso, la determinación de las velocidades y las posiciones están desfasadas por la mitad del *timestep*, de allí en nombre del algoritmo. Para un tiempo dado, es posible calcular la velocidad en un tiempo dado como el promedio de las velocidades calculadas inmediatamente antes y después del paso actual.

Usando las metodologías antes descritas, usando un *timestep* de 1 fs, y asignando vectores de velocidad aleatorios que se ajusten a una distribución de velocidades a 300 K, es posible realizar una simulación de una proteína en el vacío. Una simulación de estas características mantendría la energía total del sistema. En el caso de tener desviaciones de la energía total, es posible regular la temperatura del sistema, cuando esta pase cierto umbral, disminuyendo la velocidad de los átomos por un factor o reasignando la distribución de velocidades a una temperatura menor.

## Referencias:

- Beck DA, Daggett V. 2004. Methods for molecular dynamics simulations of protein folding/unfolding in solution. *Methods* 34(1):112-20.
- Davis H, Snider A. 1992. *Análisis Vectorial*. Mexico: McGraw-Hill Interamericana.
- Hanna M. 1985. *Mecánica cuántica para Químicos*. Tlalpan: Fondo Educativo Interamericano.

- Jorgensen WL, Tirado-Rives J. 1988. The OPLS Force Field for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* 110:1657-1666.
- Leach AR. 2001. *Molecular modelling : principles and applications*. Harlow, England ; New York: Prentice Hall. xxiv, 744.
- Mackereell J, A. D. , Brooks B, Brooks III CL, Nilsson L, Roux B, Won Y, Karplus M. 1998. CHARMM: The Energy Function and Its Parameterization with an Overview of the Program. Schleyer PvR, editor. Chichester: John Wiley & Sons.
- Ponder JW, Case DA. 2003. Force fields for protein simulations. *Adv. Prot. Chem.* 66:27-85.
- Schuler LD, Daura X, van Gunsteren WF. 2001. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *Journal of Computational Chemistry* 22(11):1205-1218.



## ANEXOII

### Funcion de Energía Efectiva 1 (EEF1)

Las propiedades de las macromoléculas biológicas están altamente influenciadas por el medio acuoso en el cual actúan. Por ello, estudios de dinámica molecular realistas requieren tomar en cuenta el efecto del solvente. El modo más simple de considerar estos efectos es utilizar solvente explícito. En este formalismo, la proteína es sumergida en una caja de agua, y se integran las ecuaciones de movimiento de Newton tomando en cuenta las interacciones proteína-proteína, proteína-solvente y solvente-solvente. El problema de esta aproximación, es que las interacciones solvente-solvente aumentan rápidamente al aumentar el tamaño del sistema, reduciendo el rango de problemas que pueden ser estudiados con solvente explícito. Para subsanar esta dificultad, se han desarrollado varias metodologías de solvatación implícita, cuya finalidad es considerar el efecto del solvente sin tener que calcular las interacciones solvente-solvente.

EEF1 es un método basado en *exclusión de solvente*, es decir, calcula la componente de solvatación como en un ciclo termodinámico, llevando el soluto desde el vacío al estado acuoso y calculando el cambio de energía libre del proceso. Lo que calcula este formalismo es la *energía efectiva o potencial de fuerza media* de una macromolécula en solución, la cual es la energía libre de la macromolécula y el promedio sobre todos los grados de libertad del solvente a una temperatura dada. Esta energía efectiva está dada por:

$$W(R) = H_{\text{int } ra}(R) + \Delta G^{\text{slv}}(R)$$

Donde  $R$  corresponde a las coordenadas de la macromolécula,  $H_{\text{int } ra}$  es la energía intramolecular y  $\Delta G^{\text{slv}}$  es la energía libre de solvatación. El cálculo de la energía efectiva requiere entonces de un formalismo que describa el cambio de energía libre de solvatación de una molécula tomando en cuenta solo sus coordenadas. Se ha descrito que la energía libre de solvatación de una molécula puede definirse como una integral alrededor del espacio de la molécula:

$$\Delta G^{\text{slv}} = \int g(r) dr$$

Donde  $g(r)$  es la densidad de energía libre de solvatación alrededor del punto  $r$ . Esta densidad toma en cuenta el cambio de la componente entálpica del sistema, como el cambio de energía de interacción entre solvente y soluto y el cambio de energía de interacción entre moléculas de solvente producido por la aparición del soluto, así como también una componente entrópica, como el cambio en entropía soluto-solvente y la entropía del solvente, tomados como el cambio de conformación estos tras la solvatación. Los cambios más evidentes en entalpía y entropía de interacción soluto-solvente se producen en las primeras capas de solvatación (la presencia del soluto no afecta las propiedades del solvente lejos de la superficie

molecular), por lo cual  $g(r)$  debe decaer rápidamente cuando  $r$  se aleja de la superficie molecular.

Para calcular  $\Delta G^{slv}$ , se considera que es posible descomponer la energía libre como una sumatoria de las contribuciones de  $i$  grupos atómicos dentro de la molécula:

$$\Delta G^{slv} = \sum_i \Delta G_i^{slv}$$

Cada una de las contribuciones de los grupos puede calcularse como la energía libre de solvatación para un compuesto modelo menos la reducción en solvatación debido a los grupos circundantes:

$$\Delta G_i^{slv} = \Delta G_i^{ref} - \sum_j \int_{V_j} g_i(r) dr$$

Donde  $\Delta G_i^{ref}$  corresponde a la energía de solvatación de un compuesto modelo completamente expuesto a solvente, la sumatoria se hace sobre todos los grupos  $j$  circundantes al grupo  $i$ , tomando en cuenta el volumen de exclusión  $V_j$  del grupo  $j$ . Para simplificar el cálculo de la integral, se aproxima a través del producto de la función de densidad del grupo  $i$  y el volumen del grupo  $j$ :

$$\Delta G_i^{slv} = \Delta G_i^{ref} - \sum_{j \neq i} g_i(r_{ij}) \cdot V_j$$

donde  $r_{ij}$  es la distancia entre los centros de los grupos  $i$  y  $j$ . Para permitir la derivación analítica de fuerzas a partir de la energía efectiva, la función de densidad de energía libre corresponde a una función gaussiana:

$$g_i(r) = \frac{\alpha_i}{4\pi r^2} e^{-x_i^2}$$

$$x_i = \frac{r - R_i}{\lambda_i}$$

donde  $R_i$  corresponde al radio de Van der Waals del grupo  $i$ ,  $\lambda_i$  es el *largo de correlación* que se consideró como el grosor de la capa de hidratación (3.5 Å para grupos neutros o polares y 6Å para los grupos cargados) y  $\alpha_i$  es una constante de proporcionalidad dada por

$$\alpha_i = \frac{2\Delta G_i^{free}}{\sqrt{\pi} \lambda_i}$$

donde  $\Delta G_i^{free}$  corresponde a la energía de solvatación del grupo  $i$  aislado. De este modo, es posible obtener una función analítica de la energía efectiva de una molécula, la cual está basada en la energía intramolecular  $H_{int ra}$  obtenida a partir de un campo de fuerza (en este caso CHARMM19), y la energía de solvatación  $\Delta G^{slv}$  que se obtiene como un modelo de exclusión de solvente basado en funciones gaussianas. Esta función analítica puede ser utilizada para métodos de minimización, así como también dinámica molecular.

### ANEXO III

**Lista de residuos que componen los núcleos hidrofóbicos de MjFtsZ y PaFtsZ.**  
Ya que los residuos escondidos se definieron en base a los residuos equivalentes entre MjFtsZ y PaFtsZ, el número de residuos en los núcleos correspondientes de MjFtsZ y PaFtsZ son los mismos.

Núcleo Hidrofóbico	Nº residuos	proteína	Lista de residuos
Interfase	55	MjFtsZ	40 42 44 49 50 52 53 56 61 66 68 76 79 81 124 126 128 153 155 157 159 162 186 188 190 192 195 207 211 215 216 218 219 221 222 223 224 225 234 239 242 243 246 249 251 274 284 286 288 320 322 334 336 338 339
		PaFtsZ	14 16 18 23 24 26 27 30 35 40 42 50 53 55 98 100 102 127 129 131 133 136 160 162 164 166 169 182 186 190 191 193 194 196 197 198 199 200 209 214 217 218 221 224 226 249 260 262 264 296 298 310 312 314 315
Dominio amino	41	MjFtsZ	41 43 45 67 69 85 87 104 107 108 111 115 119 125 127 129 131 132 134 136 137 138 139 140 141 142 143 144 146 147 152 154 156 158 173 176 177 180 183 187 189
		PaFtsZ	15 17 19 41 43 59 61 78 81 82 85 89 93 99 101 103 105 106 108 110 111 112 113 114 115 116 117 118 120 121 126 128 130 132 147 150 151 154 157 161 163
Dominio carboxilo	25	MjFtsZ	252 254 261 262 265 266 269 272 275 282 285 287 289 291 292 300 303 306 307 311 321 331 333 335 337
		PaFtsZ	227 229 236 237 240 241 244 247 250 258 261 263 265 267 268 276 279 282 283 287 297 307 309 311 313

## ANEXO IV

### Programas perl

Nombre	Descripción
RefNatMat.pl	Toma las 20 trayectorias a 300 K y la matriz de frecuencia de contactos ( $F$ ). Se considera que se tienen los archivos pdb de cada cuadro de la trayectoria, esto se puede lograr utilizando el script <b>sepPdb.tcl</b> dentro de VMD.
NatCont.pl	Lee las matrices generadas por <b>RefNatMat.pl</b> y calcula el número de contactos nativos dentro de cada núcleo, tanto en los cristales, la simulación de la proteína silvestre y las simulaciones de los dominios. Genera una distribución de frecuencias de la frecuencia de los contactos Utiliza la lista de residuos de los núcleos hidrofóbicos listada en el ANEXOIV. Utilizado para generar la tabla 4.
ListNatCont.pl	A partir de las matrices de las simulaciones nativas, calcula las matrices de contactos binarias con los contactos que tienen más de 70% de ocupancia durante las simulaciones. utilizado para generar figura 15.
ContNat.pl	Toma las matrices binarias de contactos para calcular el número de contactos nativos en los núcleos hidrofóbicos durante las simulaciones. Se calcularon los núcleos hidrofóbicos de la proteína silvestre y los dominios aislados. Lo que se obtuvo para cada simulación es una tabla con el número de contactos de cada uno de los núcleos para cada cuadro de la simulación.
secStructResvStride.pl	Toma matrices creadas por <b>allSecStructTraj.tcl</b> y calcula frecuencias de estructura secundaria.
Tm.pl	Calcula el perfil de valores $S_{1/2}$ a partir de las matrices R

### Programas Tcl

Nombre	Descripción
sepPdb.tcl	Separa archivos de trayectoria binarios generados por CHARMM en archivos pdb de cada uno de los cuadros de la simulación.
secStructTraj.tcl	Genera matrices de estructura secundaria a partir de simulaciones de dinámica. Las matrices generadas pueden ser vistas con MATLAB. utilizado para generar la figura 7.
allSecStructTraj.tcl	Cicla sobre varias simulaciones y utiliza el procedimiento <b>secStructTraj</b> para calcular matrices de estructura secundaria.
sasaMutantes.tcl	Para las proteínas completas, para cada simulación, y para cada temperatura, calcula las accesibilidades a solvente de los residuos correspondientes a las mutantes de E. coli
Rmsf.tcl	Calcula el rmsf para una selección de carbonos alfa de una proteína. Utilizado para generar la figura 8.

