

UNIVERSIDAD DE CHILE

FACULTAD DE CIENCIAS QUÍMICAS Y FARMACÉUTICAS



Caracterización biofísica a nivel de molécula individual del plegamiento del regulador transcripcional RfaH

Tesis presentada a la Universidad de Chile para optar al grado de Magíster en Bioquímica área de Especialización en Proteínas y Biotecnología y Memoria para optar al Título de Bioquímico por:

PABLO ANTONIO GALAZ DAVISON

**Director de Tesis: Dr. César A. Ramírez Sarmiento
Co-Director de Tesis: Dr. Christian A. M. Wilson Moya**

Santiago-CHILE

Octubre 2017

UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS QUÍMICAS Y FARMACÉUTICAS

INFORME DE APROBACIÓN DE TESIS DE MAGISTER

Se informa a la Dirección de la Escuela de Graduados de la Facultad de Ciencias Químicas y Farmacéuticas que la Tesis de Magíster y Memoria de Título presentada por el candidato

PABLO ANTONIO GALAZ DAVISON

Ha sido aprobada por la Comisión de Evaluadora de Tesis como requisito para optar al grado de Magíster en Bioquímica, Área de Especialización: Proteínas y Biotecnología y Título de Bioquímica, en el examen público rendido el día

Director de Tesis:

Dr. César A. Ramírez Sarmiento _____

Co-director de Tesis:

Dr. Christian A. M. Wilson Moya _____

Comisión Evaluadora de Tesis:

Dr. Mauricio Báez Larach _____

Dra. Carolina Mascayano Collado _____

Dra. María A. Valenzuela Pedevila _____

Agradecimientos

Esta tesis fue realizada con el apoyo del Laboratorio de Bioquímica de la Universidad de Chile y el Laboratorio de Ingeniería de Proteínas y Microorganismos de la Pontificia Universidad Católica de Chile. Profundos agradecimientos al grupo de investigación del Dr. Jorge Babul y Dra. Victoria Guixé por toda la ayuda instrumental, metodológica y teórica que me han proporcionado, al Instituto de Ingeniería Biológica y Médica de la Pontificia Universidad Católica de Chile por el apoyo logístico y financiamiento, y al Fondo Nacional de Desarrollo Científico y Tecnológico de Chile e Instituto Antártico Chileno por el financiamiento. Especial agradecimiento a Javiera Reyes por toda la ayuda experimental y paciencia, y a mi tutor César Ramírez por todo el apoyo y motivación para mi trabajo y formación. También agradezco a mi co-director, el Dr. Christian Wilson por amablemente ofrecerme un espacio e instrumental para llevar a cabo mi investigación, junto con todo el apoyo teórico y metodológico necesario para finalizar este trabajo.

A mis amigos y cercanos con quienes hemos festejado las glorias y pasado las penas, nuestro grupo de laboratorio, los mandriles de Ciencias, compañeros de carrera y especialmente a Cecilia Muster y Beatriz Salas que me han hecho disfrutar estar en la Universidad. A mis amigos que tengo lejos, especialmente a Sebastián Retamal y Evelyn Carrasco, que me recuerdan que, al final del día, somos los mismos después de 10 años.

A mis dos familias, mis papás Margarita Davison y Juan Galaz por darme todo lo necesario para poder estar en la Universidad y hacer lo que me gusta, y a mis tíos Nancy Davison y Sergio Escobedo por toda la ayuda y adopción en su casa y familia.

Agradezco a la comisión por la supervisión y guía de mi trabajo de tesis, por darme la flexibilidad para realizar mi investigación de distintos modos para terminar este trabajo.

Especial agradecimiento a César Ramírez, por dejarme llevar a cabo mi investigación del modo que más disfruto, apoyar incondicionalmente mi formación científica y toda la confianza que ha puesto en mi trabajo, resultados y propuestas.

También agradecer a los profesores y ayudantes de la carrera de Bioquímica de la Facultad de Ciencias Químicas y Farmacéuticas por la formación que me han

entregado, que he apreciado y ha sido valiosa cada momento desde que salí de la facultad.

Finalmente, gracias a todos los biofísicos y biólogos computacionales por desarrollar el campo más entretenido de la ciencia.

Índice

Índice	5
Índice de Elementos.....	7
Abreviaturas.....	9
Resumen.....	11
Abstrac	13
I. Introducción	14
Estructura y función de RfaH.....	14
Estudios computacionales	17
Estudios de molécula individual	24
I. Hipótesis.....	27
II. Objetivo General.....	27
Objetivos Específicos	28
III. Teoría y Métodos	29
1) Computacionales	29
2) Experimentales	36
IV. Resultados.....	40
Objetivo 1: Estudio global del desplegamiento mecánico utilizando SBM.....	40
Objetivo 2: Caracterización del plegamiento de RfaH mediante SBM	47
Trabajo al equilibrio de desplegamiento de α - y β CTD.....	50
Objetivo 3: Análisis termodinámico de RfaH a través de CCR.....	54
V. Discusión	59
Relación de estiramientos in silico/in vitro.	59
Desplegamiento mecánico de RfaH	61
Desplegamiento reversible de RfaH.....	63
Cálculos de energía por confinamiento-conversión	64
Resultados experimentales	65
Conclusiones y perspectivas futuras	66
VI. Referencias.....	68
Anexo.....	74
Construcción de mutantes para pinzas ópticas	74
Modificación de construcciones para pinzas ópticas	76

Índice de Elementos

Figuras

Figura 1 Estructura primaria, secundaria y terciaria de RfaH.....	14
Figura 2 Granularidad en dinámica molecular	17
Figura 3 Modelo de estado de Markov de RfaH	18
Figura 4 Transformación de RfaH por dinámica dirigida.....	19
Figura 5 Panorama energético de la transformación de RfaH	21
Figura 6 Panorama bidimensional de la transición de RfaH.....	22
Figura 7 Espectroscopía de fuerza por pinza o rampa de fuerza.....	24
Figura 8 Confinamiento y conversión	31
Figura 9 Análisis de traza de fuerza	42
Figura 10 Correlación entre desplegamiento mecánico computacional y experimental	43
Figura 11 Efecto de la velocidad de estiramiento.....	44
Figura 12 Efecto de temperatura	44
Figura 13 Desplegamiento mecánico de α RfaH.....	46
Figura 14 Desplegamiento mecánico de β RfaH.....	47
Figura 15 Desplegamiento químico del CTD de RfaH.....	48
Figura 16 Desplegamiento reversible del CTD y CTD-RfaH.....	50
Figura 17 Energía al equilibrio del desplegamiento del CTD.....	51
Figura 18 Confinamiento de RfaH	53
Figura 19 Diferencia de energía de RfaH	54
Figura 20 Descomposición de energía de transformación	55
Figura 21 Comparación computacional-experimental de estabilidad local	56
Figura 22 Purificación de RfaH-2C	74
Figura 23 Cromatograma de exclusión molecular de RfaH.....	75
Figura 24 Reactividad de cisteínas en RfaH	76
Figura 25 Espectro de masas de reacción de marcación.....	78
Figura 26 Electroforesis de mangas de ADN.....	79

Ecuaciones

Ecuación 1 Potencial clásico Amber	16
Ecuación 2 Potencial de estiramiento	29
Ecuación 3 Fluctuaciones cuadráticas globales.....	30
Ecuación 4 Energía libre de estado confinamiento	31
Ecuación 5 Energía de un oscilador armónico	32

Ecuación 6 Energía libre por confinamiento.....	32
Ecuación 7 Descomposición de energía libre	33
Ecuación 8 Energía libre de Helmholtz	33
Ecuación 9 Teorema de fluctuaciones de Crooks.....	34

Tablas

Tabla 1 Proteínas estudiadas computacionalmente (al 8/9/2016)	40
Tabla 2 Construcciones utilizada durante el trabajo de tesis	73
Tabla 3 Recuperación de proteína tras reacción con DTDP	77

Esquemas

Esquema 1 Hipótesis de la transformación de RfaH.....	26
Esquema 2 Esquema global del plegamiento de RfaH	65
Esquema 3 Desarrollo de quimeras proteína-ADN	73
Esquema 3 Protocolo de marcación de RfaH-2C	79

Abreviaturas

A/BA	Acrilamida/Bis-acrilamida
AbsCO	<i>Absolute contact-order</i> (Orden de contacto absoluto)
ACN	acetonitrilo
AFM	<i>Atomic force microscope</i> (Microscopio de fuerza atómico)
ARNP	Polimerasa de ARN
Bio	Biotina
CCR	<i>Confinement-conversion-release</i> (Confinamiento-conversión-liberación)
CD	<i>Circular dichroism</i> (Dicroísmo circular)
CGMD	<i>Coarse-grained molecular dynamics</i> (Dinámica molecular de grano grueso)
CMD	<i>Conventional molecular dynamics</i> (Dinámica molecular convencional)
CTD	<i>C-terminal domain</i> (Dominio carboxilo-terminal)
Dig	Digoxigenina
DMSO	dimetilsulfóxido
DRMS	Diferencia en RMSD
DTDP	2,2-ditiopiridina
DTNB	ácido di-tionitrobenzoico
DTT	ditiotreitól
EtOH	Etanol
F _m	Fuerza de ruptura media
GDL	Grados de libertad
HDX	<i>Hydrogen-deuterium exchange mass spectrometry</i> (Espectrometría de masas de intercambio hidrógeno-deuterio)
I	Intermediario
IMAC	<i>Immobilized metal affinity chromatography</i> (Cromatografía de afinidad a metal inmovilizado)
IPTG	isopropil β-D-1-tiogalactopiranosido
LB	<i>Lysogeny broth</i> (Medio de lisogenia)
MALDI-TOF	<i>Matrix-assisted laser desorption/ionization - time of flight</i> (desorción-ionización ayudada por matriz – tiempo de vuelo)
MMPBSA	<i>Molecular Mechanics Poisson-Boltzmann Surface Area</i> (Area superficial de mecánica molecular Poisson-Boltzmann)
MSM	<i>Markov state model</i> (Modelo de estado Markov)
NMR	<i>Nuclear magnetic resonance</i> (Resonancia magnética nuclear)
NTD	<i>N-terminal domain</i> (Dominio amino-terminal)
PCR	<i>Polymerase chain reaction</i> (Reacción en cadena de polimerasa)
PDB	<i>Protein Data Bank</i> (Banco de datos de proteína)

PMSF	fluoruro de p-metilsulfóxido
R2C	RfaH doble mutante Q2C, R160C
REMD	<i>Replica exchange molecular dynamics</i> (Dinámica molecular de intercambio de réplica)
RMSD	<i>Root mean-squared deviation</i> (Raíz de la desviación cuadrática media de distancias)
SBM	<i>Structure-based models</i> (Modelos basados en estructura)
SDS-PAGE	<i>Sodium dodecyl sulphate – polyacrylamide gel electrophoresis</i> (Electroforésis en gel de poliacrilamida en presencia de dodecilsulfato de sodio)
SEC	<i>Size exclusion chromatography</i> (Cromatografía de exclusión molecular)
SMD	<i>Steered molecular dynamics</i> (Dinámica molecular de estiramiento)
TB	<i>Terrific broth</i>
TEC	<i>Transcription elongation complex</i> (Complejo de elongación de transcripción)
TMD	<i>Targeted molecular dynamics</i> (Dinámica molecular dirigida)
Tris	trimetanoamina
UV	<i>Ultra-violet light</i> (luz ultravioleta)
β° CH	β° -clamp helix
β GL	β -gate loop

Resumen

La proteína RfaH de *E. coli* es un regulador transcripcional perteneciente a la ampliamente conservada familia NusG, que incrementa la expresión de genes distales en operones estrechamente relacionados con virulencia. Esta expresión controlada toma lugar de manera posterior al reclutamiento específico de RfaH a una hebra de ADN expuesta en la superficie del complejo de elongación de transcripción, pausado en la secuencia nucleotídica *ops*. RfaH posee un dominio C-terminal plegado en forma de α -hélice (α CTD) compactada contra la superficie de unión a polimerasa de ARN (ARNP) del dominio N-terminal e inhibiéndolo, mientras que el de su parálogo NusG corresponde a un plegamiento barril- β (β CTD) sin interacción a la ARNP y con alta movilidad. Sin embargo, cuando ambos dominios de RfaH se disocian (evento necesario para unirse a la ARNP), el dominio C-terminal sufre un cambio de plegamiento desde α -hélice a un barril- β idéntico estructuralmente al de NusG, permitiendo la activación del NTD, y además reclutando la proteína ribosómica S10 para acoplar la transcripción con la traducción.

RfaH es la primera proteína metamórfica cuyo cambio estructural está asociado a un cambio funcional. Debido a la escasa estabilidad de RfaH en solución la mayoría de los estudios termodinámicos han sido computacionales, sugiriendo la presencia de un intermediario que conecta ambos estados; sin embargo, a la fecha no se ha realizado un análisis profundo que conecte las dinámicas observadas computacionalmente con datos experimentales contrastables.

En este trabajo de tesis se evaluó si el plegamiento del dominio C-terminal de RfaH ocurre previo al desplegamiento del NTD, con especial énfasis en las características termodinámicas de aquél proceso. Esto se realizó mediante distintos enfoques computacionales incluyendo modelos basados en estructura y dinámica molecular de confinamiento.

Computacionalmente, se utilizaron modelos nativo-céntricos para simular el desplegado mecánico de RfaH en ambos estados nativos. A pesar de que estos modelos se han usado consistentemente para estudiar fenómenos de plegamiento, las magnitudes computacionales o “reducidas” no se encuentran calibradas respecto a su contraparte experimental. Por ello, inicialmente se estudió el desplegado mecánico de 35 proteínas depositadas en el PDB para poder estimar la capacidad de estos modelos de reproducir características experimentales. Utilizando ponderación por orden de contacto, se obtuvo una aproximación lineal de las fuerzas estimadas/experimentales ($R=0,934$) con un error estándar de 56 pN. Al realizar simulaciones de estiramiento de RfaH en estas condiciones, se observó consistentemente que el CTD se despliega independiente del NTD. A una velocidad de $100 \text{ nm}\cdot\text{s}^{-1}$, la disociación y desplegado del α CTD ocurre a $43 \pm 6 \text{ pN}$, mientras que el β CTD se despliega a los $35 \pm 5 \text{ pN}$ dando lugar a la formación de un intermediario β -extendido, que expone su núcleo hidrofóbico y se despliega a los $29 \pm 4 \text{ pN}$. Utilizando este enfoque también se modeló el replegamiento del CTD en ambos

estados. Mostrando alta histéresis para el caso del α CTD, y distribuciones de trabajo solapantes con las del desplegamiento para ambos pasos de desplegamiento del β CTD, permitiendo estimar un ΔG de 12,0 y 8,7 kcal mol⁻¹ para la reacción β CTD \rightleftharpoons I \rightleftharpoons desplegado, respectivamente.

Por otro lado, se calculó la estabilidad de ambos estados mediante dinámica molecular de confinamiento, entregando un $\Delta G = 6,5$ kcal·mol⁻¹ para la reacción α RfaH \rightleftharpoons β RfaH. Una descomposición a nivel de residuo de esta diferencia de energía muestra alta correlación con la estabilidad relativa calculada a partir de espectrometría de masas de intercambio hidrógeno-deuterio. Esto permitió identificar una estabilidad diferencial de segmentos al interior de RfaH, en la cual los residuos 129-146 están altamente estabilizados en el α CTD asociado al NTD, mientras que las otras dos regiones 115-128 y 147-162 favorecen enormemente un plegamiento β CTD, y apoyan la predicción energética predicha por las dinámicas de confinamiento.

Este trabajo corresponde a la primera investigación termodinámica de la transición de RfaH, dando luces además de los directores en secuencia de esta transformación, y caracterizando su estabilidad y vía por la cual se conecta con el estado desplegado.

Abstract

The idea of a singular native state has been overruled by metamorphic proteins such as RfaH. This is a transcription factor composed of two domains: The N-terminal (NTD, 100 residues) and C-terminal (52 residues) domains. In solution this protein is autoinhibited, and it becomes activated by completely refolding its C-terminal domain from its initial α -hairpin to a β -barrel when binding to the transcriptional machinery. Through multiple molecular dynamics approaches, it was possible to computationally characterize and predict the folding pathways of both RfaH states, and estimate the free-energy difference required for converting them. Using structure-based models, the single-molecule mechanical unfolding of α -folded RfaH was simulated, which takes place by two sequential events: the simultaneous dissociation and unfolding of its α -helical C-terminal domain (α CTD), and a secondary unfolding of the NTD. On the other hand, when starting from the β -folded C-terminal domain (β CTD), RfaH unfolding occurs mediated by two different intermediates: first, the β -barrel opens and becomes solvent exposed turning into an extended- β intermediate. In a tertiary transition, this intermediate is unfolded and only the NTD remains structured, which shares the same folding characteristics as when unfolding from the α -state RfaH. Analysis of the unfolding of 35 different proteins using structure-based models allowed for estimation of forces from simulations. Using this approach, reversible unfolding of CTD was performed, yielding high hysteresis between the un- and refolding of the α CTD; and overlapping work distributions for such processes in the β CTD, enabling equilibrium free energy calculation through Crooks fluctuation theorem. Thus, the predicted ΔG° values for the reaction β CTD \rightleftharpoons Intermediate \rightleftharpoons unfolded are 12.0 and 8.7 kcal mol⁻¹ respectively. By using confinement molecular dynamics, it was possible to estimate the free energy of the RfaH metamorphosis (α RfaH \rightleftharpoons β RfaH) calculated to be 6.5 kcal mol⁻¹. The further decomposition of this free energy is highly correlated with thermodynamic analysis of hydrogen-deuterium mass spectrometry experiments performed for both CTD states. Altogether, this research provides insights into the folding mechanism and thermodynamics of the metamorphic RfaH protein, constituting a significant step forward in understanding the molecular determinants behind the duality of metamorphic proteins.

I. Introducción

Estructura y función de RfaH

La proteína RfaH de *Escherichia coli* (acceso UniProt: P0AFW0) es un factor de elongación, perteneciente a la familia de factores transcripcionales ampliamente conservados NusG [1]. Al igual que su parálogo NusG, RfaH se une a la polimerasa de ARN (ARNP) mientras ésta forma el complejo de elongación de transcripción (TEC) al asociarse con el ADN. Esta interacción entre el TEC y RfaH o NusG permite el aumento de la procesividad de la transcripción y la reducción de pausas en sitios génicos propensos a la inducción de retrocesos [1]–[3].

A pesar de pertenecer a la misma familia de proteínas, NusG y RfaH tienen diferencias fundamentales respecto a su función: mientras NusG es una proteína esencial que regula la transcripción de la mayoría de los genes de *E. coli* de manera inespecífica [4], [5], RfaH reconoce un elemento de doce pares de bases (5'-GGCGGTAGnnTG-3') [6] denominado supresor de polaridad de operon (*ops*), el cual se encuentra particularmente en operones que contienen genes asociados a la virulencia bacteriana y la síntesis de lipopolisacáridos [3]. Este reconocimiento se lleva a cabo en la superficie del TEC, donde el elemento *ops* en la hebra no codificante es expuesto en la superficie de ARNP, lo que permite el reclutamiento y unión de RfaH al complejo, convirtiéndolo en un regulador específico [1]–[3]. Otra gran diferencia es que NusG promueve la terminación de transcripción mediada por la proteína Rho (ρ) en genes foráneos [7], mientras que RfaH inhibe la acción de ρ y por tanto activa la expresión de genes transferidos horizontalmente [2].

Esta diferencia funcional entre NusG y RfaH se explica en gran medida por sus diferencias estructurales. Al igual que NusG, RfaH consta de dos dominios (Figura 1a) [8]–[10]. Su dominio amino terminal (NTD) es conservado estructural y funcionalmente entre los miembros de la familia NusG en bacteria, arqueobacteria y eucariotas [9], [11], [12] (Figura 1b). Este dominio tiene una arquitectura de sándwich de dos láminas que contiene motivos de residuos que permiten la unión a dos sitios de ARNP: la punta de las hélices pinza de la subunidad β' (β' *Clamp-Helix*, β' CH) y el bucle de entrada β (β *Gate*

Loop, β GL), además de poseer un dominio de unión a ADN [13]–[15]. El motivo de unión a ADN y a β' CH son responsables en mayor medida de la afinidad por ARNP, mientras que el motivo de unión a β GL (también llamado motivo HTT) no afecta la unión, pero sí el aumento de procesividad y actividad anti-pausa [12], [16], [17]. Además, la unión de éste dominio al TEC permite aislarlo físicamente de la subunidad σ , y evitar pausas inducidas por ella, incluso cuando su unión involucra contactos distintos [18].

En el otro extremo de la proteína y separado del NTD por un espaciador flexible, se encuentra el dominio carboxilo terminal (CTD), el cual es sorprendentemente distinto en NusG y RfaH. Por un lado, NusG posee un CTD plegado en forma de un barril- β anti paralelo (β CTD), característico de los miembros de la familia [10], mientras que el CTD de RfaH posee una estructura de horquilla de α -hélice (α CTD; Figura 1c)

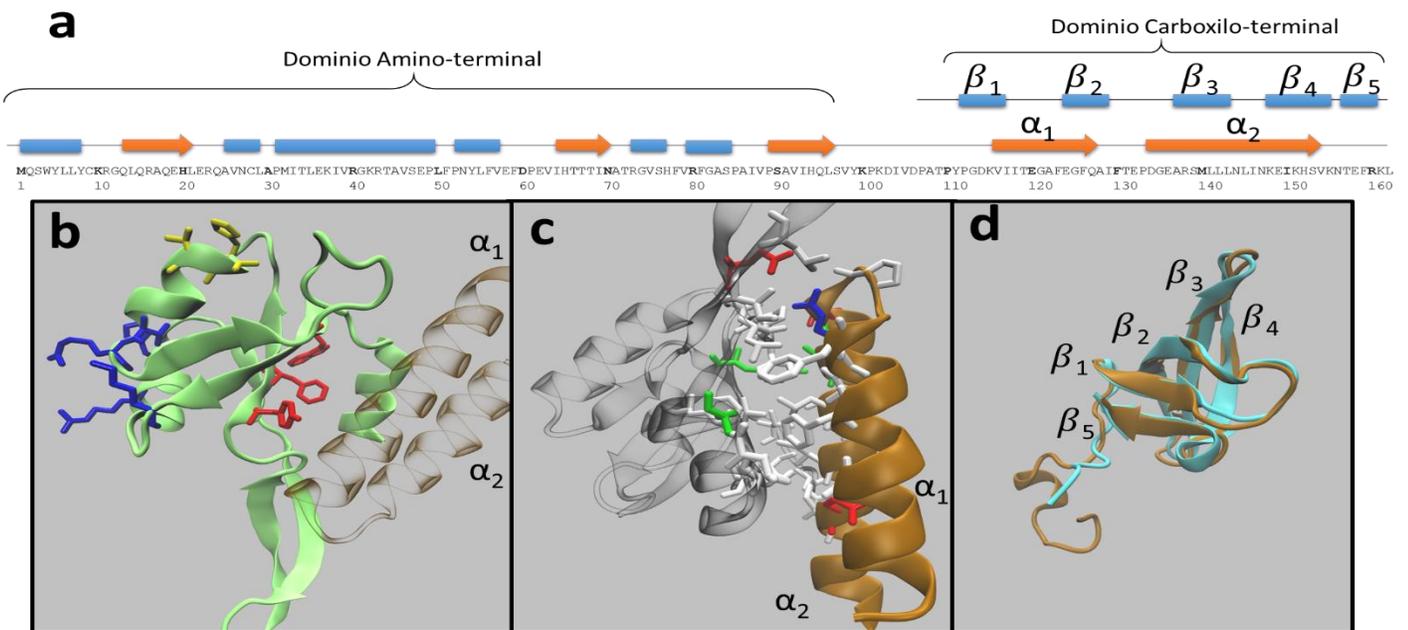


Figura 1: Estructura primaria secundaria y terciaria de RfaH. (a) Estructura primaria y secundaria de RfaH. En azul se muestra las hojas- β y en naranja las α -hélices. El NTD comprende del residuo 1 al 100, el espaciador flexible del 101 al 114 y el α CTD del 115 al 162. Sobre el α DCT se muestra la estructura del β DCT [9]. (b) Representación de cintas de la estructura cristalográfica de RfaH (PDB 2UOG), con el NTD en verde y las cadenas laterales de los residuos pertenecientes al motivo de unión a ADN (azul), motivo HTT (amarillo) y motivo de unión a β' CH (rojo). También se muestra transparente el α CTD (café) denotando las hélices 1 y 2. El espaciador no fue resuelto en la estructura cristalográfica de RfaH. (c) Estructura de RfaH similar a (b), pero indicando los residuos que median las interacciones entre el α CTD (café) y NTD (transparente). En blanco se muestran los residuos apolares, en verde los polares no cargados, y en azul y rojo los con carga positiva y negativa, respectivamente. Se aprecia claramente el bolsillo hidrofóbico formado entre ambos dominios, y el puente salino entre el Glu48 (rojo) y Arg138 (azul). (d) Alineamiento estructural del β CTD de NusG (celeste) y RfaH (café, PDB: 2LCL). La numeración de las hojas- β de amino a carboxilo terminal está indicada en la figura [10].

fuertemente unida al NTD [8], [9]. Debido a este empacamiento, el α CTD enmascara el sitio de unión a ARNP ubicado en el NTD, dejándole sólo la posibilidad de interactuar mediante el sitio de unión a ADN tras el reconocimiento del elemento *ops* en el TEC, constituyendo así un estado estructural auto-inhibido y la causa de la especificidad de RfaH frente a la función transcripcional ubicua de NusG [3]. Sin embargo, el α CTD tiene otra característica que lo hace único: algunos experimentos de transcripción *in vitro* demostraron que el reclutamiento de RfaH en el TEC mediante el elemento *ops* es acompañado de un aumento de la procesividad, sugiriendo que este evento permite la disociación de los dominios de RfaH [5]. A nivel estructural se ha demostrado que este evento de disociación entre ambos dominios causa que el CTD experimente una transformación desde el estado α -helicoidal a un barril- β esencialmente idéntico en su estructura al plegado canónico del CTD de NusG [9] (Figura 1d). Esta última observación se realizó en solución, mediante experimentos de NMR, tras reducir la interacción entre ambos dominios mediante la eliminación de un puente salino que los une, y también tras la modificación del espaciador entre dominios para efectuar el corte proteolítico del mismo, que provocó una disminución de la población de α CTD con la aparición concomitante de la población β CTD [9]. Estos aspectos convierten a RfaH en la primera proteína *transformer* descrita [9], es decir, proteínas que desafían la idea de *una secuencia – una estructura – una función* del plegamiento de proteínas, al alcanzar dos estados nativos disímiles tanto en su estructura como función a partir de una misma secuencia aminoacídica. Y es porque, además de adquirir un plegamiento completamente distinto, el β CTD recluta la proteína ribosómica S10 para acoplar la transcripción con la traducción [9]. Esto da lugar a los tres mecanismos por los que RfaH inhibe la acción de la proteína ρ : 1.- aumentando la procesividad de la ARNP, 2.- evitando la unión de NusG y con ello la terminación de transcripción por parte de ρ , y 3.- asociando físicamente la maquinaria transcripcional y traduccional.

Estudios computacionales

Ha resultado difícil realizar experimentos de RfaH para analizar su transformación α -a β -plegado, debido a que el NTD es inestable una vez dissociado del CTD y tiende a agregar [9]. Es por ello que la mayoría de los estudios que se han enfocado en estudiar termodinámicamente éste fenómeno, lo han hecho mediante distintas variantes de una técnica computacional llamada dinámica molecular. Esta herramienta permite estudiar el movimiento físico de conglomerados atómicos y moleculares, acorde a funciones de energía potencial que componen un campo de fuerza, el cual describe cómo las partículas interactúan entre sí en una simulación (Ecuación 1) [19], normalmente derivado de información experimental o simulaciones a nivel cuántico. La dinámica molecular convencional (CMD) corresponde a una simulación de un sistema descrito a nivel atómico (PMID: 22257762).

$$\begin{aligned} U = & \sum_{bonds} k_r (r - r_0)^2 \\ & + \sum_{angles} k_\theta (\theta - \theta_0)^2 \\ & + \sum_{dihedrals} k_\phi [1 + \cos(n\phi + \phi_0)] \\ & + \sum_{atom\ i} \sum_{j \neq i} 4 \epsilon_{i,j} \left[\left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{12} - \left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^6 \right] \\ & + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\epsilon_0 r_{i,j}} \end{aligned}$$

Ecuación 1: Potencial clásico de Amber. Función de energía potencial descrita para la primera versión del hamiltoniano AMBER [19]. Los primeros tres términos corresponden a la diferencia entre la geometría puntual y la ideal de la partícula, los dos siguientes representan interacciones no covalentes de Van der Waals e interacciones electrostáticas.

En general, encontramos alrededor de 2000 átomos para una proteína globular pequeña (en promedio 19 átomos por residuo). Dado que los parámetros de los campos de fuerza son obtenidos generalmente en soluciones acuosas, estos sistemas requieren la adición de átomos que involucran el medio circundante (solvente o membrana), pudiendo entonces superar fácilmente los 10.000 átomos por sistema. Cada paso de dinámica molecular es normalmente del orden de los femtosegundos (10^{-15} s), y debido al enorme esfuerzo computacional requerido por cada paso, estas simulaciones rara vez superan 100

nanosegundos (10^{-9} s). Sin embargo, la escala temporal en la que ocurren cambios macroestructurales en proteínas, como es el caso de fenómenos de plegamiento y cambios conformacionales, es normalmente del orden de los milisegundos (10^{-3} s) a segundos (PMID: 22257762). Para superar esta enorme barrera computacional se han desarrollado alternativas que permiten reducir el número de átomos del sistema, como por ejemplo el uso de solvente implícito. En este escenario, se considera que el solvente actúa como un continuo, de manera uniforme sobre todos los átomos del sistema, por lo que es integrado en el campo de fuerza como un potencial de solvatación (PMC4045398). Otra manera de reducir el número de átomos es utilizando una representación de grano grueso (CGMD), en la que cada partícula representa un grupo de átomos (Figura 2). Para proteínas, normalmente se utiliza este método para acceder a grandes tiempos de simulación, reduciendo la representación del agua o de cada residuo a una o dos partículas (PMID: 27333362). Para validar este tipo de simulación simplificada, se suele buscar correlación experimental o con su contraparte atomística.

Algunos experimentos de dinámica molecular de la estructura cristalina de los

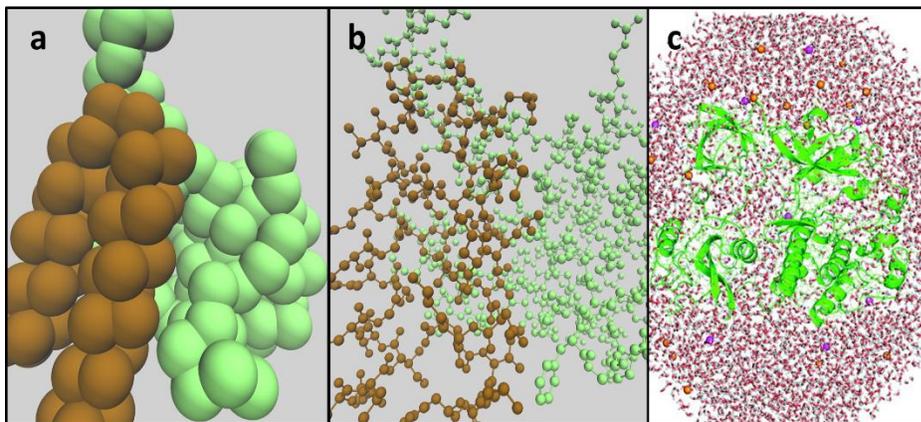


Figura 2: Granularidad en dinámica molecular. Representaciones de dinámica molecular. (a) y (b) corresponde a la estructura de RfaH en estilo grano grueso (162 partículas), donde sólo se representa el carbono- α por residuo, y atomística (aprox. 1300 átomos), donde se representan todos los átomos pesados, respectivamente. (c) Simulación de proteína solvatada (aprox. 50.000 átomos) [19]

estados α CTD y β CTD en ausencia del NTD, utilizando CMD o CGMD muestran que el estado β -plegado es estable durante la simulación. Por otro lado, el estado α CTD es incapaz de mantener puentes de hidrógeno consistentes y pierde su característica helicoidal en el tiempo, lo que permite explicar por qué el estado β -plegado del CTD predomina en solución al estar dissociado del NTD [9], [20], [21]. Esto se contrasta con

simulaciones atómicas con solvente explícito de la proteína RfaH completa en ambos estados, donde las interacciones entre dominios estabilizan la conformación α CTD, mientras que β CTD tiene una alta fluctuación debido al espaciador flexible, pero mantiene establemente su característica de plegamiento- β [22].

Utilizando simulaciones en solvente implícito, se construyó un modelo de estado Markov (*Markov State Model*, MSM), el que utiliza un muestreo estadístico de simulaciones del CTD aislado para generar un esquema de replegamiento de este dominio desde la conformación α a β [23]. Un análisis del MSM para la transición irreversible desde la conformación α a β del CTD permitió dar luces sobre su transformación, sugiriendo que esta ocurre principalmente en una conformación pobremente estructurada (Figura 3), siendo el primer paso un desenrollamiento de la hélice α_1 seguida por la pérdida de toda estructura helicoidal, posterior formación de las hebras β_2 y β_3 (que los autores sugieren son particularmente estables), hasta formar la estructura de barril- β , tomando un tiempo total calculado de 100 μ s. En este proceso, hay varios posibles intermediarios en las múltiples vías observadas para la transición entre ambas estructuras, pero tres de ellos son estadísticamente significativos [21].

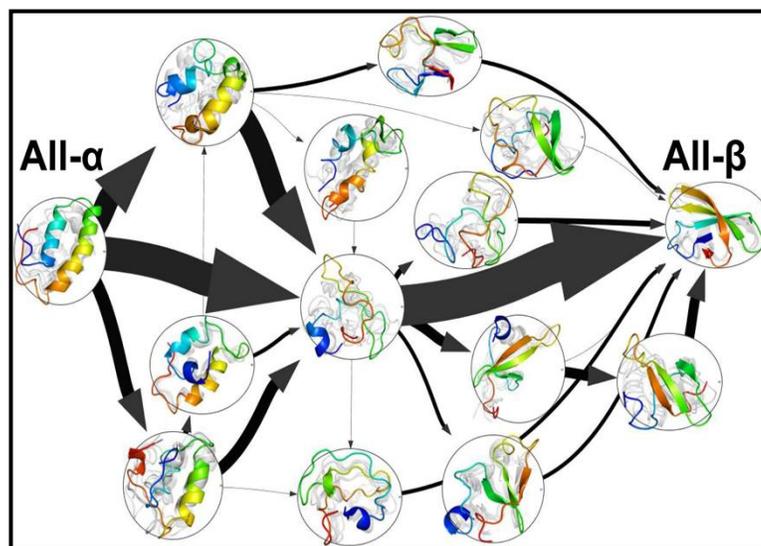


Figura 3: Modelo de estado de Markov de RfaH. Flujos de las transiciones entre intermediarios para el CTD de RfaH de acuerdo a MSM. Se aprecia la multiplicidad de vías que pueden mediar la transformación α CTD a β CTD, y aquellos intermediarios que presentan el flujo predominante para la transición, mediante estados intermediarios pobremente estructurado (centro). Extraído de [22].

Otro análisis de esta conversión estructural se realizó mediante dinámica molecular de intercambio de réplica (*Replica Exchange Molecular Dynamics*, REMD). En este método, múltiples copias de un sistema (llamadas réplicas) son simuladas en paralelo a distintas temperaturas. Cada cierto intervalo de tiempo, configuraciones de pares de estas réplicas a distintas temperaturas son intercambiadas entre sí. Este intercambio tiene una probabilidad basada en un criterio tipo Metropolis, que permite lograr un recorrido aleatorio de las temperaturas y con ello mejorar la exploración del espacio configuracional, disminuyendo el tiempo de cálculo requerido para la transición [24]. Los análisis del CTD usando REMD atómica en solvente implícito permitieron una rápida conversión estructural de α CTD a ovillo al azar a β CTD, sin ser posible observar el proceso inverso (Figura 4). Al igual que el MSM, se observó un rápido desenrollamiento de la hélice α_1 , y posterior pérdida de estructura secundaria, con la diferencia que las hebras β_3 y β_4 forman una horquilla que nuclea las otras hebras para la formación del barril- β [25]. Sin embargo, tanto simulaciones de MSM como este método no predicen el comportamiento de la proteína completa, dada la ausencia del NTD.

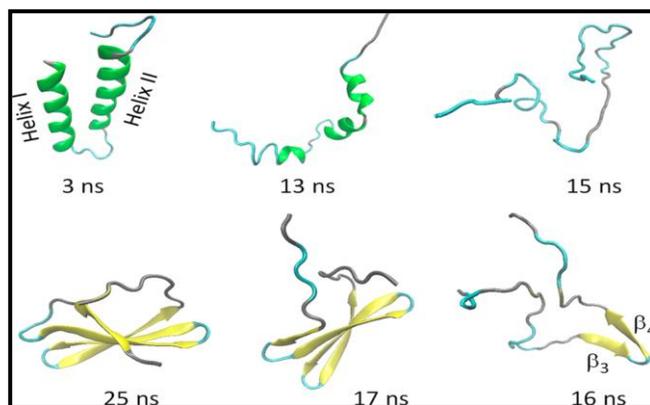


Figura 4: Transformación de RfaH por dinámica molecular dirigida. Imágenes representativas del cambio estructural del CTD de RfaH mediante REMD. Todas las réplicas presentan una trayectoria similar, y ésta logra una estructura final más parecida al β CTD. Se observa la aparición de un intermediario pobremente estructurado (15 ns). Extraído de [25].

Un análisis de la transición del CTD de RfaH en el contexto de la proteína completa se realizó recientemente, utilizando un modelos basados en estructura (SBM), es decir, es dirigido hacia el estado nativo mediante la inclusión de la topología en la función de energía potencial, de modo que todas las interacciones nativas resultan

estabilizantes, reduciendo la exploración del espacio conformacional necesaria para llegar al estado nativo, y con ello el tiempo de cálculo [26]. Utilizando CGMD de éste modelo se recapituló el efecto de las interacciones entre dominios de RfaH sobre el plegamiento del CTD reportado experimentalmente, y además se observó reversibilidad de la transición entre estados al reducir la fuerza de los contactos entre dominios a la mitad ($\epsilon^{IF} \cong 0,5$) [27]. Esta reversibilidad es apoyada experimentalmente por ensayos de intercambio de dominios, que consiste en la inversión del orden de los dominios en el gen de RfaH, permitiendo que se pliegue el CTD previo al NTD en el péptido naciente desde el ribosoma. Si no existe reversibilidad, entonces RfaH mantendrá la conformación β CTD característica de éste dominio en solución; sin embargo, se demostró estructural y funcionalmente la presencia del estado α CTD [28]. Más aún, el estudio estructural de RMN- ^1H y ^{13}C demostró que al desestabilizar la interacción entre dominios (usando una construcción que carece del puente salino E48-R138) ambos estados se observan en solución, lo que sugiere que existe una activa interconversión del dominio carboxilo-terminal [9]. Adicionalmente, el paisaje energético en estas condiciones ($\epsilon^{IF} \cong 0,5$) también muestra que las cuencas correspondientes a los estados α CTD y β CTD están conectadas mediante configuraciones de dos intermediarios obligados (I_1 e I_2) (Figura 5) [27]. Al igual que los intermediarios propuestos anteriormente [21], [25] I_1 existe en ausencia (y presencia) del NTD y conserva múltiples interacciones entre las hojas- β similares al β CTD. Por otro lado, la formación de I_2 requiere de contactos entre dominios que estabilizan la hélice α_1 , mientras que la hélice α_2 presenta cierto desenrollamiento, contrastando con lo observado previamente en ausencia del NTD [29]. Adicionalmente, un residuo de fenilalanina en el CTD (F130) se muestra responsable en gran medida de formar el núcleo 3hidrofóbico del estado β y de interactuar con el NTD en el estado α , pues una mutación *in silico* que elimina las interacciones de su cadena lateral, desestabiliza el β CTD y los intermediarios I_1 e I_2 [28].

En otro estudio sobre la proteína RfaH completa se utilizó dinámica molecular dirigida (*Targeted Molecular Dynamics*, TMD) y dinámica molecular de tracción (*Steered Molecular Dynamics*, SMD). En la primera, un conjunto de átomos del sistema inicial es dirigido hacia las coordenadas de un sistema final predeterminado [29], mientras que en la segunda se utiliza una fuerza de tracción sobre átomos seleccionados de modo de

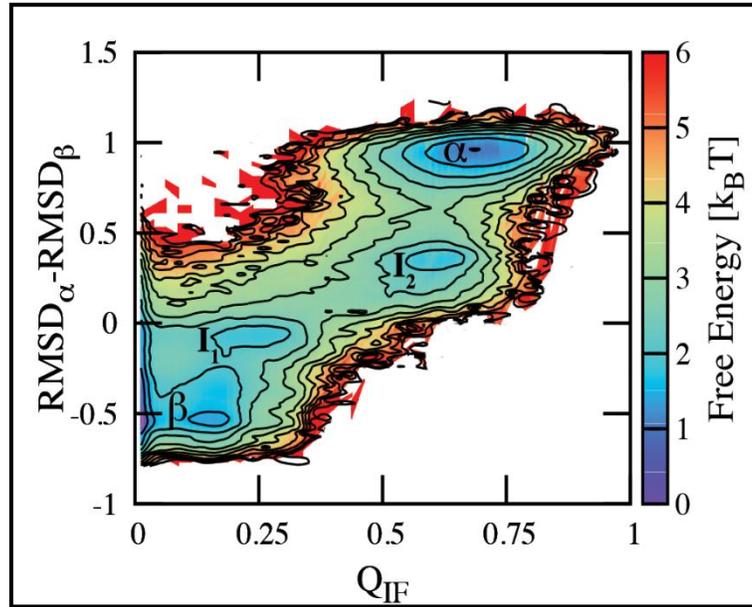


Figura 5: Paisaje energético de RfaH para la transición espontánea. Este paisaje se muestra frente al RMSD del CTD (estado α o β) y fracción de contactos de la interfaz NTD- α DCT (Q_{IF}), cuando la fuerza de los contactos entre dominios se ha reducido a la mitad en un SBM. En el gráfico se observan ambos estados y los intermediarios, teniendo I1 contactos similares a β DCT. Extraído de [29].

estudiar respuestas mecánicas del sistema [30]. Explorando la transformación entre estados α CTD y β CTD de RfaH completa utilizando TMD atómica con solvente explícito, se observa un desenrollamiento inicial de la hélice α_2 y posteriormente de la hélice α_1 [22], similar a lo observado anteriormente mediante CGMD basada en estructura [27]. Esta transición continúa con una pérdida de estructura secundaria, sin embargo, no tiene características de ovillo al azar, pues tras abandonar el estado helicoidal, el CTD mantiene un incipiente núcleo hidrofóbico, haciéndolo más similar a un glóbulo fundido. Luego, similar a lo analizado mediante MSM, se observa la formación inicial de las hebras β_2 y β_3 , para terminar en el barril- β . La conversión inversa es algo distinta, pues tras la ruptura de la conformación β ocurre el colapso y rearrreglo del núcleo hidrofóbico, que da lugar a múltiples interacciones no-nativas antes de la formación de la estructura helicoidal [22].

En el mismo artículo, se utilizó una estrategia diferente llamada Fuerzas Directoras Adaptables (*Adaptive Biasing Forces*, ABF), que permite acelerar la exploración del espacio configuracional mediante la aplicación de fuerzas opuestas de naturaleza difusiva

que reducen las barreras energéticas, de modo que sean accesibles a través de fluctuaciones térmicas [31], [32]. Tanto las simulaciones de SMD, estirando en direcciones opuestas los extremos amino y carboxilo terminal de α RfaH, así como simulaciones de ABF utilizando RMSD como coordenada de reacción, muestran que la disociación interdominio ocurre previa al desplegamiento del α CTD (Figura 6), contrastando con todos los reportes previos. Adicionalmente, estas técnicas muestran que el desenrollamiento de la hélice α_2 ocurre previo al de la hélice α_1 [22].

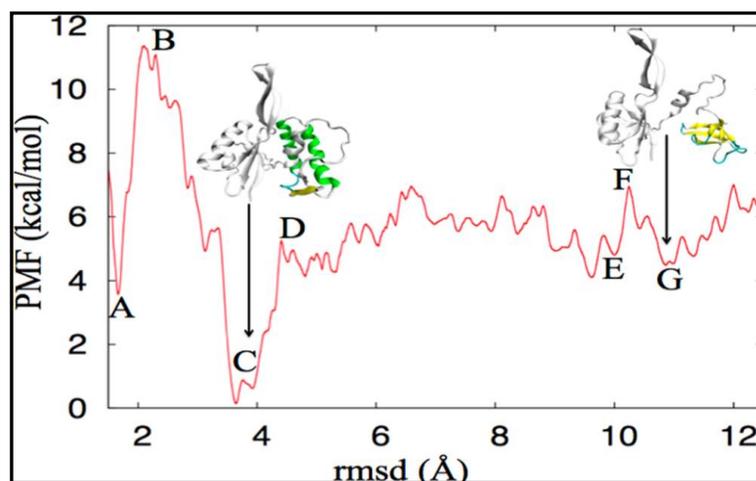


Figura 6: Panorama energético de la transición de RfaH obtenido por ABF. PMF (potential of mean force) corresponde al potencial que representa el trabajo ejercido por las fuerzas difusivas. En A se observa el estado nativo, B el estado de transición hacia el estado C cuyo α DCT está parcialmente disociado, en E, F y G se observan conformaciones similares al β DCT. Extraído de [24].

Una publicación reciente estudió el efecto de las zonas intrínsecamente desordenadas, correspondiente al bucle que conecta ambos dominios y a la sección carboxilo terminal no estructurada, en la estabilidad de RfaH. Dicho trabajo encontró una estabilización del dominio carboxilo terminal tanto en el estado α RfaH a partir de la estructura cristalográfica como en el dominio β CTD aislado [33], la cual es principalmente originada por interacciones favorables que establecen estas regiones mediante sus cadenas laterales tanto con la interfaz interdominio como con el CTD.

Estudios de molécula individual

Las simulaciones de SMD mencionadas anteriormente [22] retratan computacionalmente lo que se podría observar mediante espectroscopía de fuerza, una técnica de manipulación de molécula única que permite estudiar el plegamiento de proteínas a través del uso de fuerza como agente desnaturalizante [34]. Los diseños experimentales de molécula individual o *in singulo*, permiten estudiar el comportamiento de cada molécula (o complejo funcional) de manera aislada, para obtener información inaccesible utilizando técnicas de asamblea estadística o *in multiplo*. Un ejemplo de ello son los procesos estocásticos como pausas y retrocesos, cuya ocurrencia es apantallada por el enorme número de moléculas actuando simultáneamente, provocando que el observable represente simplemente el valor promedio del comportamiento. A nivel de molécula individual es posible entonces conocer la frecuencia con la que ocurren estos eventos y, dado un tamaño muestral lo suficientemente grande, replicar el comportamiento observado en asamblea estadística. Un instrumento que permite realizar espectroscopía de fuerza son las pinzas ópticas, que permiten realizar tracción mediante el control por atrapamiento óptico de esferas micrométricas de poliestireno conectadas a la proteína de estudio [35]. Estas esferas normalmente están funcionalizadas con algún anticuerpo que permita su unión a “mangas” de ADN modificadas, para que éstas a su vez se unan a la proteína [34], [36]. La naturaleza flexible del ADN permite comunicar la fuerza ejercida desde la microesfera hacia la proteína, la que está unida al ADN mediante puentes disulfuro. Esta unión continua de los componentes entre ambas esferas permite que el sistema se comporte esencialmente como un elástico, ya que el ADN aumenta su extensión gradualmente acorde se ejerce mayor tracción desde las microesferas. Sin embargo, los fenómenos de plegamiento de proteínas son altamente cooperativos, por lo que las transiciones estructurales ocurren por cambios violentos en el grado de compactación de la cadena peptídica. Por ello, cuando ocurren estos eventos, se evidencian por cambios bruscos en la extensión del sistema (Figura 7) [37]. Estos cambios son estudiados principalmente de dos maneras:

- i) realizando una rampa de fuerza, que consiste en separar las microesferas a una velocidad constante (análogo a SMD), por lo que aumenta gradualmente la fuerza sobre

el sistema (para superar el potencial elástico del ADN). Esto permite observar las fuerzas a las que ocurre el desplegamiento y replegamiento, y extrapolar las energías al equilibrio sin perturbación mediante el teorema de Crooks [38], [39].

- ii) Realizando una pinza de fuerza: que consiste en generar una perturbación a fuerza constante en el sistema para observar cómo se van poblando, acorde a su estabilidad, los distintos estados de plegamiento de la proteína en el tiempo. Gracias a que las mediciones son respecto al tiempo y distancia, podemos conocer tanto la energía libre del proceso como las constantes cinéticas involucradas, gracias al modelo de Bell-Zhurkov [40], [41]. Esto presenta una gran ventaja experimental respecto a estudios *in múltiplo*, debido a que en las trayectorias se puede observar cada estado de manera independiente, por lo que es posible identificar intermediarios de manera directa, diferenciarlos de aquellos fuera de la vía de plegamiento, e incluso diferenciar entre estados estructuralmente similares mediante información cinética en un mismo experimento [42].

Como se mencionó previamente, la transición entre los estados α CTD a β CTD ha sido explorada computacionalmente usando fuerzas difusivas (ABF, Figura 6), indicando la ruptura de la interacción entre la horquilla- α y el NTD como necesaria para el proceso de transformación, pero el panorama energético en tal experimento no se observa claramente definido lejos de la conformación α CTD, seguramente por las altas

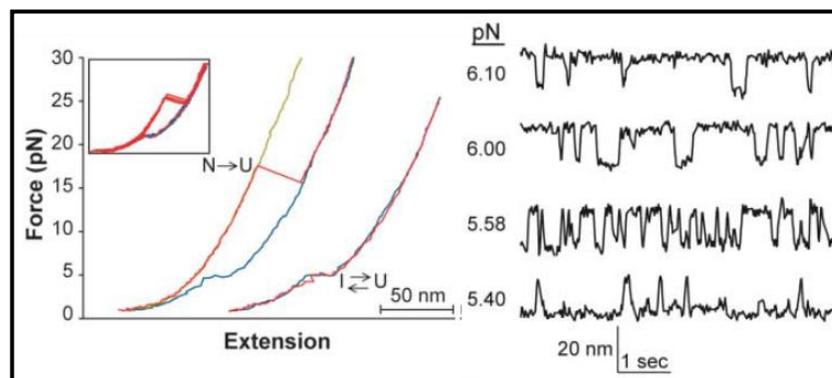


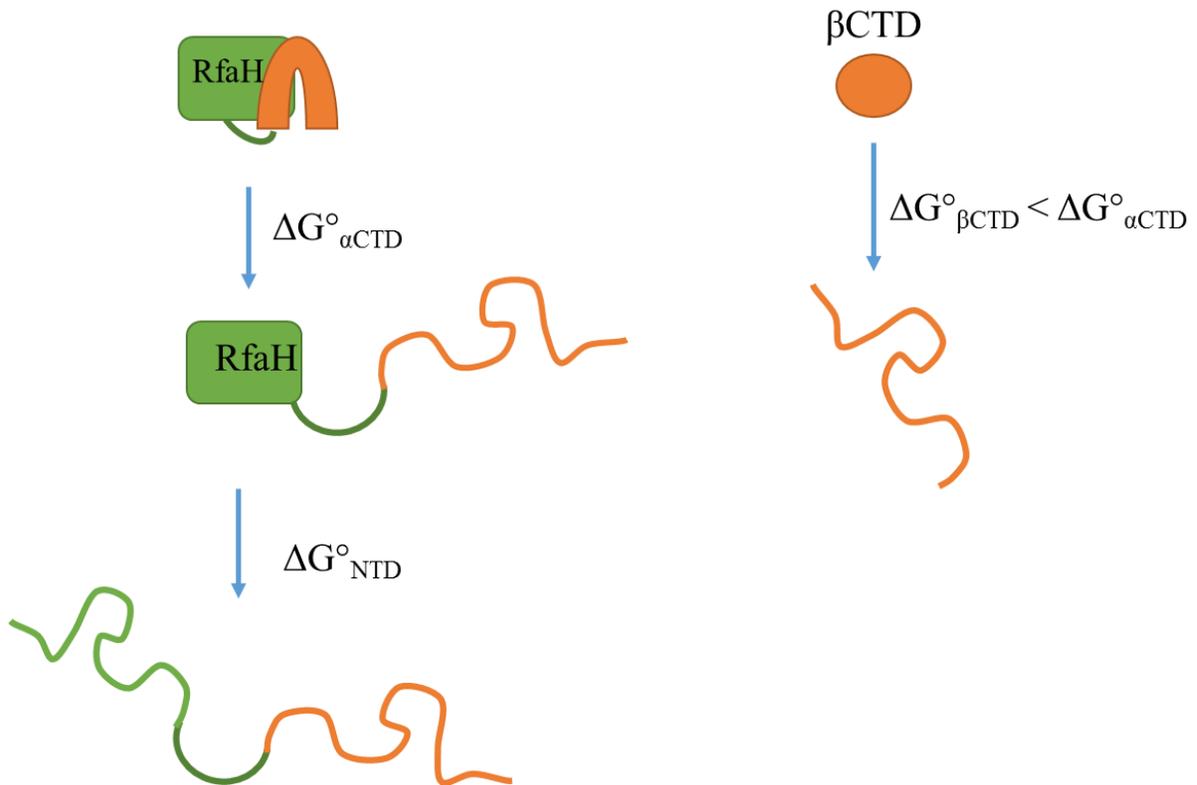
Figura 7: Espectroscopía de fuerza por pinza o rampa de fuerza. (izq) traza experimental de rampa de fuerza. En gris se observa el estiramiento del ADN solo, en rojo desplegamiento de la proteína ribonucleasa H de *E. coli*, y en azul su replegamiento en régimen de alta y baja fuerza. Se observa que el estiramiento a altas fuerzas induce el desplegamiento de la proteína, mientras que a bajas fuerzas o replegando se observa una transición distinta. El inserto muestra el mismo experimento en multiplicado. (der) Experimento de pinza de fuerza, se observa que, a mayores fuerzas, el tiempo que la proteína visita el estado compacto disminuye, evidenciando la desestabilización de la estructura inducida por fuerza. Modificado de [37].

fluctuaciones de RMSD luego de su separación [22]. Esto es parcialmente distinto a lo observado al suavizar el panorama energético mediante SBM, en el cual ambos estados nativos se encuentran bien definidos, si bien existe consistencia en la necesidad de romper las interacciones interdominio para facilitar el proceso de transformación [27]. Considerando las discrepancias predichas computacionalmente durante el estudio del plegamiento de RfaH, es difícil aseverar si la transformación espontánea entre ambos estados en un panorama energético suavizado (i.e. mediante SBM) involucra vías comunes al desplegamiento mecánico. Esto es, porque las simulaciones de desplegamiento mecánico utilizando campos de fuerza convencionales suelen ejecutarse a velocidades que sobrepasan en órdenes de magnitud la experimental [43]. Por otro lado, las simulaciones con SBM atomísticos [44] generan un re-escalamiento de la unidad de tiempo tal que permite explorar velocidades contrastables a las empleadas experimentalmente [45].

Por otra parte, la presencia de interacciones inter-dominio en α RfaH y la ausencia de tales interacciones en el estado nativo de diversas estructuras de NusG sugieren que en un experimento de tracción de los extremos de la proteína RfaH completa, su CTD debería disociarse y desplegarse simultáneamente y sin perturbar la estructura del NTD. Experimentos de tracción con fuerza permitirían elucidar cómo es la estabilidad del dominio carboxilo-terminal de RfaH afectada por la presencia del dominio amino-terminal y proveer de información experimental de cómo podría ocurrir este proceso en solución. Debido a esto se propone la siguiente hipótesis de investigación:

I. Hipótesis

El desplegamiento mecánico de la proteína RfaH completa sigue un mecanismo de tres estados, en el que la diferencia de energía libre del desplegamiento del dominio carboxilo terminal en el estado α es mayor que el mismo aislado en el estado β .



Esquema 1: Mecanismo hipotético de desplegamiento mecánico de la proteína RfaH. En verde se muestra el dominio Amino-terminal (NTD) y en naranja el dominio Carboxilo-terminal (CTD). A la izquierda se representa la horquilla helicoidal correspondiente al α CTD, y a la derecha el barril β -plegado que corresponde a la estructura del β CTD aislado en solución.

II. Objetivo General

Determinar computacionalmente los parámetros termodinámicos del desplegamiento mecánico de la proteína RfaH y del dominio carboxilo terminal aislado.

Objetivos Específicos

1) Estudio global del desplegamiento mecánico utilizando SBM

Con el fin de generar una predicción del plegamiento de RfaH, primero y a través de modelos nativo-céntricos se estudiará cómo responden modelos basados en estructura a experimentos de estiramiento equivalentes a manipulación de molécula individual mediante pinzas ópticas o AFM. Todas las proteínas estudiadas a través de esas técnicas y disponibles en el PDB serán simuladas, y se evaluarán parámetros como fuerza de ruptura o F_m , eventos de ruptura o *rips* y cómo cambia la fuerza frente a cambios en la velocidad de estiramiento y cambios de temperatura, comparando los resultados computacionales con su contraparte experimental.

2) Caracterización del plegamiento de RfaH mediante SBM

Conociendo las limitaciones y fortalezas de los modelos basados en estructura frente a su contraparte experimental de molécula individual, se procederá a caracterizar el plegamiento de RfaH en ambos estados α - y β -plegados. Para ello se realizará el desplegamiento y replegamiento computacional de RfaH en condiciones equivalentes a las experimentales, con el fin de i) determinar las vías de plegamiento de ambos estados de RfaH, y ii) estimar la energía involucrada en el plegamiento reversible del CTD. Esto se realizará a través de SMD desde el estado nativo para desplegar, y desde estados desplegados (o intermediarios en caso de existir) para lograr el replegamiento, estimando directamente la energía requerida a través del teorema de fluctuaciones de Crooks.

3) Análisis termodinámico de RfaH a través de CCR

Para estimar la energía al equilibrio de transformación entre los estados α - y β -plegados de la proteína RfaH completa se utilizará dinámicas de confinamiento a través del método CCR, que predice con alta exactitud la energía libre de diferentes estados en un entorno de dinámica convencional, y que además es comparable con datos de intercambio hidrógeno-deuterio disponibles en el laboratorio.

III. Teoría y Métodos

1) Computacionales

1.a) **Modelos Basados en Estructura** (*Structure-Based Models*, SBM): Para éstas dinámicas se utilizó modelos basados en estructura en una versión modificada de Gromacs [26]. Para la simulación de cada proteína, se utilizó la estructura depositada en el *Protein Data Bank* (PDB) correspondiente a la misma, y se definió como mínimo energético descrito a nivel atómico [46]. Para llevar a cabo la simulación, se utiliza una función de energía potencial modificada, donde la energía de distancias, ángulos y diedros atómicos corresponden a funciones armónicas centradas en los valores para cada grupo atómico observado en la estructura cristalina. Asimismo, funciones no enlazantes presentan un potencial de contacto gaussiano obtenidas mediante el algoritmo de sombra [47] para todos aquellos átomos presentes en la estructura cristalográfica cuyas distancias se encuentran en un radio corte de 6 Å y un radio de sombra de 1 Å. Adicionalmente, se utiliza un potencial repulsivo Lennard-Jones 12 a todos los pares atómicos que no se contacten en la estructura cristalina de acuerdo al criterio de distancia y sombra anteriormente indicado [48].

Debido a que los contactos provienen de estructuras en solución, el efecto de la solvatación está implícita en la simulación. El paisaje energético suavizado por estos potenciales y la ausencia de interacciones con agua permite que las simulaciones en estos modelos exploren cambios conformacionales de escalas temporales órdenes de magnitud más rápido que sus contrapartes convencionales [49]. Una calibración reciente utilizando éste mismo diseño computacional de SBM atómicos estimó una relación entre tiempo computacional y tiempo real de 1:50.000, es decir, una aceleración intrínseca de 4 órdenes de magnitud [45].

1.b) **Dinámica Molecular de Estiramiento** (*Steered Molecular Dynamics*, SMD): Estas simulaciones se realizaron utilizando SBM y una versión modificada de Gromacs 4.5 [50]. En estas se realizaron simulaciones de estiramiento a velocidad constante, en las que se utiliza una función de energía potencial modificada del tipo $U_{\text{pot}} = U_{\text{SBM}} + U_{\text{est}}$ [51],

donde el segundo corresponde a un potencial armónico dependiente del tiempo y de la distancia entre los residuos terminales x , definido como

Ecuación 2 Potencial de estiramiento

$$U_{est} = K_s[(V_p \cdot t + x_0) - x]^2$$

donde K_s corresponde a la constante de resorte o dureza del potencial en $\text{kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$, V_p a la velocidad de estiramiento en $\text{nm}\cdot\text{ps}^{-1}$, t el tiempo en ps y x_0 la distancia inicial entre los residuos terminales en nm. Para estas simulaciones se utilizó una constante de resorte de $3,6 \text{ pN}\cdot\text{nm}^{-1}$, que corresponde a un valor de dureza de trampa intermedio entre dispositivos de pinzas ópticas ($< 1 \text{ pN}\cdot\text{nm}^{-1}$) y AFM ($> 10 \text{ pN}\cdot\text{nm}^{-1}$). En literatura se ha explorado la relación entre fuerzas de desplegamiento experimentales y obtenidas por SMD utilizando SBM de carbono alfa, encontrando una cercana correlación entre ambos [49], [52], [53]. Sin embargo, en la gran mayoría de dinámicas computacionales reportadas se utiliza una rampa de fuerza ($K_s \cdot V_p$) órdenes de magnitud mayor que la experimental, debido al costo computacional que requiere evaluar explícitamente escalas de tiempo macroscópicas [43].

En otra publicación más reciente, se aprovechó de la estrecha relación entre la fuerza de ruptura (F_m) y V_p dada inicialmente por el modelo fenomenológico de Bell [40] que propone que en regímenes experimentales, existe una dependencia lineal entre F_r y $\ln(V_p)$. Utilizando esto, los autores realizaron una extrapolación desde fuerzas computacionales en regímenes cercanos a $\text{mm}\cdot\text{s}^{-1}$, hacia su equivalente al utilizar valores de V_p experimentales de cientos a miles de $\text{nm}\cdot\text{s}^{-1}$ [54]. Este método mejoró significativamente la correlación entre las fuerzas observadas y las experimentales. Es por esto que, para realizar los estiramientos atomísticos, utilicé la calibración de tiempo reportada previamente para SBM [45], de modo de estimar V_p computacional que corresponda al V_p experimental para cada una de las 35 proteínas estudiadas. Estas simulaciones se realizaron utilizando un paso de tiempo de 0,2 fs, sin definición de bordes periódicos, utilizando temperaturas reducidas desde 0,58 a 1,16 unidades (en estos modelos, las proteínas se pliegan entre 0,83 y 1 unidades reducidas). Los replegamientos

de ambos estados de RfaH se realizaron con el mismo código, invirtiendo el vector de estiramiento a partir de un estado total o parcialmente desplegado.

1.d) **Confinamiento-Conversion:** Éste corresponde a un método de cálculo de energía libre por integración termodinámica de simulaciones de dinámica molecular [55]. Confinamiento se refiere a restringir las posiciones atómicas globales de los átomos de la proteína. Al utilizar simulaciones con restricciones crecientes a las posiciones, generamos simulaciones con exploración configuracional cada vez más reducida, hasta que eventualmente los movimientos de cada átomo de la proteína se limitan a su vecindad próxima, con amplitudes del orden de mÅ. En esta condición “confinada” los átomos no presentan movimientos amplios, sino pequeñas oscilaciones armónicas en su vecindad acorde a la función de energía potencial (Figura 8).

Es posible conocer la energía requerida para lograr el estado confinado mediante el cálculo de las fluctuaciones globales definidas como

Ecuación 3 Fluctuaciones cuadráticas globales

$$\chi_k^2 = \langle RMSD \rangle^2 \cdot N$$

donde χ_k corresponde a las fluctuaciones globales atómicas (Å^2) con una constante de restricción k armónica ($\text{Kcal}\cdot\text{mol}^{-1}\cdot\text{Å}^{-2}$), $\langle RMSD \rangle$ corresponde al valor de la desviación cuadrática media de las posiciones atómicas de la proteína o RMSD en el ensamble promedio, y N el número de átomos totales de la proteína, con los que se calcula el RMSD [56]. Al realizar dinámicas de confinamiento a través de distintos valores de k es posible conocer pares $[\chi_k, k]$, cuya área bajo la curva define la energía de confinamiento. Una publicación reciente estableció la dependencia del RMSD en un mínimo energético respecto a una constante de restricción de posiciones como una función exponencial simple, permitiendo calcular a través de integración trapezoidal la energía de confinamiento desde un macroestado Ω como:

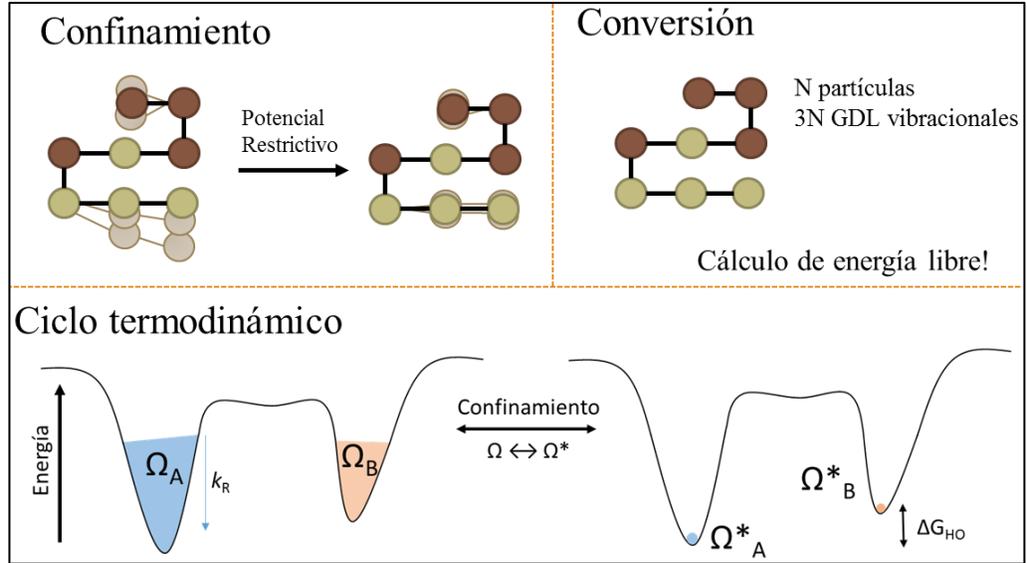


Figura 8: Representación gráfica del proceso de confinamiento. El paso de confinamiento se realiza a través de potenciales restrictivos que reducen la desviación de posición media durante la simulación. El paso de conversión requiere conocer la frecuencia de cada uno de los $3 \cdot N$ grados de libertad (GDL) vibracionales del sistema para computar la energía libre. El ciclo termodinámico permite conocer la diferencia de energía libre entre dos o más macroestados definidos por pozos energéticos a partir de sus componentes confinadas y vibracionales.

Ecuación 4 Energía libre de estado confinamiento

$$\Delta F_{\Omega \rightarrow \Omega^*} = \int_{k_i}^{k_f} \chi_k dk = \int_{k_i}^{k_f} a k^b dk = \frac{a k^{b+1}}{b+1} \Big|_{k_i}^{k_f} = \frac{\chi_k k}{b+1} \Big|_{k_i}^{k_f}$$

Donde a y b son valores constantes para una función exponencial (i.e. $y = ax^b$). Por otro lado, el comportamiento oscilatorio observado en dinámicas confinadas es muy bien modelado por análisis de modos normales [57], que calcula los diferentes componentes armónicos de los movimientos de conglomerados atómicos, y permite calcular la frecuencia con la que ocurren. A partir de estas frecuencias es posible conocer analíticamente la energía libre, considerando al sistema como compuesto por N partículas con $3N-6$ grados de libertad vibracionales. Esto permite computar la función de partición z del sistema y su energía libre como:

Ecuación 5 Energía de un oscilador armónico

$$z_{\Omega^*} = e^{-\beta \langle E_{\Omega^*} \rangle} \cdot \prod_i^{3N-6} (\beta h \nu_i)^{-1}; F = -\beta^{-1} \ln(z)$$

Donde $\langle E_{\Omega^*} \rangle$ corresponde a la energía potencial en el mínimo confinado, h la constante de Planck, β al inverso del producto entre la constante de Boltzmann y la temperatura absoluta, y ν la frecuencia de cada uno de los $3N-6$ modos normales.

Este análisis permite, a partir de sucesivas simulaciones de confinamiento, obtener la energía libre absoluta del macroestado tal Ω como [58]:

Ecuación 6 Energía libre por confinamiento

$$F_{\Omega} = \Delta F_{\Omega \rightarrow \Omega^*} + F_{\Omega^*}$$

Donde F_{Ω} corresponde a la energía libre del macroestado en $\text{kcal} \cdot \text{mol}^{-1}$, $\Delta F_{\Omega \rightarrow \Omega^*}$ la energía de confinar el macroestado a un microestado en $\text{kcal} \cdot \text{mol}^{-1}$, y F_{Ω^*} la energía libre del microestado considerado como un oscilador armónico en $\text{kcal} \cdot \text{mol}^{-1}$. Calcular la energía libre de macroestados confinables, es decir con un profundo mínimo energético, permite conocer diferencias absolutas de energía entre estados configuracionalmente distintos, como es el caso de α - y β RfaH.

Para realizar las simulaciones de confinamiento se utilizaron 23 constantes de restricciones globales desde $0,000025$ ($\text{kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$), duplicando su valor hasta llegar a 104.8 ($\text{kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$). Estas simulaciones se realizaron en solvente implícito utilizando el modelo de pares generalizado de Born de Hawkins, Cramer y Truhlar [59], junto con algoritmo SHAKE [60] para restringir la posición de hidrógenos. Cada simulación de confinamiento corresponde a un monoplificado de 30 ns de extensión.

1.e) **Minimización de Energía:** Para realizar el confinamiento se requiere de estructuras suficientemente minimizadas de modo que el potencial de confinamiento siempre realice trabajo a favor del potencial y así represente bien el comportamiento de los modos normales [55], [56]. Debido a que la proteína RfaH posee dominios conectados

por un *loop* flexible, cuya estructura no se observa a través de cristalografía, se utilizó Rosetta (PMID: 27490953) para modelar tal región. La rutina denominada Loop Refinement de Rosetta permite reconstruir bucles a partir de una biblioteca de fragmentos generada con la secuencia de RfaH y una base de datos de estructuras conocidas. La selección aleatoria de fragmentos en el código de Rosetta permite generar múltiples configuraciones minimizadas con resolución atómica [61]. 500 modelos fueron generados para la región que comprende 15 residuos en α RfaH y 10 en β RfaH. La configuración de menor energía fue evaluada en una simulación de 20 ns en solvente explícito en AMBER 16. Esta estructura fue posteriormente minimizada mediante gradiente conjugada hasta fluctuación de RMSD (Δ RMS) $< 10^{-5}$, y posteriormente por minimización Newton-Raphson hasta Δ RMSD $< 10^{-12}$ [56].

1.f) **Energía libre por residuo:** Una de las grandes ventajas del protocolo de confinamiento es poder estimar el aporte de energía de cada residuo al confinamiento estableciendo que

Ecuación 7 Descomposición de energía libre

$$(R_k - R_0)^2 = \sum_1^R (r_k - r_0)^2$$

señalando que la diferencia cuadrática de las coordenadas atómicas de una proteína que posee un número R de residuos bajo una constante de confinamiento (R_k) respecto a las posiciones minimizadas (R_0) es igual a la suma de las diferencias cuadráticas individuales de cada residuo entre los estados confinados (r_k) y minimizado (r_0) [62].

Por otro lado, hasta la fecha no hay una forma fácil de descomponer los componentes de los modos normales en modos y frecuencias por residuo. Sin embargo, es posible estimar la energía libre por residuo considerando que:

Ecuación 8 Energía libre de Helmholtz

$$\Delta A(T, V) = \Delta U - T\Delta S$$

donde $A(T, V)$ corresponde al potencial de energía libre de Helmholtz, U la energía interna, T la temperatura absoluta y S la entropía. Debido a que el origen de la entropía en el estado confinado es principalmente vibracional, y ambos estados están compuesto del mismo número de osciladores armónicos, la diferencia del término entrópico es despreciable [62]. Es por esto que evaluar la diferencia de energía interna por residuo entre ambos estados corresponde a una aproximación de la diferencia de energía por residuo entre ambas configuraciones. La descomposición de la energía libre se realizó mediante el módulo *Molecular Mechanics Poisson-Boltzmann Surface Area* (MMPBSA) de AMBER con la simulación correspondiente a la mayor constante de restricción.

1.g) **Cálculo de trabajo al equilibrio:** En los experimentos de estiramiento es común observar histéresis entre plegamiento y desplegado, evidenciando que el sistema no alcanza el equilibrio a medida que aumenta la perturbación mecánica. Desarrollos de física estadística han planteado marcos de trabajo que permiten extraer propiedades de equilibrio a partir de datos fuera del equilibrio, como el teorema de fluctuaciones de Crooks [63]:

Ecuación 9 Teorema de fluctuaciones de Crooks

$$e^{\beta(W-\Delta F)} = \frac{p_f(+W)}{p_r(-W)}$$

Donde W corresponde al trabajo externo realizado sobre el sistema y F la energía libre del proceso al equilibrio. Los términos el numerador a la derecha del teorema de Crooks corresponde a la probabilidad de que ocurra el proceso directo (desplegado) respecto al trabajo externo, y el denominador la probabilidad de que ocurra el proceso inverso (plegado) respecto al trabajo del sistema. Utilizando estas ecuaciones es posible conocer la energía libre del plegamiento a partir de la distribución de trabajos dado muestreo suficiente. El cálculo computacional de trabajos se realizó mediante el producto de la fuerza de ruptura o plegamiento, y la extensión involucrada en el estiramiento de

una cadena peptídica correspondiente al número de residuos desplegados. Este cálculo se realizó para 27 simulaciones independientes de α RfaH y β CTD.

2) Experimentales

2a) **Construcción de RfaH:** con el fin de realizar espectroscopía de fuerza a nivel de molécula individual, se requiere construir una doble mutante de RfaH donde residuos terminales fueran reemplazados por cisteína. Esta construcción se realizó mediante mutagénesis por PCR en el plásmido pIA777 que contiene RfaH, generosamente donado por Irina Artsimovitch de Ohio State University, Columbus, Ohio, EEUU. Esta construcción fue publicada por ella [64] y corresponde a un vector de expresión que codifica para resistencia a kanamicina y expresa RfaH bajo el promotor inducible Lac. La proteína RfaH posee un sitio de corte por la proteasa TEV en el bucle interdominio (secuencia ENLYFQG) y un repetido de hexa-histidina posterior al dominio carboxilo terminal:



La integridad de esta construcción, RfaH Q2C, R160C (en adelante RfaH-2C), fue confirmada por secuenciación del plasmidio (Macrogen USA, Rockville, Maryland, EEUU) y almacenada en colonias de *E. coli* DH5- α quimiocompetentes.

Debido a que la proteína RfaH posee dos cisteínas internas (posición 9 y 28), se realizaron nuevas construcciones con la eliminación de una o ambas cisteínas. Adicionalmente, se generó también la mutante Q2GC que podría ayudar a favorecer la reacción de la cisteína para protocolos posteriores.

2b) **Purificación y mutación de plasmidios:** Las construcciones que se realizaron en el plasmidio reportado pIA777 (sección 2a). Las mutantes de RfaH silvestre se realizaron mediante mutagénesis por PCR en base al protocolo descrito en el kit Genetailor (Invitrogen Corp, Carlsbad, California, EEUU), utilizando como molde el plasmidio mencionado. Para realizar este protocolo se requiere utilizar dos partidores para la PCR que posean una zona de solapamiento, inmediatamente fuera de esta zona de solapamiento, el partidor *forward* o 5'→3', posee un codón mutagénico seguido de una zona de

extensión, lo que permite que una enzima como la polimerasa de ADN Pfu amplifique incorporando la mutación.

Las reacciones de PCR se realizaron en un volumen de 50 μ L de una mezcla en agua estéril compuesta por: partidor a 0,1 μ M, solución amortiguadora 10x Pfu Ultra II (Agilent Technologies, Santa Clara, California, EEUU) a una dilución 10% v/v, enzima Pfu Ultra II Fusion HS ADN (Agilent Technologies) a 0,05 μ L, y desoxinucleótidos (Agilent Technologies) a 0,2 uM. La reacción de PCR se llevó a cabo utilizando un programa estándar de amplificación: inicio a 95 °C por 2 minutos, 18 ciclos de i) denaturación a 95 °C por 30 segundos, ii) apareamiento en gradiente de 55-68 °C dependiendo de la temperatura de fusión de los partidores por 1 minuto, iii) extensión a 72 °C por 90 segundos, y al terminar los ciclos de PCR una extensión final de 10 minutos.

Los productos de PCR son analizados por inspección visual en electroforesis en gel de agarosa al 0,5% m/v, para digerir posteriormente el ADN parental mediante la enzima de restricción DpnI (Thermo Fischer Scientific Inc, Walwham, Massachussets, EEUU).

2c) Expresión y purificación de RfaH doble-mutante de cisteína: El plásmido purificado que contiene RfaH-2C fue transformado químicamente a *E. coli* BL21 para evaluar su expresión. Se realizó una prueba de expresión creciendo las bacterias en medio LB estándar [65] con kanamicina 35 μ g ml⁻¹. Para todas las proteínas utilizadas se realizó pruebas de expresión en condiciones cercanas a las utilizadas para este plasmidio [64]:

Temperatura	C. Inductor (IPTG)	Turbidimetría a 600 nm	Tiempo de inducción
25 °C	0,2 mM	0,5 Abs	2, 4, 6 y 18 Horas
25 °C	0,2 mM	1,0 Abs	2, 4, 6 y 18 Horas
35 °C	0,2 mM	1,0 Abs	2, 4, 6 y 18 Horas

La expresión se midió tras centrifugar 10 ml en cada condición a 10.000 g x 15 min, posterior lisis por sonicación y centrifugación a 18.000 g por 10 minutos. La pella es

resuspendida en solución de urea 6 M y la masa de proteína tanto en pella como sobrenadante es cuantificada por test de Bradford [66] y evaluados por electroforesis reductora y desnaturalizante (SDS-PAGE) a través de gel de poliacrilamida al 17% A/BA.

La purificación posterior se realiza tras crecer 500 a 1.000 ml del clon bacteriano en medio apropiado, LB para BL21 o TB [67] para C41, y tras llevar a cabo la condición de inducción, se centrifuga a 15.000 g por 15 minutos, lisis por sonicación, y posterior centrifugación a 28.000 g por 20 minutos para recuperar (preferencialmente, dependiendo del test de inducción) el sobrenadante.

Este lisado se purifica por cromatografía de afinidad a catión metálico inmovilizado, en este caso níquel, utilizando una columna HisTRAP (GE Healthcare Bio-Sciences, Pittsburgh, Pennsylvania, EEUU) equilibrada en 150 mM NaCl, TrisHCl pH 7,8, glicerol 5%, Imidazol 30 mM, DTT 1 mM y PMSF 0,1 mM, lavando y finalmente eluyendo en gradiente del mismo amortiguador con concentraciones crecientes hasta 500 mM de imidazol, acorde a instrucciones del proveedor.

2d) **Ensayo de reactividad de cisteínas:** Debido a que la proteína RfaH tiene 4 cisteínas totales, se evaluó su reactividad a través del ensayo de Ellman [68] y mediante DTDP [36]. Para el ensayo de Ellman se realizó una curva de calibración del reactivo DTNB 100 μ M con concentraciones crecientes desde 1 a 10 μ M en TrisHCl pH 8,1, midiendo un duplicado de R2C 2 μ M.

También se evaluó la modificación por DTDP, que se oxida modificando covalentemente residuos de cisteína expuestos. Esta molécula debe ser disuelta en solventes apolares como ACN o DMSO; sin embargo, ambos solventes en concentraciones cercanas al 1% inducen la agregación de R2C y RfaH silvestre (datos no mostrados). Por esto se utilizó EtOH, que por sí solo no acelera la agregación de R2C (hasta concentraciones del 2%). Sin embargo, incluso en estas condiciones se observa la rápida agregación de RfaH al añadir concentraciones iguales o superiores a 1 mM de DTDP. Posteriormente se realizó cromatografía de exclusión molecular utilizando la columna Superdex 75 10/300 GL de 24 mL (GE Healthcare Bio-Sciences) montada en un equipo HPLC Waters 1525 acoplado a un detector UV dual Waters 2487 (Waters Corp, Milford, Massachussets, EEUU) programado para medir absorbancia a dos longitudes de

onda (220 y 280 nm), para evaluar la elución de la muestra en amortiguador 100 mM NaCl y 50 mM TrisHCl pH 7,2. Se recolectó el pico correspondiente al tamaño de RfaH dada una previa calibración con estándares de masa molecular y posteriormente se realizó análisis por espectrometría de masas MALDI-TOF. Para ellos, se mezcló 1 μ l de la muestra con 1 μ l de matriz de ácido sinapínico en una placa porta muestra Micro Scout (Bruker Daltonics Inc, Massachussets, EEUU). Se realizó el espectro en un equipo MALDI-TOF Microflex en modo ión positivo mediante detección lineal, en una calibración con estándar externo correspondiente a una mezcla de proteínas de masas de 6 a 17,5 kDa.

2e) **Síntesis de mangas de ADN:** Las mangas de ADN fueron sintetizadas acorde a lo publicado en literatura [69][36]. Brevemente, se utiliza el plasmidio pGEMEX que posee un segmento que es amplificado por dos partidores. Sin embargo, se utiliza dos partidores sentido, uno con una modificación por digoxigenina y el segundo por una modificación con biotina, y siempre un partidor antisentido con una modificación con sulfhidrilo. Esto permite generar dos mangas de ADN lineales de \approx 550 pb, que poseen un extremo sulfhidril-reactivo y un biomarcador molecular en el otro. Estas mangas son las responsables de conectar las cisteínas expuestas en la proteína a las esferas de poliestireno funcionalizadas con streptavidina y anticuerpo anti-digoxigenina. Una vez realizada la amplificación por PCR, los productos son purificados mediante precipitación (QIAGEN Inc., Germantown, Maryland, EEUU), y posteriormente cuantificados y concentrados por filtración y por centrifugación.

IV. Resultados

Objetivo 1: Estudio global del desplegamiento mecánico utilizando SBM.

Correlación entre estiramientos experimentales y simulaciones SBM

Hasta la fecha sólo un trabajo ha descrito el comportamiento de la proteína completa RfaH frente al uso de fuerzas de tracción computacionalmente [22]; sin embargo, en él se observa que las hélices del CTD se disocian del NTD previo a su desplegamiento, contradiciendo la estabilidad observada en múltiples instancias de dinámica molecular convencional [20], [21], [25]. Una de las posibles causas expuestas en el trabajo para aquel comportamiento, es la velocidad de estiramiento utilizada de $2 \cdot 10^6 \text{ nm} \cdot \text{s}^{-1}$, cerca de un millón de veces más alta que la utilizada normalmente en AFM, considerando que su constante de resorte de $21 \text{ pn} \cdot \text{nm}^{-1}$ está en el rango de aquel diseño experimental.

Para poder estudiar cómo es el comportamiento de RfaH frente a fuerzas mecánicas, se diseñó el mismo experimento de estrés mecánico, pero utilizando modelos basados en estructura [44], [46], que permiten lograr velocidades de estiramiento más cercanas a las utilizadas experimentalmente, permitiendo observar fenómenos de plegamiento y desplegamiento en tiempos accesibles computacionalmente. Debido a que los modelos basados en estructura utilizan un campo de fuerza y Hamiltoniano modificado, las unidades con las que se llevan a cabo estas simulaciones no son equivalentes en magnitud a su contraparte convencional, requiriendo entonces de una calibración previa en función de observaciones experimentales.

Con el fin de poder predecir la fuerza y trayectoria de desplegamiento mecánico de RfaH, se utilizó estos modelos basados en estructura para caracterizar el desplegamiento mecánico de proteínas reportadas en la base de datos PDB (un requisito para llevar a cabo estas simulaciones) y que han sido estudiadas mediante espectroscopía de fuerza, para realizar dinámica molecular de estiramiento o *Steered Molecular Dynamics* (SMD). Realizando una búsqueda exhaustiva de la literatura y PDB, se simuló todas las proteínas cuyas curvas de desplegamiento mecánico por las técnicas mencionadas han sido

reportadas, y que poseen al menos una estructura en el PDB, dando un total de 35 proteínas con tamaños entre 80 y 600 residuos (Tabla 1).

Tabla 1: Proteínas estudiadas computacionalmente (al 8/9/2016). De izquierda a derecha: nombre genérico de proteína, código de acceso PDB, Fuerza de ruptura media reportada experimentalmente, Fuerza de ruptura media observada durante la simulación con SBM atómicos, Relación entre los rips observados en la simulación respecto a los observados en la traza experimental, Orden de contacto absoluto, fuerza simulada al ponderar por el orden de contacto absoluto [70], Referencia del cristal, Referencia del estiramiento.

Proteína	PDB	F_{ex} p pN	F_{obs} C·n m ⁻¹	Rip s S/E	AbsC O	F_{sim} C·n m ⁻¹	Referencia
MHC C1	1a1n	146	60,8	2	16,133	980	[71][72][73]
Barstar	1bta	20	47,2	2	10,88	513	[74] [75]
CSP	1g6p	50	28,1	1	11,687	365	[76][77]
Laminin	1ifr	56	44,5	2	18,974	533	[78][79]
Hp35	1unc	12	29,1	1	9,94	442	[80][81]
LBP apo	1usg	76	57,0	1	24,17	703	[82][83]
M-crystallin	2k1w	90	26,0	2	12,8	729	[84][85]
Ankirina	2qyj	27	27,0	2	8,33	216	[86][87]
Gelsolin		23.					[88][89]
	3ffn	9	44,1	1	12,7	342	
GolB	4y2i	165	18,0	2	12,5	551	[90]
Top7	1qys	150	74,0	1	9,91	178	[91][75][92]
B-Catenina	1qz7	95	32,0	2	9,4	695	[93][94]
NCS-1	2lcp	6.5	21,0	1	8,8	281	[95][96]
Spectrin	1aj3	68	34,0	1	8,15	171	[97][98]
Cohesin-2	1anu	214	76,0	1	27,29	927	[99][53]
Cohesin-2	1aoh	480	87,3	1	29,20	2219	[100][53]
T4-Lysozyme	1b6i	64	22,0	1,5	11,48	1001	[101]
Barnasa	1bni	70	30,0	3	12,25	269	[102][103]
Barnasa	1bnr	70	23,0	3	12,78	383	[102][103]
Barnasa	1bny	70	27,3	1	13,96	321	[102][103]
Synaptotagmin							[104][105]
	1byn	60	32,5	1	20,79	567	
Calmodulin	1cfc	20	30,5	2	8,02	260	[106][105]
Synaptotagmin							[107][105]
	1dqv	60	25,0	1	20,86	636	
GFP	1emb	104	43,0	4	27,36	683	[108][109]
Fibronectin ₁₀							[110][111][112]
	1fnf	75	63,7	1	16,94	728	[110][111][112]
Hep-Int-BD FN							[113][112]
	1fnh	89	43,0	1	17,01	1083	
Titin	1g1c	230	62,5	1	17,82	933	[114][115]
Cohesina	1g1k	425	95,0	1	27,84	1739	[100][53]
GB1	1gb1	190	94,9	1	9,58	910	[116]
Prot L	1hz5	152	113,4	1	9,74	924	[117][118]
Prot L	1hz6	152	110,7	1	9,96	1129	[117][118]
Ankirina	1n11	37	26,2	-	10,66	1179	[119]

Fibronectin III	loww	220	51,0	1	15,14	396	[120][112]
GB1	1pga	190	86,0	1	9,70	494	[121][116]
Synaptotagmin	1rsy	60	31,0	2	18,99	1633	[122]
Tenascin	1ten	135	49,5	1	15,44	478	[123][112][124]
Titin	1tit	204	68,2	1	15,83	783	[125]
Fibronectin III	1ttf	75	40,5	2	12,23	834	[126][127]
Fibronectin III	1ttg	75	44,0	2	11,48	464	[126][127]
Ubiquitin	1ubq	230	76,7	1	11,48	504	[128][129]
Prot L	2ptl	152	33,0	1	10,00	767	[130][118]

Los modelos basados en estructura utilizan una función de energía potencial que incluye de manera explícita la topología del estado nativo [46], de manera que las energías y tiempos de simulación para acceder a fenómenos biológicos son alterados. Por ello, se utilizó una calibración de tiempo reportada que estima una relación entre tiempo simulado:real de 1:50,000 [45]. Por otro lado, la temperatura fue calibrada comparando fluctuaciones de cadena principal observadas en CMD a 298 °K y a distintas temperaturas reducidas en SBM para la misma proteína. A 70 grados reducidos, o $0,6 [k_B T \cdot \epsilon^{-1}]$ se observó la menor desviación entre ambos. El efecto de esta elección de temperatura es evaluado más adelante. Utilizando esta información se realizaron SMD de estas proteínas a la velocidad de estiramiento correspondiente a cada reporte experimental. De las trazas de desplegamiento se obtuvo la fuerza máxima y número de *rips* para cada proteína tal como se explica en la Figura 9.

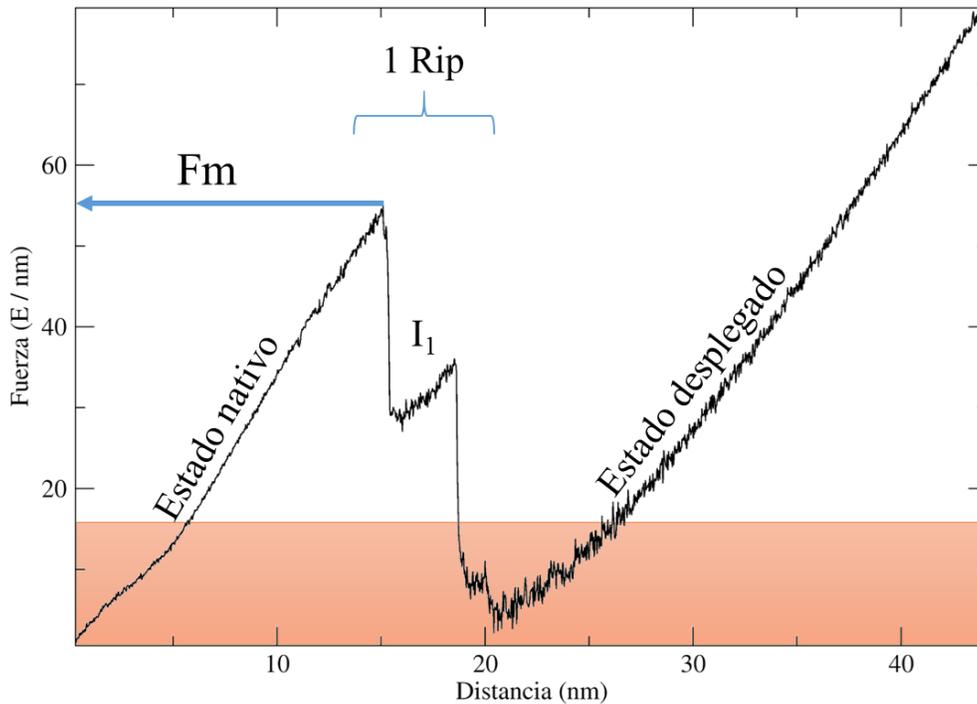


Figura 9: Traza representativa de los datos obtenidos al realizar SMD mediante SBM atómicos. En el eje X se muestra la distancia total, análoga a la observada en AFM, en el eje Y el cambio en la función de energía potencial dada por tracción mecánica. En naranja se marca lo estimado al límite de resolución obtenido por equipos de AFM convencional, la flecha azul representa la fuerza máxima obtenida antes de la transición al estado intermedio y rápido desplegamiento. Se observan dos eventos de ruptura consecutivos, que en equipos convencionales de AFM no sería posible observar, debido a limitaciones en resolución temporal, por esto se considera como un solo “rip” o evento de ruptura.

Los datos reportados experimentalmente para estas proteínas fueron comparados con los resultados de las simulaciones utilizando la fuerza media de ruptura, F_m , junto a la profundidad del pozo energético de contacto (ϵ) y el orden de contacto absoluto (AbsCO) [70] (Figura 10), obteniendo la mejor correlación al utilizar la fuerza promedio máxima de desplegamiento ponderada por su orden de contacto como parámetro de entrada.

El error estándar de la correlación entre fuerzas experimentales y teóricas tiene un valor de 56 pN, el cual si bien es significativo, es aproximadamente la mitad al reportado previamente utilizando un modelo sin detalle atómico [54]. Para evaluar en qué medida las curvas de desplegamiento simuladas representan las observadas experimentalmente, se cuantificó el número de eventos de ruptura o *Rips* visibles en las

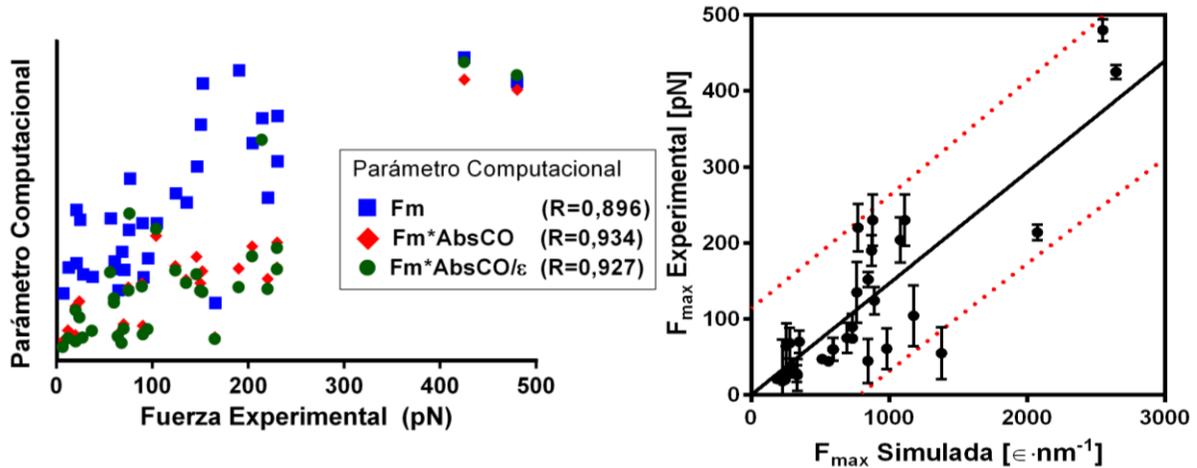


Figura 10: (izquierda) Correlación entre las mediciones de fuerza experimental y computadas para cada proteína ajustado frente a tres parámetros computacionales. Cada set de puntos fue ajustado linealmente forzando al origen y se muestra el coeficiente de correlación lineal de Pearson (R) junto a cada uno. (derecha) Utilizando el set de mayor correlación, correspondiente al que pondera cada fuerza media por el orden de contacto propio de cada proteína se realizó un ajuste más detallado, que incluye la desviación estándar de medición experimental. La línea continua representa el mejor ajuste lineal, y la línea punteada el intervalo de 95% predicción. El ajuste tiene una pendiente de $0,147 \pm 0,009 \text{ pN} \cdot \text{nm}^{-1} \cdot \text{ε}^{-1}$, y un R^2 de 0,88, junto con un error estándar de 56 pN.

curvas experimentales y en las simulaciones, considerando también las limitaciones de resolución temporal y de fuerza comúnmente encontradas en equipos de AFM.

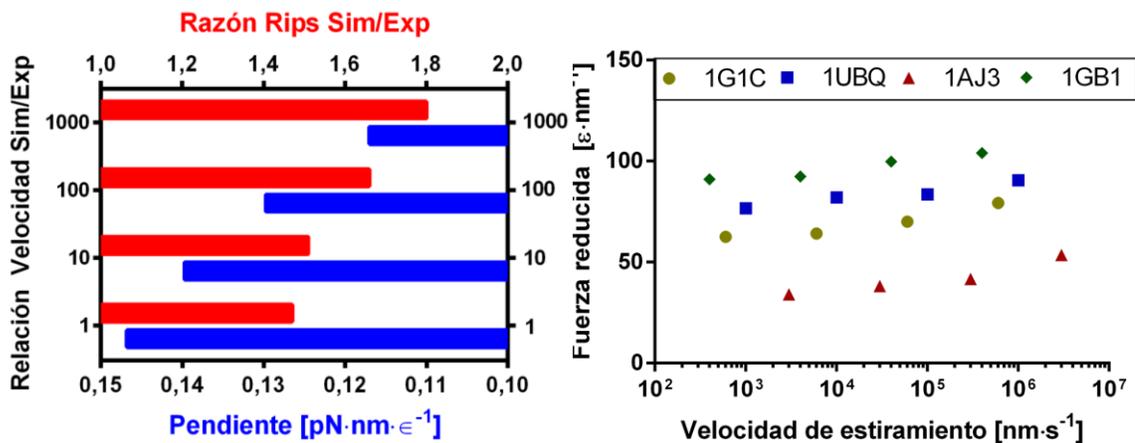


Figura 11: (Izquierda) Efecto del aumento de la velocidad de estiramiento sobre la fuerza y trazas de desplegamiento. Para evaluar la presencia de un evento de ruptura se consideró un umbral de 20 pN, cercano al límite de resolución de fuerza de equipos de AFM estándar, junto con las limitaciones temporales asociadas a la presencia de dos eventos consecutivos de desplegamiento. (Derecha) Ejemplo de cambio de fuerza de ruptura observada computacionalmente respecto al aumento de su velocidad de estiramiento. Los ejemplos mostrados corresponden a titina (1G1C), ubiquitina (1UBQ), espectrina (1AJ3) y proteína G (1GB1).

A velocidades de estiramiento comparables con las experimentales [45], se generan trazas de desplegamiento que presentan 42% más *rips* que su contraparte

experimental (Figura 11). Esto corresponde a una mejora comparado con las trazas reportadas en literatura utilizando un enfoque similar, donde aquellas trazas muestran 298% de incremento respecto al número de *rips* experimentales [54], [52]. Al aumentar exponencialmente la velocidad de estiramiento, se observa un aumento lineal en las fuerzas de ruptura, consistente con la descripción fenomenológica propuesta por Bell-Zhurkov [40], [41]. Sin embargo, a velocidades de estiramiento muy altas ($\geq 1 \text{ mm}\cdot\text{s}^{-1}$) se comienza a observar un aumento no lineal de las fuerzas de ruptura evidenciado para el estiramiento de la proteína espectrina (Figura 11). Este comportamiento es consistente con la descripción propuesta por Hummer & Szabo para extender el modelo de Bell-Zhurkov a estiramientos a alta velocidad [131].

Debido a que a altas velocidades de estiramiento (1000 veces la velocidad experimental) se observó pérdida de cooperatividad durante el desplegamiento mecánico de algunas proteínas (Figura 11), se evaluó el efecto de la temperatura sobre la correlación entre las fuerzas de desplegamiento.

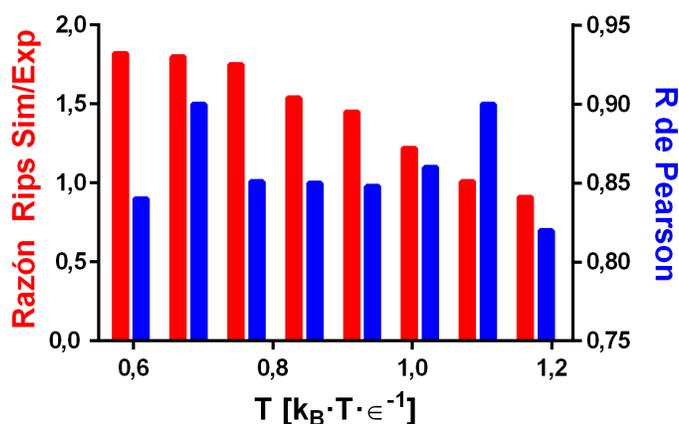


Figura 12: Efecto de la temperatura sobre la correlación (R^2) y trazas de desplegamiento. Las temperaturas exploradas corresponden al rango de 0,6 a 1,2. Este experimento se realizó a la mayor velocidad de desplegamiento para minimizar eventos de desplegamiento térmico observados comúnmente a la máxima temperatura utilizada.

Estos resultados muestra que la temperatura no genera una tendencia obvia sobre la correlación (Figura 12, azul), probablemente porque muchos de los cambios que ocurren al aumentar la temperatura, como el desplegamiento térmico, disminución de fuerza máxima, y vector de estiramiento dependen de cada proteína y su estabilidad térmica [49]. Lo que sí se observa para todas, es una disminución del número de eventos de ruptura

observadas en las trazas de desplegamiento computacional (evaluado por la razón rips Sim/Exp). Esto ocurre principalmente porque i) a mayor temperatura se observa mayores fluctuaciones de la fuerza en el tiempo y por lo tanto el ruido, impidiendo detectar con certeza el rip; y ii) a temperaturas altas se observa desplegamiento térmico o fuerza-independiente. Esto se observa gracias a que se realizaron 10 simulaciones independientes por cada proteína a cada temperatura, permitiendo observar trazas que se desvían significativamente en la fuerza de ruptura del promedio, generalmente con fuerzas significativamente menores.

La temperatura inicial de $0,6 [k_B T \cdot \epsilon^{-1}]$ corresponde a la temperatura donde se realizó la calibración descrita a distintas temperaturas; sin embargo, temperaturas más bajas (cerca de $0,1 [k_B T \cdot \epsilon^{-1}]$) se han propuesto en literatura como ideales para el estudio de desplegamiento, debido a que representan las características cinéticas esperadas para proteínas a temperatura ambiente. No obstante, esto último depende enormemente de la función de energía potencial utilizada, criterio de contacto nativo y de la temperatura de fusión de cada proteína [49].

Objetivo 2: Caracterización del plegamiento de RfaH mediante SBM

Desplegamiento mecánico de RfaH utilizando SBM

Para evaluar las características del desplegamiento mecánico de RfaH se utilizó modelos basados en estructura con las mismas características que se evaluaron las proteínas interiores. En el caso de RfaH, las simulaciones se ejecutaron sobre la proteína completa con su dominio carboxilo terminal estructurado tanto en el estado α -helicoidal como β -plegado (Figura 13 y 14):

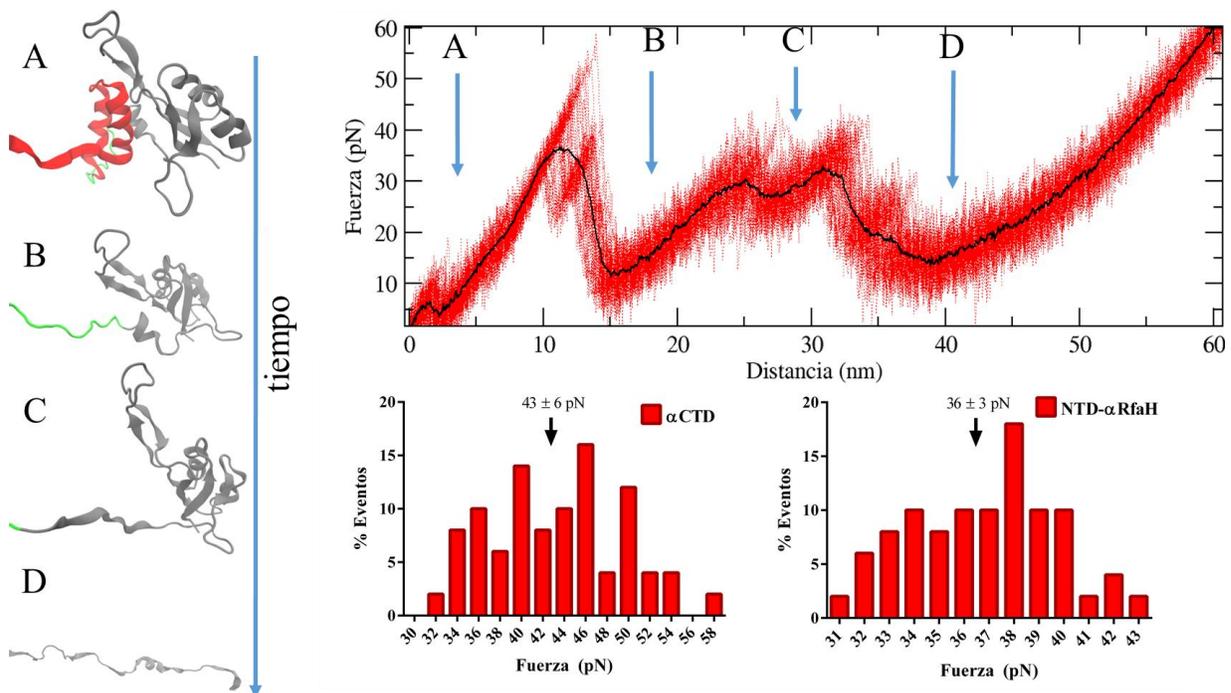


Figura 13: Desplegamiento mecánico de RfaH desde el estado inicial α -plegado utilizando modelos basados en estructura. A la izquierda se observan imágenes representativas de los estados estables de plegamientos observados desde el estado nativo (A) hasta el estado desplegado (D). A la derecha arriba se observa la traza de desplegamiento de 50 eventos independientes de desplegamiento en rojo y en negro la traza media de desplegamiento, señalizando cada estado estructurado (A→D) observado en la derecha. Abajo se muestra la distribución de la fuerza máxima de desplegamiento involucrada en el desplegamiento del dominio carboxilo-terminal (izquierda, transición A→B) y del desplegamiento del amino-terminal (derecha, transición C→D). La fuerza ha sido convertida a unidades reales utilizando la correlación para las otras proteínas previamente mencionadas.

En el desplegamiento desde el estado α -plegado se observa que el desplegamiento y la disociación del dominio carboxilo terminal ocurren simultáneamente, contrario a lo observado mediante SMD convencional [22]. En estas simulaciones se observan altas

fluctuaciones de las hélices del CTD previo al desplegamiento, concordante con lo reportado en literatura [22], [25], [27], [33]; particularmente en la hélice α_2 , que muestra algún grado de desenrollamiento de su estructura helicoidal previo al desplegamiento global del α CTD.

Por otro lado, el dominio N-terminal se despliega en dos eventos de ruptura: un primer desenrollamiento de la hélice pequeña que conecta ambos dominios y un desplegamiento final del NTD en un gran paso de desplegamiento. Al utilizar la correlación de fuerzas simuladas-teóricas para estimar la fuerza a la que ocurren ambos eventos limitantes (Figura 10, B), se observa que el desplegamiento del CTD ocurre con una fuerza cercana a 43 pN a la velocidad estimada de estiramiento ($\approx 100.000 \text{ nm s}^{-1}$), siendo tal el evento de ruptura mecanoresistente, seguido por el NTD que se despliega a una fuerza cercana a 36 pN.

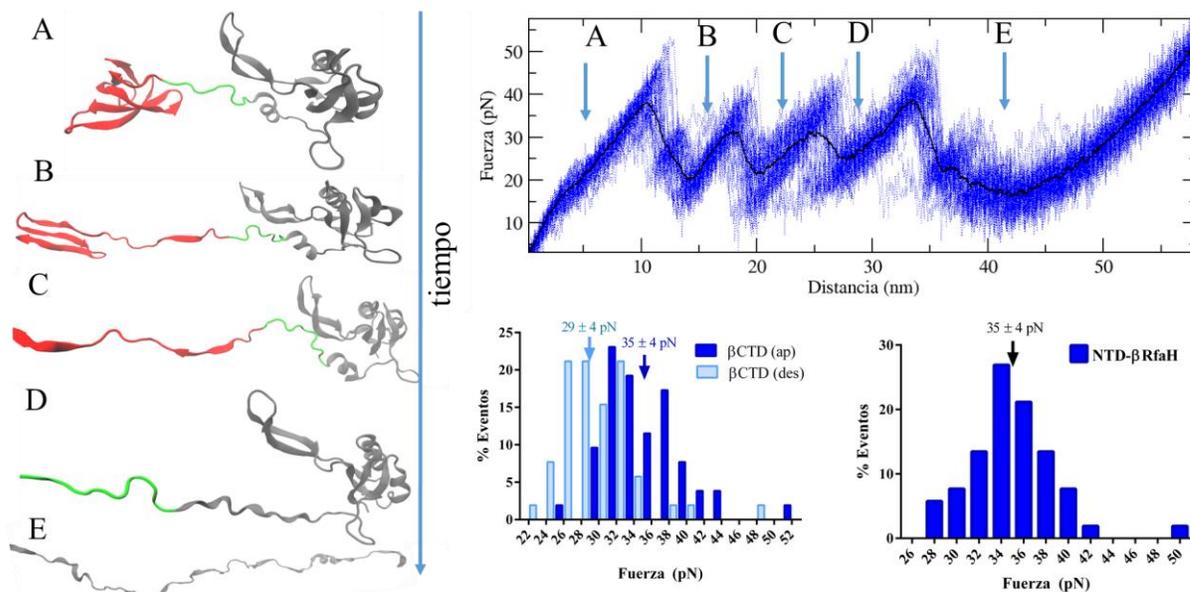


Figura 14: Desplegamiento mecánico de RfaH desde un estado inicial β -plegado utilizando modelos basados en estructura. Análogo a la Figura 13, se utilizó un estado optimizado en el estado donde el CTD está plegado como barril- β obtenido mediante códigos de modelamiento de loops de Rosetta y SBM. A la izquierda se observa los estados representativos observados durante el desplegamiento (A→E) donde C, D y E son pasos comunes con el desplegamiento desde la forma α RfaH. A la derecha arriba se observa la traza de 50 eventos independientes de desplegamiento y en negro la traza promedio. Abajo se muestra a la izquierda la distribución de fuerzas ponderadas a unidades reales requeridas para el desplegamiento del CTD en los dos pasos observados: (i) apertura inicial del barril beta (A) para dar lugar al intermediario (B) en azul (β Ctd (ap)), y (ii) en celeste el desplegamiento de este intermediario para dar lugar sólo al NTD plegado (β CTD (des)). A la derecha se observa la distribución de fuerzas para el desplegamiento del NTD (transición D→E).

En el caso del desplegamiento de RfaH desde el estado β -plegado (Figura 14) se observa que, al igual que para α RfaH, el CTD es quien experimenta el primer evento de

desplegamiento observado. En este caso, el barril- β se abre en un primer paso con una fuerza ponderada a unidades reales de 35 pN para dar lugar a un intermediario, el que luego se despliega completamente en un segundo evento con una fuerza estimada de 29 pN.

Estos resultados guardan estrecha relación con datos obtenidos en nuestro laboratorio por José Molina, señalando que en experimentos de desplegamiento químico mediante la adición de cloruro de guanidinio al CTD de RfaH aislado en solución (Figura 15), identifica un intermediario con estructura secundaria a ~ 1 M de agente caotrópico. La combinación de resultados experimentales y simulados muestra que el desplegamiento del CTD parece ocurrir a través de una apertura inicial del barril- β plegado en donde se expone al menos parte del núcleo hidrofóbico. Esto es observado experimentalmente por la única transición de cambio de fluorescencia intrínseca de una mutante de fenilalanina por triptófano (F126W) en el centro hidrofóbico (figura 15). Esta primera transición se sobrelapa con el primer cambio de estructura secundaria observada a través de CD, que además muestra una segunda transición no observada por fluorescencia intrínseca que da lugar al estado desplegado. A pesar de que no es evidencia directa, éstos fenómenos guardan estrecha similitud con lo observado computacionalmente para el desplegamiento del β CTD, según lo cual el desplegamiento mecánico da lugar primero a un intermediario (Figura 14, B) donde las hebras más grandes del barril- β están expuestas al solvente, pero

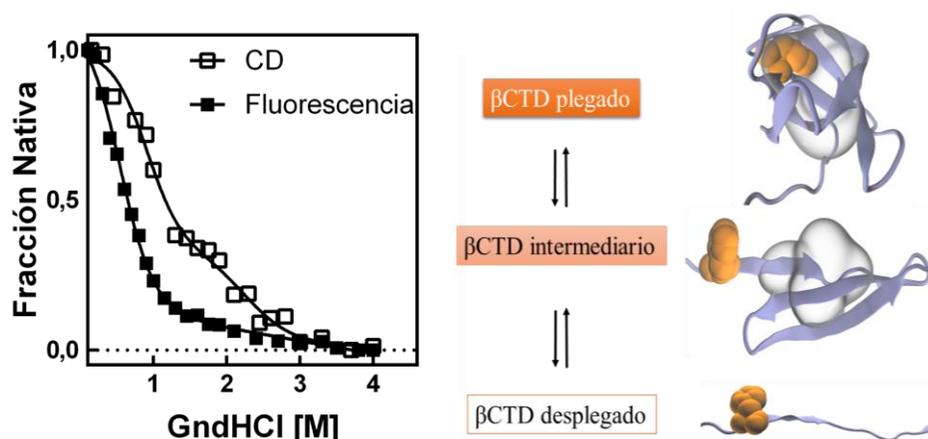


Figura 15: Desplegamiento químico del CTD de RfaH. El desplegamiento se realizó mediante la adición de concentraciones crecientes de cloruro de guanidinio a $5 \mu\text{M}$ del CTD de RfaH y fue seguido por fluorescencia intrínseca mediante una mutante F126W (cuadrados negros) y por dicroísmo circular. A la derecha se observa el intermediario del mismo observado a través de SMD con SBM, marcando en naranja el residuo F126 y en blanco el núcleo hidrofóbico. Obtenidos por José Molina, Instituto de Ingeniería Biológica y Médica, Pontificia Universidad Católica de Chile.

manteniendo su estructura secundaria. Esta intermediario β -extendido se despliega en un evento independiente sucesivo.

Por otro lado, el desplegamiento del NTD conserva estrecha relación con lo observado previamente para el caso de α RfaH, manteniendo las características del desplegamiento inicial de la hélice pequeña, seguido por una ruptura del NTD involucrando una fuerza ponderada a unidades reales equivalente entre ambas simulaciones.

Trabajo al equilibrio de desplegamiento de α - y β CTD

Para poder estimar la energía al equilibrio involucrada en el desplegamiento del CTD en ambos estados nativos, se optimizó un protocolo de replegamiento utilizando SMD. A diferencia del protocolo recién mencionado de desplegamiento mecánico, éste se realiza a mayor temperatura para evitar la formación artificial de estructura secundaria en el α CTD previo a su asociación al NTD, y a velocidades de estiramiento muy lentas ($\approx 1.000 \text{ nm s}^{-1}$) tanto para el desplegamiento como replegamiento. Debido a que estas simulaciones tienen un alto costo computacional, para estudiar el α CTD se simuló sólo la disociación del CTD, por lo que durante todas las simulaciones el NTD siempre estuvo estructurado. Por otro lado, debido a que β CTD existe aislado en solución, se simuló este dominio en ausencia del NTD (Figura 16).

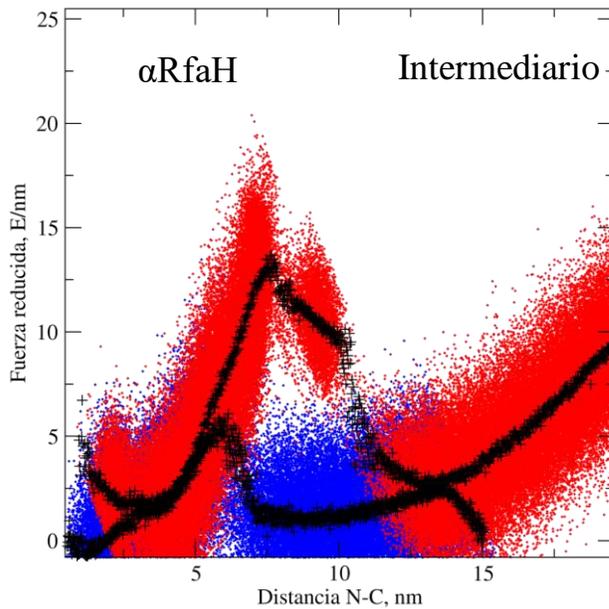
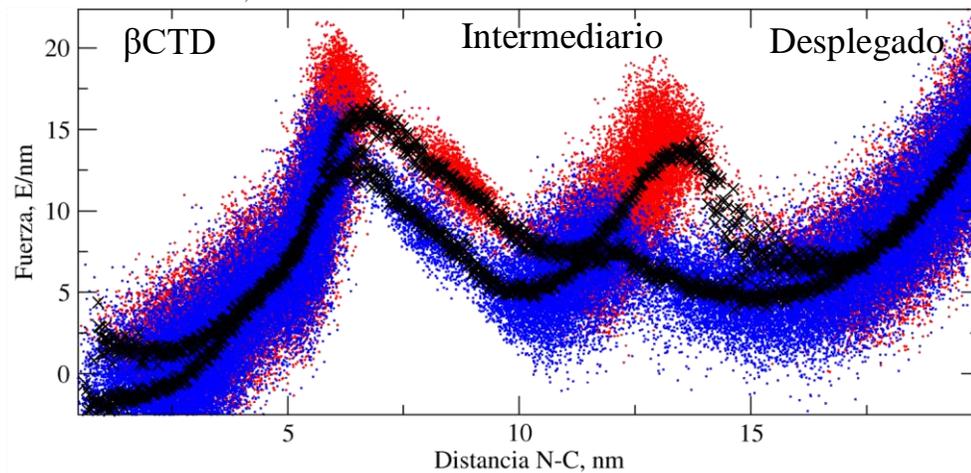


Figura 16: Desplegamiento reversible de α - y β CTD. Utilizando el diseño experimental descrito en el texto, se muestran en rojos los desplegamientos, en azul el repliegamiento y en negro la curva promedio para el desplegamiento reversible del α CTD (izquierda) y β CTD (abajo) tras 30 eventos independientes de desplegamiento y repliegamiento. Para la proteína completa en estado α RfaH (izq.) sólo se evaluó la transición desde el estado nativo al intermediario NTD-plegado, mientras que para el CTD aislado como barril- β (der.) se observó el desplegamiento completo del mismo mediado por un intermediario estable observado en ambos, desplegamiento y repliegamiento.

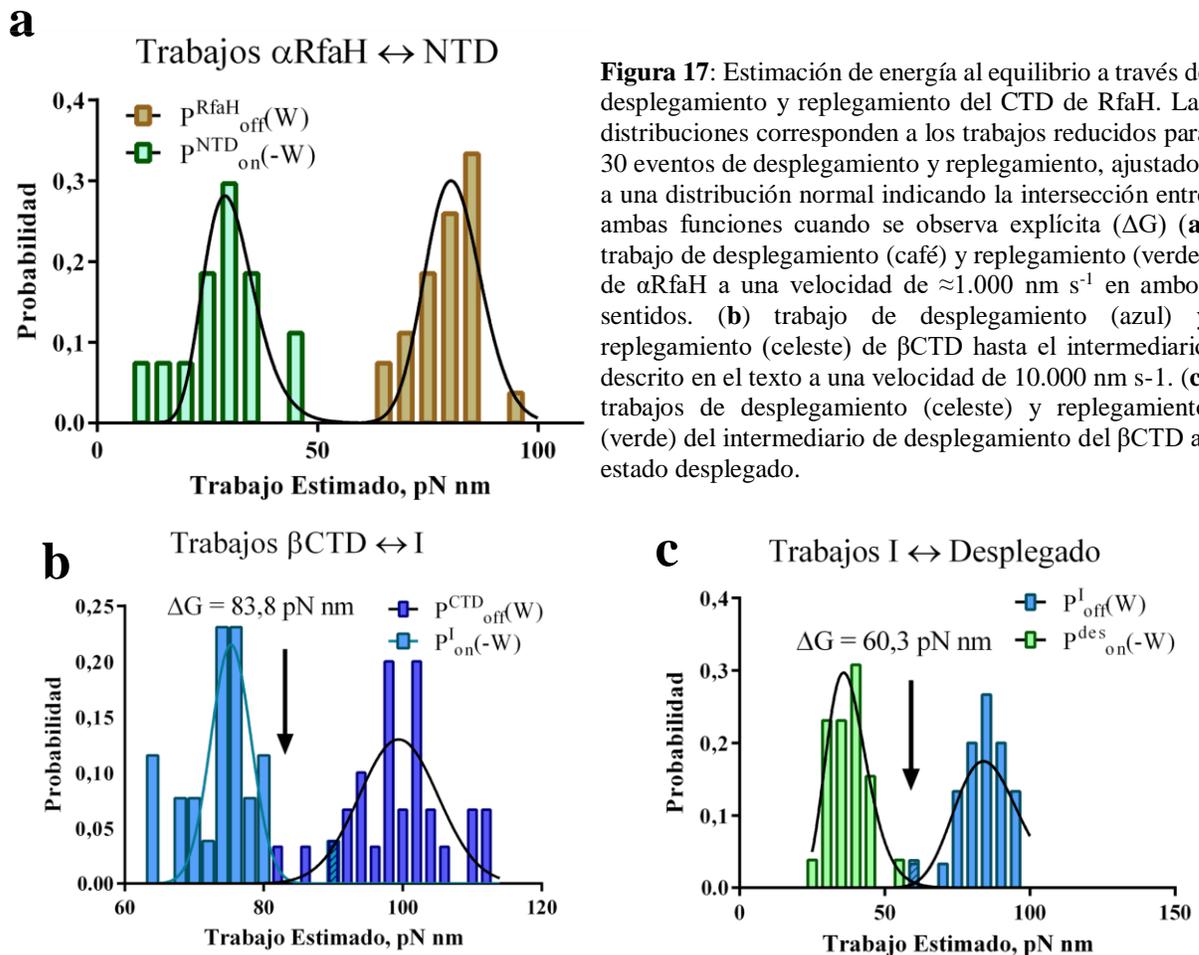


En estos experimentos se observa alta histéresis para el repliegamiento del CTD en el contexto de α RfaH (Figura 16, izquierda), esto es seguramente a que la mayoría de los contactos de este interdominio son lejanos a nivel de secuencia y la función de energía utilizada para los modelos basados en estructura corresponde a una función tipo Lennard-Jones cuyo radio máximo es $\approx 8 \text{ \AA}$ (sobre esa distancia la función adquiere valores numéricos despreciables), produciendo una baja probabilidad de formación de contacto desde un estado estirado.

Por otro lado, el desplegamiento reversible del dominio carboxilo terminal como barril- β ocurre en dos transiciones al igual que en el contexto de la proteína completa. A pesar de que ambas presentan histéresis entre desplegamiento y repliegamiento, la mayor

se observa para la formación del intermediario desde el estado desplegado, mientras que la formación del estado nativo desde aquél intermediario es mucho más rápida. Utilizando los cambios de distancia entre los carbonos- α del primer y último residuo para cada caso (puntos de anclaje y estiramiento respectivamente) se calculó los trabajos asociados a las transiciones y se resumen en la Figura 17, junto con la energía al equilibrio estimada a partir del teorema de fluctuaciones de Crooks [38]:

Al analizar las distribuciones de trabajos de desplegamiento y replegamiento, observamos que efectivamente la histéresis observada durante el estiramiento de la proteína RfaH completa se conserva, impidiendo el uso del teorema de Crooks para poder calcular la energía al equilibrio del desplegamiento del α CTD. Por otro lado, para ambos desplegamientos del β CTD aislado se logra observar superposición entre la distribución de los trabajos, dando valores de energía al equilibrio para la transición $N \rightleftharpoons I$ y la transición $I \rightleftharpoons D$ de 12,0 y 8,7 kcal·mol⁻¹ respectivamente. A pesar de que no se obtiene el valor de



energía al equilibrio, el desplegamiento del α CTD involucra trabajo similar al de cada evento del β CTD, por lo que acorde a lo estimado por este enfoque, el estado β -plegado de RfaH es más estable termodinámicamente que su contraparte α -helicoidal, contradiciendo la evidencia experimental.

Objetivo 3: Análisis termodinámico de RfaH a través de CCR

Desconvolución de la energía de RfaH

Para explorar en más detalle la transformación de RfaH, se utilizó el protocolo de confinamiento-conversión [55], [56] para estimar la diferencia de energía libre entre los estados α - y β RfaH en un entorno de dinámica molecular convencional. Para llevar a cabo este procedimiento se siguió las fluctuaciones espaciales a distintos grados de restricción de posiciones globales (Figura 18). En éstas, se observa una dependencia mayormente lineal, tal como lo reportado en literatura para otros sistemas de simulación [56]. Por esto, se calculó el área bajo la curva utilizando integración trapezoidal tal como está descrito en la ecuación 4 para obtener la energía involucrada en el confinamiento. De acuerdo a estos cálculos, se observa que basta con constantes de restricción cercanas a 10^1 kcal·mol⁻¹·Å⁻² para lograr el estado confinado, en el que las fluctuaciones se independizan de la fuerza de restricción (datos no mostrados).

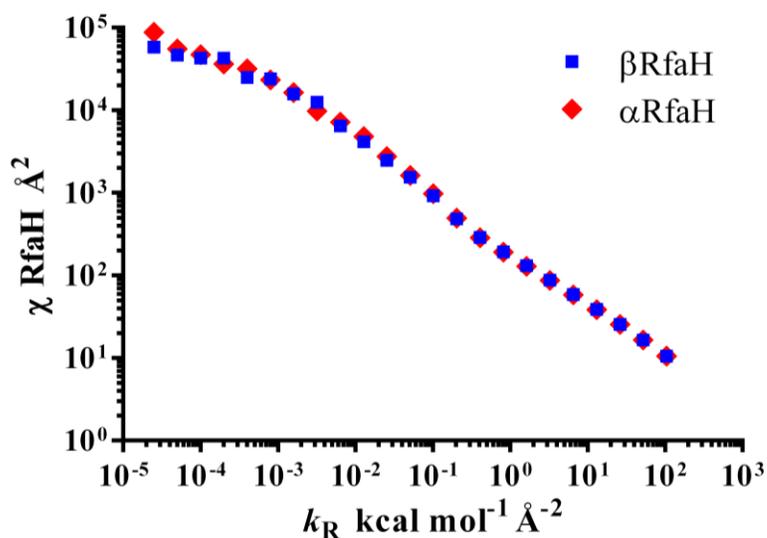


Figura 18: Dependencia de las fluctuaciones globales respecto a la constante de restricción por átomo. Las variaciones en RMSD oscilan entre ≈ 5 Å en la menor constante de restricción (equivalente al sistema libre) hasta $0,002$ Å en el microestado confinado. En esta curva el mayor aporte a la diferencia de energía de confinamiento (mayor diferencia en la integral) está en las fluctuaciones a valores de constante entre $0,001$ a 1 kcal·mol⁻¹·Å⁻², tal como se observa en la figura 13.

Utilizando el enfoque de confinamiento-conversión descrito en la sección de Teoría y Métodos, se estimó la diferencia de energía que requeriría la transformación de RfaH entre sus dos estados (Figura 19).

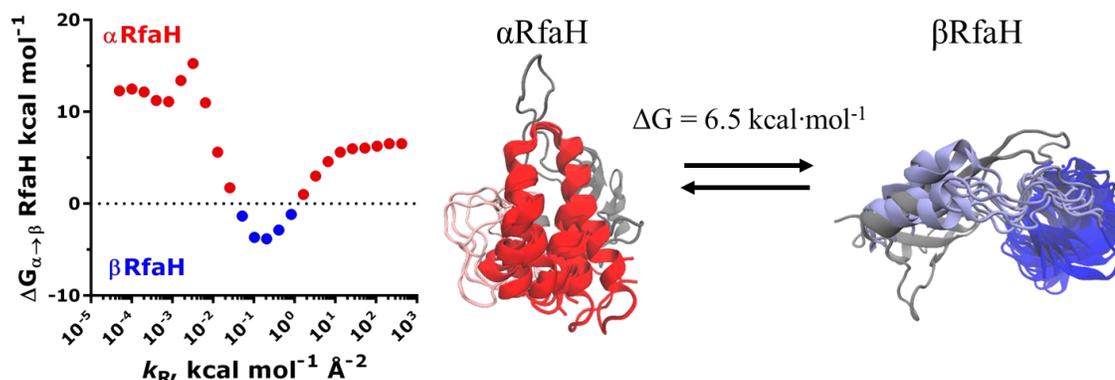


Figura 19: Ciclo de confinamiento y energía libre del microestado mínimo para α - y β -RfaH. El confinamiento se logró con 24 simulaciones de restricción creciente desde 0,000025 a 204 kcal mol⁻¹·Å⁻² por átomo. La energía libre del estado confinado se obtuvo por modo normal utilizando las 3N-6 frecuencias no negativas. A la izquierda se observa la diferencia de energía entre ambos estados a distintos grados de confinamiento, y convergencia en constantes superiores a 10¹ kcal·mol⁻¹·Å⁻².

Las mayores de las fluctuaciones de α RfaH se encuentran en la base de las hélices, incluyendo el *loop* que conecta ambos dominios, y la horquilla que yace del NTD y se posa sobre el CTD (datos no mostrados). Por otro lado, el estado mayormente minimizado del β CTD corresponde a un estado asociado al NTD mediante varias interacciones de puentes de hidrógeno y dos puentes salinos, apoyando simulaciones previamente reportadas en literatura que muestran que el estado- β y las interacciones interdominio no son mutuamente excluyentes [27]. Sin embargo, en simulaciones mayormente libres, se observan grandes fluctuaciones del CTD frente al NTD, presentando una desviación de posiciones promedio mayor que el α CTD asociado. Al computar la diferencia de energía global entre ambos estados se observa una convergencia estable a 6,5 kcal·mol⁻¹ aplicando constantes armónicas mayores a 10 kcal·mol⁻¹·Å⁻².

Utilizando el enfoque de confinamiento-conversión-liberación (CCR) [62] se descompuso aproximadamente el aporte energético a nivel de residuo hacia cada estado, y se calculó la diferencias de energía relativa o $\Delta\Delta G$ (Figura 20). En estos, se observa que los mayores determinantes de la estabilidad del α CTD se encuentran en la cabeza de las hélices, similar a la estructura mecanoresistente observada durante los estiramientos con SBM. Por otro lado, los residuos involucrados principalmente en la estabilidad del β CTD se encuentran en los residuos formando la base de las hélices en α RfaH. Ya que cada residuo está más favorecido energéticamente en un estado α - o β -plegado, se han descrito en literatura como “conductores moleculares del plegamiento” [62]. En este caso, se

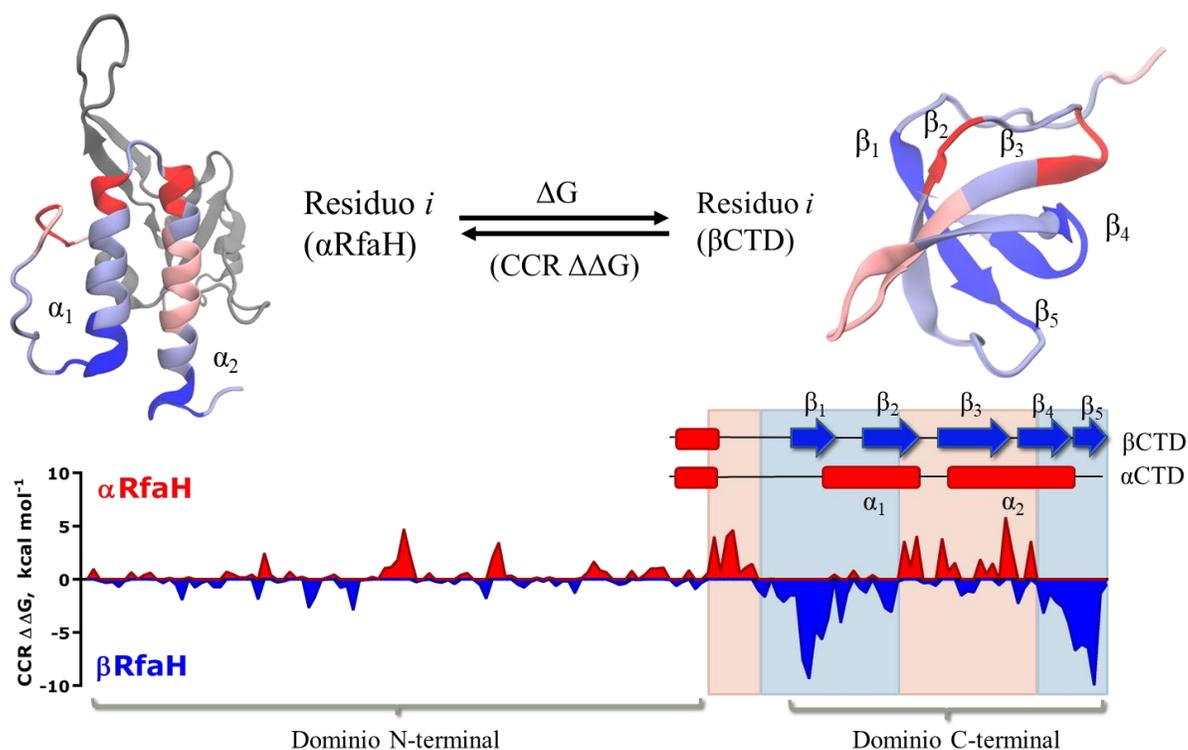


Figura 20: Descomposición de la energía de confinamiento y conversión a nivel de residuo. Al descomponer los aportes de confinamiento, y considerar la diferencia de energía entre los estados confinados a través de un cambio en energía interna a nivel de residuo, se puede aproximar la distribución de la energía libre de CCR. Arriba se muestra la energía libre por residuo en la estructura primaria, y abajo, en la estructura terciaria en ambos estados. Rojo representa una preferencia por el estado alfa, y azul por el estado beta.

aprecia claramente como regiones dentro de RfaH prefieren un estado estructural. En particular, la cabeza de la horquilla- α rica en contactos interdominio [22], [27] estabiliza preferencialmente el estado α RfaH, mientras que el resto del CTD estabiliza profundamente el estado β RfaH (Figura 20).

Debido a que nuestro laboratorio posee datos de experimentos de espectrometría de masas de intercambio hidrógeno/deuterio (HDX) obtenidos antes del inicio de mi trabajo de tesis (por César Ramírez-Sarmiento, en colaboración con la Dra. Elizabeth Komives de la Universidad de California en San Diego, La Jolla, California, EEUU), se utilizó estos datos para proveer comparación experimental de lo observado a través de dinámica molecular y CCR. Para ello se calculó las constantes cinéticas de incorporación de deuterio a la amida del enlace peptídico con el fin de obtener las constantes de equilibrio de apertura [132] y la energía libre asociada. Se calculó la diferencia de energía libre asociada a la apertura, que en el contexto de estructura secundaria es proporcional a la

estabilidad de la misma, entre el CTD unido a RfaH (α CTD) y el mismo aislado (β CTD) y se comparó con la descomposición observada por CCR (Figura 21).

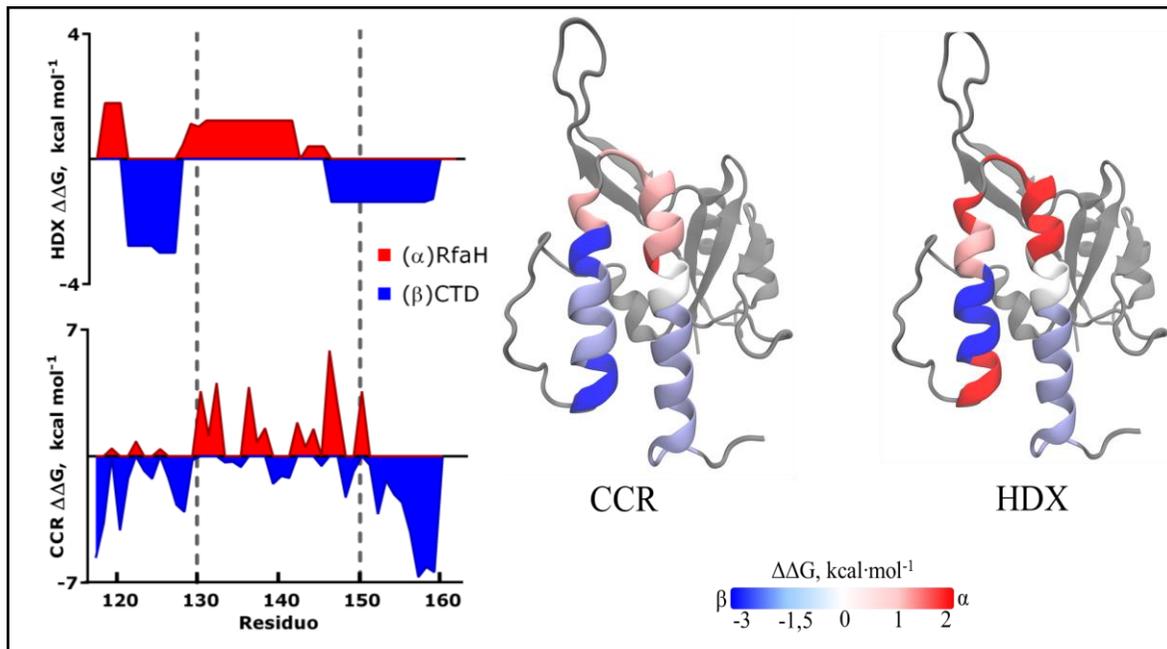


Figura 21: Comparación entre la diferencia de energía entre el estado α - y β CTD obtenidas a través de CCR y HDX. (izquierda) arriba se muestran las diferencias de energía a nivel de resolución de péptido obtenido a través de espectrometría de masas de incorporación de deuterio para el CTD de RfaH completa, en el estado α -helicoidal, y el CTD aislado, en el estado β -plegado. Se calculó las constantes de incorporación experimentales para los péptidos en ambos estados y se restó la diferencia de energía de apertura entre ambos ($\Delta G_{ap\alpha} - \Delta G_{ap\beta}$). Abajo, las diferencias obtenidas a través de la descomposición de energía libre de CCR como se indicó anteriormente. (derecha) Promedio de la diferencia de energía a nivel de residuo representada en la estructura de α RfaH utilizando el nivel de resolución peptídico a través de HDX. La escala de energía es la misma para ambos métodos.

Estos resultados muestran que hay una gran similitud entre las diferencias en estabilidad local calculadas computacionalmente y los resultados experimentales de espectrometría de masas de incorporación de deuterio. En breve, los resultados combinados corroboran que, efectivamente, la base de las hélices del CTD prefieren energéticamente el estado β -plegado, mientras que la cabeza de la horquilla helicoidal es más preferente energéticamente por el estado α -helicoidal. Es posible estimar la constante de equilibrio de desplegamiento mediante intercambio hidrógeno-deuterio, considerando que el intercambio de las amidas más lentas corresponde a transiciones globales de la proteína, y que efectos de apertura local o sub-globales son despreciables o no existentes [132]. Se observa que para el CTD aislado, el péptido correspondiente a los tres residuos principales de la hoja $\beta 2$ (127-129) formando puentes de hidrógeno con $\beta 1$ y $\beta 3$, tiene una

diferencia energética de apertura de $3,9 \text{ kcal}\cdot\text{mol}^{-1}$. Este mismo péptido posee la menor constante de intercambio en el CTD asociado al NTD, con una diferencia energética de apertura de $4,9 \text{ kcal}\cdot\text{mol}^{-1}$, y el péptido que le sigue tiene una diferencia de energía de apertura similar, con $4,7 \text{ kcal}\cdot\text{mol}^{-1}$, estando ambos en la cabeza de la horquilla α -helicoidal (residuos 127-129, 130-141). Sin embargo, debido a que éste péptido en ambos casos está expuesto al solvente, es muy probable que su constante de apertura refleje más bien la estabilidad local de la estructura secundaria que el desplegamiento global del CTD.

Resumiendo, utilizando dinámicas de confinamiento se pudo calcular la diferencia de energía libre entre los estados α - y β -plegado de RfaH, entregando un valor de $\Delta G = 6,5 \text{ kcal}\cdot\text{mol}^{-1}$ para la reacción $\alpha\text{RfaH} \rightleftharpoons \beta\text{RfaH}$. Este resultado es apoyado por la posterior descomposición de aquella energía libre a nivel de residuo, que entrega valores similares en magnitud y distribución a los calculados experimentalmente a través de experimentos de espectrometría de masas del intercambio hidrógeno-deuterio.

V. Discusión

Para predecir el comportamiento mecánico de RfaH se utilizaron modelos basados en estructura [44], que a través de un campo de fuerza y Hamiltoniano modificado utilizan las interacciones nativas observadas en la estructura cristalina o en solución mediante ^2H - ^{13}C -NMR para llevar a cabo simulaciones de plegamiento y desplegamiento en tiempos computacionalmente accesibles. Debido a estas modificaciones, variables como masa, temperatura, energía y tiempo, junto a todas sus derivadas, no son directamente contrastables con las magnitudes convencionales, y por ello reciben el nombre de magnitudes reducidas. Para poder correlacionar unidades reducidas con su contraparte convencional, un trabajo estudió la cinética de desplegamiento de un cierre de leucina LZ26 mediante modelos basados en estructura y su contraparte experimental [45], estableciendo que en modelos atomísticos estándar, una unidad de paso de tiempo o τ equivale a aproximadamente 50.000 veces la misma cantidad en unidades convencionales. Para poder entonces estudiar el desplegamiento mecánico de RfaH utilizando velocidades de estiramiento contrastables con las experimentales, y que a su vez entreguen fuerzas contrastables a las magnitudes físicas, se estudió la relación entre dinámicas de estiramiento utilizando modelos basados en estructura, y su equivalente en diseño experimental de pinzas ópticas o AFM.

Relación de estiramientos *in silico/in vitro*.

En este estudio fue posible simular 35 diferentes modelos experimentales, que van desde sistemas pequeños, como fibronectinas individuales ≈ 9 kDa, hasta de gran tamaño, como son los repetidos de ankirina o la β -catenina ≈ 70 kDa, todos con detalles atomísticos y en múltiples repeticiones (n total = 160 simulaciones / estructura). A pesar de la alta heterogeneidad del conjunto de proteínas estiradas, con diversidad de contenido de estructura secundaria tipo alfa y/o beta, ya sea globulares o fibrilares, el comportamiento global de estas proteínas en nuestras simulaciones comparte características de plegamiento contrastables con la evidencia experimental. Si bien a nivel de las trazas de desplegamiento, en promedio estas simulaciones sobreestiman en un 43% el número de

eventos de ruptura, se logra establecer una correlación lineal entre las fuerzas de estos eventos mecánicos experimentales y computacionales con el menor error estándar reportado en literatura [53], [54]. Observar más *rips* computacionalmente que en el experimento puede deberse a que i) estos intermediarios se encuentran bajo el límite de resolución temporal y/o espacial de instrumentos utilizados para experimentos de manipulación de molécula individual, como pinzas ópticas y AFM; o bien ii) las condiciones computacionales de estiramiento no simulan con suficiente exactitud las experimentales.

Una evidencia a favor de lo último son las diferencias observadas al modificar la temperatura de estiramiento. En literatura se indica que a bajas temperaturas (0,1 a 0,2 T_m , donde T_m es aproximadamente 140 grados reducidos) los SBM representan mejor las propiedades cinéticas asociadas a la cooperatividad de plegamiento observada *in vitro* [49]. Mientras que nuestros resultados indican que, al comparar las fluctuaciones atómicas de cadena principal, es una temperatura más elevada equivalente a 0,5 T_m la que mejor se ajusta al comportamiento de dinámicas moleculares convencionales. Por otro lado, al mirar las trazas de desplegamiento, son temperaturas incluso mayores, cercanas a 0,8 T_m , las que mejor reproducen las trayectorias experimentales. Sin embargo, la variación observada en el comportamiento respecto a la temperatura es proteína-dependiente, pues a pesar de que existe una temperatura que minimiza el error promedio de los parámetros evaluados, como la fluctuación de cadena principal, la mayoría de las proteínas posee una temperatura de simulación ideal, o que entrega la mejor correlación con dinámicas convencionales, que se diferencia en $\pm 0,1 T_m$ de la utilizada en este estudio.

Al comparar las fuerzas reducidas con las experimentales observamos una clara tendencia lineal ($R=0,896$) entre ambas, que es particularmente insensible a fuerzas superiores a $200 \text{ E}\cdot\text{nm}^{-1}$. Al evaluar distintos parámetros proteína-dependientes, como son la profundidad del pozo ϵ , número de átomos, masa molecular y número de contactos, ninguno de ellos por sí solo mejoró significativamente la relación entre los resultados experimentales y los computacionales. Sin embargo, el orden de contacto absoluto [70], un parámetro que indica la proximidad a nivel de secuencia de los contactos en una proteína, mejora considerablemente la relación entre estos datos ($R=0,934$).

Las tasas de plegamiento están relacionadas linealmente con la función exponente del orden de contacto absoluto, debido a que para un polímero desplegado con monómeros de igual largo este valor está directamente relacionado con la distancia promedio entre dos átomos que pueden formar contactos nativos. No es obvio el motivo detrás de una mejora en la correlación entre fuerzas experimentales y computacionales al ponderarlas por este parámetro; sin embargo, cabe mencionar que el orden de contacto relativo, equivalente al orden de contacto normalizado por el largo del polímero, tiene un efecto adverso en la misma. Sin embargo, la mejora de las correlaciones con la utilización de este parámetro sugiere que tanto el efecto del tamaño de la proteína como la cercanía de sus contactos sobre la fuerza de ruptura están subestimados en modelos basados en estructura atomísticos. A pesar de las complicaciones mencionadas, cabe destacar que este enfoque sigue siendo más exacto para describir procesos de desplegamiento mecánico que dinámicas moleculares convencionales, ya que se demostró que i) las fuerzas de ruptura a velocidades aproximadas a las experimentales siempre siguen el comportamiento experimental propuesto por Hummer-Szabo [131], ii) es posible distinguir entre proteínas mecanoresistentes ($F_m > 200$ pN) y mecanosensibles ($F_m < 100$ pN), iii) se observan eventos de ruptura discretos que se asemejan a los observados experimentalmente, y iv) es posible de obtener una gran cantidad de datos con pocos recursos computacionales.

Desplegamiento mecánico de RfaH

Al estirar la proteína RfaH completa en ambos estados se observa que los dominios CTD y NTD que forman RfaH llevan a cabo desplegamiento independiente, acorde a lo propuesto en la hipótesis. Esto es bastante interesante puesto que, a pesar de que en su estado beta-plegado RfaH es esencialmente idéntica a NusG, observada como dos dominios independientes en solución, su estado alfa-helicoidal está caracterizado como una proteína globular, y de sólo un dominio aparente. A pesar de estas grandes diferencias estructurales, el desplegamiento mecánico de RfaH siempre se inicia por el CTD y termina con el NTD, siendo la fuerza y eventos de desplegamiento de este último dominio independiente de si el estado inicial es alfa- o beta-plegado.

Estos resultados están de acuerdo con resultados computacionales publicados por nuestro laboratorio utilizando SBM de carbono alfa, que indican que la transformación espontánea y reversible de RfaH no requiere de una perturbación en la estructura del NTD [27]. Adicionalmente, el desplegamiento de ambos estados de RfaH es distinto de lo esperado; por un lado, se observa que el α CTD no posee homogeneidad estructural, ya que los residuos finales de la hélice 2, y gran parte de la hélice 1 son irrelevantes para la estabilidad mecánica puesto que se encuentran desplegados previo al evento mecanoresistente, contradiciendo directamente una publicación de SMD convencional reciente [22]. Otra gran discrepancia con tal trabajo es que en estas simulaciones se observa que la disociación ocurre simultáneamente al desplegamiento, y en ningún momento se observa un intermediario disociado con estructura helicoidal.

Por otro lado, el desplegamiento del β CTD parece ocurrir a través de la formación de una especie intermediaria que explica datos experimentales obtenidos por nuestro laboratorio. Esta especie intermediaria corresponde a un lámina- β formada por las tres hojas más grandes del CTD, y expone su núcleo hidrofóbico al solvente. Cabe resaltar que apenas 35 residuos componen este intermediario, y corresponde a la misma región mecanoresistente predicha para α RfaH. Esto acentúa las interrogantes respecto a la estabilidad dual de RfaH, pues estos estudios predicen que corresponde a una región mínima que logra tener estructura β -plegada en solución, y que además es responsable de asociarse en forma α -helicoidal al NTD.

La formación de esta estructura β -intermediario genera un cambio respecto a lo que se pensaba del plegamiento de RfaH, pues a pesar de que en literatura se ha propuesto la formación de una especie intermediaria o no completamente desestructurada para dar lugar al β CTD desde α CTD [21], [22], [27], en ningún caso se ha observado la formación de un intermediario con éstas características de estructura secundaria tan marcadas. Esto puede originarse en los efectos que tiene el NTD sobre el plegamiento de RfaH que no es evaluado al conectar ambos estados de plegamiento mediante el estado CTD-desplegado, y enfatiza la pregunta propuesta por mucho tiempo respecto a si la vía de transformación de RfaH es común con las vías de plegamiento de cada estado individualmente. La comparación con resultados preliminares de la transformación espontánea de RfaH con

SBM atomísticos parecen indicar que efectivamente hay diferencias importantes en la formación del β CTD desde un estado inicial desplegado respecto un estado inicial α CTD.

Desplegamiento reversible de RfaH

Al optimizar las condiciones de desplegado para lograr un replegamiento que siga la vía inversa del desplegado, se observó gran histéresis entre ambos procesos, principalmente en el caso del α CTD. Esto es probablemente debido a que gran parte de los contactos que estabilizan esta estructura se encuentran en la interfaz con el NTD, y por lo tanto con una baja probabilidad de formación de contacto durante la mayor parte del tiempo.

Es precisamente en este escenario donde se muestra una de las mayores desventajas de los modelos basados en estructura más simples: su función de energía potencial. Como se mencionó previamente, estos modelos utilizan una función tipo Lennard-Jones que se vuelve nula a distancia de 8 Å. Esto es suficiente para describir el estado nativo, donde los contactos (electrostáticos, Van der Waals o puente hidrógeno) están minimizados y en distancias que oscilan entre los 3,5 – 5 Å dependiendo de su naturaleza. Sin embargo, en el estado desplegado estos contactos se deben formar, y las interacciones electrostáticas para la mayoría de los átomos con cargas puntuales o parciales se vuelve nula a $\approx 1,6$ Å, el doble de alcance del valor utilizado durante la simulación. Esto es particularmente importante en experimentos de estiramiento, donde la coordenada de reacción maximiza la distancia entre residuos.

En este modelo en particular no se evaluó el potencial electrostático, por lo que no existen interacciones que se formen más allá de 8 Å, dando lugar a la existencia de una distancia umbral en el replegamiento de α RfaH desde la cual simplemente no puede ocurrir el replegamiento, independiente de la velocidad con que se realice el experimento. El hecho experimental de que una interacción interdominio de puente salino estabilice enormemente el estado α CTD apoya la noción de que es la falta de evaluación electrostática el origen de la dramática histéresis observada en el replegamiento de α RfaH.

Por otro lado, se observa que el replegamiento del β CTD tiene una menor histéresis que su contraparte, siendo la formación del intermediario descrito previamente durante el

replegamiento el paso más lento. Se observa, que una vez toma lugar la formación de este intermediario, el plegamiento del barril- β es bastante rápido y casi al equilibrio con el paso inverso. Estos resultados contrastan dramáticamente lo observado por *targeted molecular dynamics* y modelos de Markov, que sugieren que la formación de las hebras $\beta 2$ y $\beta 3$ o $\beta 3$ y $\beta 4$ es muy rápida desde un estado más bien desestructurado [21], [22].

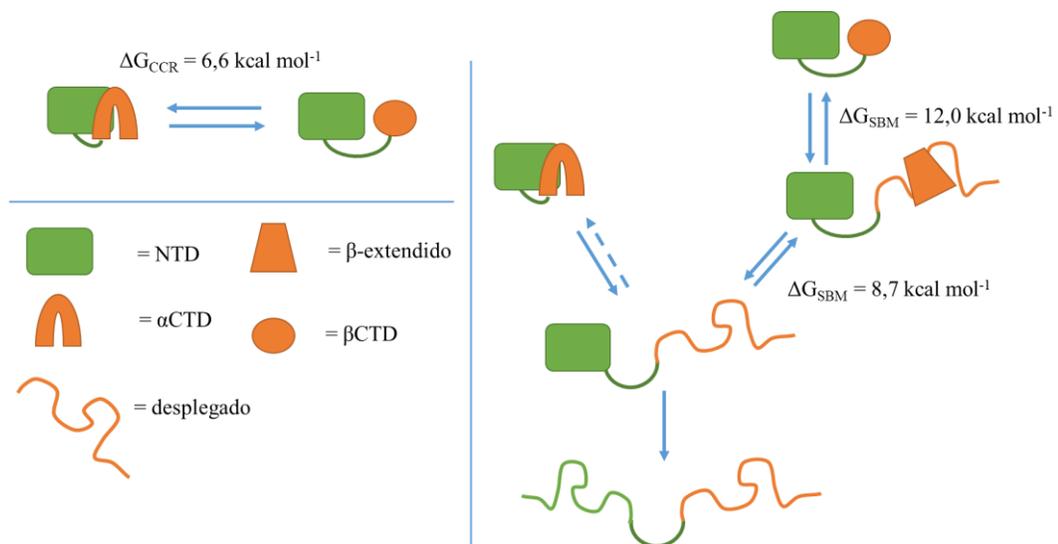
Cálculos de energía por confinamiento-conversión

A través de una integración termodinámica discontinua [55] se pudo estimar la diferencia de energía que involucra un cambio de estado como es la transformación de RfaH. A diferencia de métodos convencionales por perturbación de energía libre como TMD o SMD, el protocolo de confinamiento-conversión no requiere exploración exhaustiva y posterior reponderación para el cómputo de energía libre. Esto la exenta de la necesidad de una coordenada de reacción que defina la ruta de transformación estructural, evitando los clásicos artefactos observados computacionalmente por barreras cinéticas y exploración insuficiente [56]. La diferencia de energía predicha para la transición desde α RfaH a β RfaH corresponde a $6,5 \text{ kcal mol}^{-1}$, mostrando que este enfoque es capaz de predecir el estado más estable en solución, tal como se ha reportado previamente [55], [56], [62], entregando una diferencia de energía físicamente feasible.

En este caso, la diferencia de energía nos entrega un poco más de información que la termodinámica. Al estimar la diferencia de estabilidad de RfaH entre sus dos estados a nivel de residuo [62] ($\Delta\Delta G$), se observa que existe una acentuada polarización de la estabilidad en su estructura primaria observable en escalas de tiempo microscópicas, que es avalado por experimentos de espectrometría de masas de intercambio hidrógeno-deuterio [132], que entrega luces sobre la estabilidad local a una menor resolución, pero en escalas de tiempo macroscópicas, indicando que efectivamente las dinámicas moleculares de RfaH están explorando suficientemente el espacio configuracional y muestran características al equilibrio.

Esta información señala qué partes de la estructura primaria son mayormente responsables de estabilizar cada estado, y reportando también las que tienen mayor probabilidad relativa de ser desestabilizadas. Esto podemos considerarlo cierto porque, a

diferencia de otros fenómenos de conversión estructural como amiloidogénesis o agregación inespecífica, que tienen barreras cinéticas muy elevadas [133], las proteínas *transformer* llevan a cabo cambios estructurales y funcionales drásticos, que deben ocurrir múltiples veces durante su tiempo de vida al interior de la célula para asegurar su correcto funcionamiento [134]. Esto implica que independiente del mecanismo cinético (y si involucra o no el estado desplegado) ocurrirá una transformación estructural, donde probablemente las regiones inestables de cada estado serán las responsables de la salida del estado nativo. Una vez desestabilizada la estructura del CTD, al ocurrir una minimización del potencial de energía libre hacia el próximo mínimo energético, serán entonces los contactos nativos de las regiones comparablemente más estables de RfaH las encargadas de consolidar la próxima configuración a adoptar.



Esquema 2: Resultados generales del plegamiento de RfaH. Izquierda, diferencia de energía libre entre ambos estados de RfaH tal como está predicho por dinámicas de confinamiento. Derecha, vías de plegamiento de RfaH. Se observa que el desplegamiento desde ambos estados iniciales, α CTD y β CTD convergen en un estado intermedio NTD-plegado.

Conclusiones y perspectivas futuras

En esta investigación se exploró exhaustivamente el empleo de simulaciones computacionales utilizando modelos basados en estructura para estudiar el mecanismo y predecir las fuerzas de desplegamiento de proteínas mediante espectroscopía de fuerza, en base a correlaciones con proteínas que han sido reportados experimentalmente. Se evaluó las características mecánicas del desplegamiento de 35 proteínas bajo distintas condiciones de temperatura y velocidad de estiramiento, y hasta dónde permiten predecir fuerzas de ruptura y trazas de desplegamiento. Adicionalmente, utilizando este mismo enfoque, se logró desplegar y replegar el CTD de la proteína RfaH en sus dos estados nativos descritos en literatura [8], [9], observando características mecánicas congruentes con simulaciones equivalentes con granularidad de carbono alfa [27], mientras que con grandes discrepancias a las observadas con análogos en dinámica molecular convencional [22]. Estas simulaciones sugieren fuertemente la presencia de un intermediario en el plegamiento del barril- β de RfaH apoyado por evidencias experimentales preliminares, jamás descrito para miembros de esta familia de proteínas.

Utilizando un protocolo de integración termodinámica discontinua mediante dinámica molecular se logró estimar la diferencia de energía libre de transformación desde α RfaH a $7 \text{ kcal}\cdot\text{mol}^{-1}$. Adicionalmente, la estimación del aporte a esta energía a nivel de residuo permitió encontrar claras regiones de estabilidad diferencial al interior del CTD, entregando luces sobre los motores detrás de tal fenómeno de metamorfosis. Esta información está cercanamente relacionada con la estabilidad diferencial de péptidos conformando el CTD en ambos estados, evaluada a través de la incorporación de deuterio seguida por espectrometría de masa.

Globalmente, estos resultados presentan alta consistencia entre los distintos enfoques de dinámica molecular utilizados para evaluar la estabilidad de ambos estados de RfaH. Se observan altas fluctuaciones en los terminales libres de las hélices 1 y 2 del α CTD a través de dinámica molecular convencional, de estiramiento, acelerada y por confinamiento, que son corroboradas experimentalmente por espectrometría de masa de intercambios hidrógeno deuterio. Por otro lado, la estabilidad del NTD no parece verse afectado por el estado de plegamiento del CTD, incluso cuando la forma α -helicoidal de

este último depende completamente de la presencia del primero. Esto se observa explícitamente en simulaciones de estiramiento, donde se observa que ambos dominios presentan desplegamiento independiente, mientras que a través de confinamiento-conversión al observar que la mayor parte del NTD no tiene grandes preferencias hacia algún estado nativo en particular.

Estos resultados añaden información sobre las dinámicas de plegamiento de RfaH en escala micro- y macroscópica, al y fuera del equilibrio; indicando que efectivamente ambos dominios de RfaH tienen plegamiento independiente, en los que el estado nativo α RfaH es más estable que β RfaH. A pesar de que la hipótesis fue resuelta en un sentido amplio, aún quedan preguntas respecto a las múltiples especies involucradas en la dinámica de plegamiento de RfaH sin evidencia experimental, mientras que durante el desarrollo de la tesis han surgido incluso muchas nuevas interrogantes. A pesar de esto, gran parte de la información computacional presentada aquí tiene al menos un correlato experimental, entregando confianza para proponer con seguridad hipótesis que pueden ser resueltas por posteriores iteraciones de enfoques multidisciplinarios.

VI. Referencias

- [1] M. J. Bailey, C. Hughes, and V. Koronakis, "RfaH and the ops element, components of a novel system controlling bacterial transcription elongation," *Mol Microbiol*, vol. 26, no. 5, pp. 845–851, 1997.
- [2] G. A. Belogurov, R. A. Mooney, V. Svetlov, R. Landick, and I. Artsimovitch, "Functional specialization of transcription elongation factors," *EMBO J*, vol. 28, no. 2, pp. 112–122, 2009.
- [3] I. Artsimovitch and R. Landick, "The transcriptional regulator RfaH stimulates RNA chain synthesis after recruitment to elongation complexes by the exposed nontemplate DNA strand," *Cell*, vol. 109, no. 2, pp. 193–203, 2002.
- [4] R. A. Mooney, S. E. Davis, J. M. Peters, J. L. Rowland, A. Z. Ansari, and R. Landick, "Regulator trafficking on bacterial transcription units in vivo," *Mol. Cell*, vol. 33, no. 1, pp. 97–108, 2009.
- [5] S. L. Sullivan and M. E. Gottesman, "Requirement for E. coli NusG protein in factor-dependent transcription termination," *Cell*, vol. 68, no. 5, pp. 989–994, 1992.
- [6] I. Artsimovitch and R. Landick, "Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals," *Proc. Natl. Acad. Sci.*, vol. 97, no. 13, pp. 7090–7095, 2000.
- [7] C. J. Cardinale, R. S. Washburn, V. R. Tadigotla, L. M. Brown, M. E. Gottesman, and E. Nudler, "Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in E. coli," *Science (80-.)*, vol. 320, no. 5878, pp. 935–938, 2008.
- [8] M. N. Vassilyeva, V. Svetlov, S. Klyuyev, Y. D. Devedjiev, I. Artsimovitch, and D. G. Vassilyev, "Crystallization and preliminary crystallographic analysis of the transcriptional regulator RfaH from Escherichia coli and its complex with ops DNA," *Acta Crystallogr Sect F Struct Biol Cryst Commun*, vol. 62, no. Pt 10, pp. 1027–1030, 2006.
- [9] B. M. Burmann *et al.*, "An α helix to β barrel domain switch transforms the transcription factor RfaH into a translation factor," *Cell*, vol. 150, no. 2, pp. 291–303, 2012.
- [10] T. Steiner, J. T. Kaiser, S. Marinkovic, R. Huber, and M. C. Wahl, "Crystal structures of transcription factor NusG in light of its nucleic acid- and protein-binding activities," *EMBO J*, vol. 21, no. 17, pp. 4641–4653, 2002.
- [11] A. Hirtreiter *et al.*, "Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif," *Nucleic Acids Res*, vol. 38, no. 12, pp. 4040–4051, 2010.
- [12] A. Sevostyanova, G. A. Belogurov, R. A. Mooney, R. Landick, and I. Artsimovitch, "The β subunit gate loop is required for RNA polymerase modification by RfaH and NusG," *Mol. Cell*, vol. 43, no. 2, pp. 253–262, 2011.
- [13] F. W. Martinez-Rucobo, S. Sainsbury, A. C. Cheung, and P. Cramer, "Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity," *EMBO J*, vol. 30, no. 7, pp. 1302–1310, 2011.
- [14] G. A. Belogurov, A. Sevostyanova, V. Svetlov, and I. Artsimovitch, "Functional regions of the N-terminal domain of the antiterminator RfaH," *Mol Microbiol*, vol. 76, no. 2, pp. 286–301, 2010.
- [15] G. A. Belogurov *et al.*, "Structural basis for converting a general transcription factor into an operon-specific virulence regulator," *Mol Cell*, vol. 26, no. 1, pp. 117–129, 2007.
- [16] G. A. Belogurov, A. Sevostyanova, V. Svetlov, and I. Artsimovitch, "Functional regions of the N-terminal domain of the antiterminator RfaH," *Mol Microbiol*, vol. 76, no. 2, pp. 286–301, 2010.
- [17] I. Artsimovitch and R. Landick, "The transcriptional regulator RfaH stimulates RNA chain synthesis after recruitment to elongation complexes by the exposed nontemplate DNA strand," *Cell*, vol. 109, no. 2, pp. 193–203, 2002.
- [18] A. Sevostyanova, V. Svetlov, D. G. Vassilyev, and I. Artsimovitch, "The elongation factor RfaH and the initiation factor sigma bind to the same site on the transcription elongation complex," *Proc Natl Acad Sci U S A*, vol. 105, no. 3, pp. 865–870, 2008.
- [19] P. K. Weiner and P. A. Kollman, "AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions," *J. Comput. Chem.*, vol. 2, no. 3, pp. 287–303, 1981.
- [20] N. Balasco, D. Barone, and L. Vitagliano, "Structural conversion of the transformer protein RfaH: new insights derived from protein structure prediction and molecular dynamics simulations," *J. Biomol. Struct. Dyn.*, vol. 33, no. 10, pp. 2173–2179, 2015.
- [21] S. Li *et al.*, "Mechanism of the all- α to all- β conformational transition of RfaH-CTD: Molecular dynamics simulation and markov state model," *J. Chem. Theory Comput.*, vol. 10, no. 6, pp. 2255–2264, 2014.
- [22] J. B. Gc, B. S. Gerstman, and P. P. Chapagain, "The Role of the Interdomain Interactions on RfaH Dynamics and Conformational Transformation," *J Phys Chem B*, vol. 119, no. 40, pp. 12750–12759, 2015.

- [23] J.-H. Prinz *et al.*, “Markov models of molecular kinetics: Generation and validation,” *J. Chem. Phys.*, vol. 134, no. 17, p. 174105, 2011.
- [24] Y. Sugita and Y. Okamoto, “Replica-exchange molecular dynamics method for protein folding,” *Chem. Phys. Lett.*, vol. 314, no. 1, pp. 141–151, 1999.
- [25] Y. R. Bhandari, B. S. Gerstman, and P. P. Chapagain, “Molecular dynamics investigations of the α -helix to β -barrel conformational transformation in the RfaH transcription factor,” *J. Phys. Chem. B*, vol. 118, no. 19, pp. 5101–5108, 2014.
- [26] J. K. Noel and J. N. Onuchic, “The many faces of structure-based potentials: from protein folding landscapes to structural characterization of complex biomolecules,” in *Computational Modeling of Biological Systems*, Springer, 2012, pp. 31–54.
- [27] C. A. Ramírez-Sarmiento, J. K. Noel, S. L. Valenzuela, and I. Artsimovitch, “Interdomain contacts control native state switching of RfaH on a dual-funneled landscape,” *PLoS Comput Biol*, vol. 11, no. 7, p. e1004379, 2015.
- [28] S. K. Tomar, S. H. Knauer, M. NandyMazumdar, P. Rösch, and I. Artsimovitch, “Interdomain contacts control folding of transcription factor RfaH,” *Nucleic Acids Res.*, vol. 41, no. 22, pp. 10077–10085, 2013.
- [29] J. C. Phillips *et al.*, “Scalable molecular dynamics with NAMD,” *J Comput Chem*, vol. 26, no. 16, pp. 1781–1802, 2005.
- [30] B. Isralewitz, M. Gao, and K. Schulten, “Steered molecular dynamics and mechanical functions of proteins,” *Curr Opin Struct Biol*, vol. 11, no. 2, pp. 224–230, 2001.
- [31] J. Comer, J. C. Gumbart, J. Henin, T. Lelievre, A. Pohorille, and C. Chipot, “The adaptive biasing force method: everything you always wanted to know but were afraid to ask,” *J Phys Chem B*, vol. 119, no. 3, pp. 1129–1151, 2015.
- [32] V. Babin, C. Roland, and C. Sagui, “Adaptively biased molecular dynamics for free energy calculations,” *J Chem Phys*, vol. 128, no. 13, p. 134101, 2008.
- [33] S. Xun, F. Jiang, and Y.-D. D. Wu, “Intrinsically disordered regions stabilize the helical form of the C-terminal domain of RfaH: A molecular dynamics study,” *Bioorganic Med. Chem.*, vol. 24, no. 20, pp. 4970–4977, 2016.
- [34] C. Cecconi, E. A. Shank, C. Bustamante, and S. Marqusee, “Direct observation of the three-state folding of a single protein molecule,” *Science (80-.)*, vol. 309, no. 5743, pp. 2057–2060, 2005.
- [35] S. B. Smith, Y. CUI, and C. Bustamante, “Optical-trap force transducer that operates by direct measurement of light momentum,” *Methods Enzymol.*, vol. 361, pp. 134–162, 2002.
- [36] C. Cecconi, E. A. Shank, F. W. Dahlquist, S. Marqusee, and C. Bustamante, “Protein-DNA chimeras for single molecule mechanical folding studies with the optical tweezers,” *Eur. Biophys. J.*, vol. 37, no. 6, pp. 729–738, 2008.
- [37] M. T. Woodside and S. M. Block, “Reconstructing folding energy landscapes by single-molecule force spectroscopy,” *Annu. Rev. Biophys.*, vol. 43, p. 19, 2014.
- [38] G. E. Crooks, “Path-ensemble averages in systems driven far from equilibrium,” *Phys. Rev. E*, vol. 61, no. 3, p. 2361, 2000.
- [39] E. H. Feng and G. E. Crooks, “Length of time’s arrow,” *Phys. Rev. Lett.*, vol. 101, no. 9, p. 90602, 2008.
- [40] G. Bell, M. Dembo, and P. Bongrand, “Cell adhesion. Competition between nonspecific repulsion and specific bonding,” *Biophys. J.*, vol. 45, no. 6, pp. 1051–64, 1984.
- [41] S. N. Zhurkov, “Kinetic concept of the strength of solids,” *Int. J. Fract.*, vol. 26, no. 4, pp. 295–307, 1984.
- [42] J. Stigler, F. Ziegler, A. Gieseke, J. C. M. Gebhardt, and M. Rief, “The complex folding network of single calmodulin molecules,” *Science*, vol. 334, no. 6055, pp. 512–516, 2011.
- [43] E. H. Lee, J. Hsin, M. Sotomayor, G. Comellas, and K. Schulten, “Discovery through the computational microscope,” *Structure*, vol. 17, no. 10, pp. 1295–1306, 2009.
- [44] J. K. Noel *et al.*, “SMOG 2: A Versatile Software Package for Generating Structure-Based Models,” *PLoS Comput Biol*, vol. 12, no. 3, p. e1004794, 2016.
- [45] L. Sun, J. K. Noel, J. I. Sulkowska, H. Levine, and J. N. Onuchic, “Connecting thermal and mechanical protein (un) folding landscapes,” *Biophys. J.*, vol. 107, no. 12, pp. 2950–2961, 2014.
- [46] J. K. Noel, P. C. Whitford, K. Y. Sanbonmatsu, and J. N. Onuchic, “SMOG@ctbp: Simplified deployment of structure-based models in GROMACS,” *Nucleic Acids Res.*, vol. 38, no. SUPPL. 2, 2010.
- [47] J. K. Noel, P. C. Whitford, and J. N. Onuchic, “The shadow map: A general contact definition for capturing the dynamics of biomolecular folding and function,” *J. Phys. Chem. B*, vol. 116, no. 29, pp. 8692–8702, 2012.
- [48] C. Clementi, H. Nymeyer, and J. N. Onuchic, “Topological and energetic factors: what determines the structural details of the transition state ensemble and ‘en-route’

- intermediates for protein folding? An investigation for small globular proteins,” *J. Mol. Biol.*, vol. 298, no. 5, pp. 937–953, 2000.
- [49] J. I. Sułkowska and M. Cieplak, “Selection of optimal variants of Gō-like models of proteins through studies of stretching,” *Biophys. J.*, vol. 95, no. 7, pp. 3174–3191, 2008.
- [50] S. Pronk *et al.*, “GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit,” *Bioinformatics*, vol. 29, no. 7, pp. 845–854, 2013.
- [51] B. Roux, “The calculation of the potential of mean force using computer simulations,” *Comput. Phys. Commun.*, vol. 91, no. 1–3, pp. 275–282, 1995.
- [52] M. Sikora, J. I. Sulkowska, B. S. Witkowski, and M. Cieplak, “BSDB: The biomolecule stretching database,” *Nucleic Acids Res.*, vol. 39, no. SUPPL. 1, 2011.
- [53] A. Valbuena *et al.*, “On the remarkable mechanostability of scaffoldins and the mechanical clamp motif,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 33, pp. 13791–6, 2009.
- [54] M. Sikora, J. I. Sułkowska, and M. Cieplak, “Mechanical strength of 17 134 model proteins and cysteine slipknots,” *PLoS Comput. Biol.*, vol. 5, no. 10, 2009.
- [55] M. D. Tyka, A. R. Clarke, and R. B. Sessions, “An efficient, path-independent method for free-energy calculations,” *J. Phys. Chem. B*, vol. 110, no. 34, pp. 17212–17220, 2006.
- [56] M. Cecchini, S. V. Krivov, M. Spichty, and M. Karplus, “Calculation of free-energy differences by confinement simulations. Application to peptide conformers,” *J. Phys. Chem. B*, vol. 113, no. 29, pp. 9728–9740, 2009.
- [57] T. Ichiye and M. Karplus, “Collective motions in proteins: A covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations,” *Proteins Struct. Funct. Bioinforma.*, vol. 11, no. 3, pp. 205–217, 1991.
- [58] V. Ovchinnikov, M. Cecchini, and M. Karplus, “A simplified confinement method for calculating absolute free energies and free energy and entropy differences,” *J. Phys. Chem. B*, vol. 117, no. 3, pp. 750–762, 2013.
- [59] G. D. Hawkins, C. J. Cramer, and D. G. Truhlar, “Pairwise solute descreening of solute charges from a dielectric medium,” *Chem. Phys. Lett.*, vol. 246, no. 1–2, pp. 122–129, 1995.
- [60] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, “Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes,” *J. Comput. Phys.*, vol. 23, no. 3, pp. 327–341, 1977.
- [61] B. Qian *et al.*, “High-resolution structure prediction and the crystallographic phase problem,” *Nature*, vol. 450, no. 7167, pp. 259–64, 2007.
- [62] A. Roy, A. Perez, K. A. Dill, and J. L. MacCallum, “Computing the relative stabilities and the per-residue components in protein conformational changes,” *Structure*, vol. 22, no. 1, pp. 168–175, 2014.
- [63] G. Crooks, “Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences,” *Phys. Rev. E*, vol. 60, no. 3, pp. 2721–2726, 1999.
- [64] V. Svetlov, G. A. Belogurov, E. Shabrova, D. G. Vassilyev, and I. Artsimovitch, “Allosteric control of the RNA polymerase by the elongation factor RfaH,” *Nucleic Acids Res.*, vol. 35, no. 17, pp. 5694–5705, 2007.
- [65] G. BERTANI, “Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*,” *J. Bacteriol.*, vol. 62, no. 3, pp. 293–300, 1951.
- [66] M. M. Bradford, “A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding,” *Anal. Biochem.*, vol. 72, no. 1–2, pp. 248–254, 1976.
- [67] K. Tartof and C. Hobbs, “Improved media for growing plasmid and cosmid clones,” *Focus (Madison)*, vol. 9, no. 2, p. 10, 1987.
- [68] R. J. Simpson, “Estimation of free thiols and disulfide bonds using Ellman’s reagent,” *Cold Spring Harb. Protoc.*, vol. 3, no. 9, 2008.
- [69] E. A. Shank, C. Cecconi, J. W. Dill, S. Marqusee, and C. Bustamante, “The folding cooperativity of a protein is controlled by its chain topology,” *Nature*, vol. 465, no. 7298, pp. 637–640, 2010.
- [70] D. N. Ivankov, S. O. Garbuzynskiy, E. Alm, K. W. Plaxco, D. Baker, and A. V Finkelstein, “Contact order revisited: influence of protein size on the folding rate,” *Protein Sci.*, vol. 12, no. 9, pp. 2057–2062, 2003.
- [71] K. J. Smith, S. W. Reid, D. I. Stuart, A. J. McMichael, E. Y. Jones, and J. I. Bell, “An altered position of the $\alpha 2$ helix of MHC class I is revealed by the crystal structure of HLA-B3501,” *Immunity*, vol. 4, no. 3, pp. 203–213, 1996.
- [72] M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub, “Reversible unfolding of individual titin immunoglobulin domains by AFM,” *Science (80-.)*, vol. 276, no. 5315, pp. 1109–1112, 1997.
- [73] M. Carrion-Vazquez *et al.*, “Mechanical and chemical unfolding of a single protein: A

- comparison,” *Biophysics (Oxf)*., vol. 96, no. March, pp. 3694–3699, 1999.
- [74] M. J. Lubienski, M. Bycroft, S. M. Freund, and a R. Fersht, “Three-dimensional solution structure and 13C assignments of barstar using nuclear magnetic resonance spectroscopy,” *Biochemistry*, vol. 33, no. 30, pp. 8866–77, 1994.
- [75] D. Sharma, G. Feng, D. Khor, G. Z. Genchev, H. Lu, and H. Li, “Stabilization provided by neighboring strands is critical for the mechanical stability of proteins,” *Biophys. J.*, vol. 95, no. 8, pp. 3935–3942, 2008.
- [76] W. Kremer *et al.*, “Solution NMR structure of the cold-shock protein from the hyperthermophilic bacterium *Thermotoga maritima*,” *Eur. J. Biochem.*, vol. 268, no. 9, pp. 2527–2539, 2001.
- [77] J. Schönfelder, R. Perez-Jimenez, and V. Muñoz, “A simple two-state protein unfolds mechanically via multiple heterogeneous pathways at single-molecule resolution,” *Nat. Commun.*, vol. 7, p. 11777, 2016.
- [78] S. Dhe-Paganon, E. D. Werner, Y.-I. Chi, and S. E. Shoelson, “Structure of the globular tail of nuclear lamin,” *J. Biol. Chem.*, vol. 277, no. 20, pp. 17381–17384, 2002.
- [79] M. Bera *et al.*, “Characterization of unfolding mechanism of human lamin A Ig fold by single-molecule force spectroscopy-implications in EDMD,” *Biochemistry*, vol. 53, no. 46, pp. 7247–7258, 2014.
- [80] S.-Y. Hsieh, T.-P. Ko, M.-Y. Tseng, W. Ku, K.-F. Chak, and H. S. Yuan, “A novel role of ImmE7 in the autoregulatory expression of the ColE7 operon and identification of possible RNase active sites in the crystal structure of dimeric ImmE7,” *EMBO J.*, vol. 16, no. 6, pp. 1444–1454, 1997.
- [81] C. Lv, C. Tan, M. Qin, D. Zou, Y. Cao, and W. Wang, “Low folding cooperativity of Hp35 revealed by single-molecule force spectroscopy and molecular dynamics simulation,” *Biophys. J.*, vol. 102, no. 8, pp. 1944–1951, 2012.
- [82] U. Magnusson, B. Salopek-Sondi, L. A. Luck, and S. L. Mowbray, “X-ray structures of the leucine-binding protein illustrate conformational changes and the basis of ligand specificity,” *J. Biol. Chem.*, vol. 279, no. 10, pp. 8747–8752, 2004.
- [83] H. C. Kotamarthi, R. Sharma, S. Narayan, S. Ray, and S. R. K. Ainavarapu, “Multiple unfolding pathways of leucine binding protein (LBP) probed by single-molecule force spectroscopy (SMFS),” *J. Am. Chem. Soc.*, vol. 135, no. 39, pp. 14768–14774, 2013.
- [84] R. P. Barnwal, M. K. Jobby, K. M. Devi, Y. Sharma, and K. V. R. Chary, “Solution Structure and Calcium-Binding Properties of M-Crystallin, A Primordial ???-Crystallin from Archaea,” *J. Mol. Biol.*, vol. 386, no. 3, pp. 675–689, 2009.
- [85] V. Ramanujam, H. C. Kotamarthi, and S. R. K. Ainavarapu, “Ca²⁺ binding enhanced mechanical stability of an archaeal crystallin,” *PLoS One*, vol. 9, no. 4, 2014.
- [86] T. Merz, S. K. Wetzel, S. Firbank, A. Plückthun, M. G. Grütter, and P. R. E. Mittl, “Stabilizing Ionic Interactions in a Full-consensus Ankyrin Repeat Protein,” *J. Mol. Biol.*, vol. 376, no. 1, pp. 232–240, 2008.
- [87] W. Lee, X. Zeng, H.-X. Zhou, V. Bennett, W. Yang, and P. E. Marszalek, “Full reconstruction of a vectorial protein folding pathway by atomic force microscopy and molecular dynamics simulations,” *J. Biol. Chem.*, vol. 285, no. 49, pp. 38167–72, 2010.
- [88] S. Nag *et al.*, “Ca²⁺ binding by domain 2 plays a critical role in the activation and stabilization of gelsolin,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 33, pp. 13713–13718, 2009.
- [89] C. Lv *et al.*, “Single-molecule force spectroscopy reveals force-enhanced binding of calcium ions by gelsolin,” *Nat. Commun.*, vol. 5, p. 4623, 2014.
- [90] W. Wei *et al.*, “Structural Insights and the Surprisingly Low Mechanical Stability of the Au-S Bond in the Gold-Specific Protein GolB,” *J. Am. Chem. Soc.*, vol. 137, no. 49, pp. 15358–15361, 2015.
- [91] B. Kuhlman, G. Dantas, G. C. Ireton, G. Varani, B. L. Stoddard, and D. Baker, “Design of a novel globular protein fold with atomic-level accuracy,” *Science*, vol. 302, no. 5649, pp. 1364–8, 2003.
- [92] D. H. Goldman, C. M. Kaiser, A. Milin, M. Righini, I. Tinoco, and C. Bustamante, “Mechanical force releases nascent chain-mediated ribosome arrest in vitro and in vivo,” *Science (80-.)*, vol. 348, no. 6233, pp. 457–460, 2015.
- [93] Y. Xing, W. K. Clements, D. Kimelman, and W. Xu, “Crystal structure of a beta-catenin/axin complex suggests a mechanism for the beta-catenin destruction complex,” *Genes Dev.*, vol. 17, no. 22, pp. 2753–64, 2003.
- [94] A. Valbuena, A. M. Vera, J. Oroz, M. Menéndez, and M. Carrión-Vázquez, “Mechanical properties of β -catenin revealed by single-molecule experiments,” *Biophys. J.*, vol. 103, no. 8, pp. 1744–1752, 2012.
- [95] P. O. Heidarsson *et al.*, “The C-terminal tail of human neuronal calcium sensor 1 regulates the conformational stability of the Ca²⁺-activated state,” *J. Mol. Biol.*, vol. 417, no. 1, pp. 51–64, 2012.
- [96] P. O. Heidarsson, M. M. Naqvi, M. R. Otazo, A. Mossa, B. B. Kragelund, and C. Cecconi,

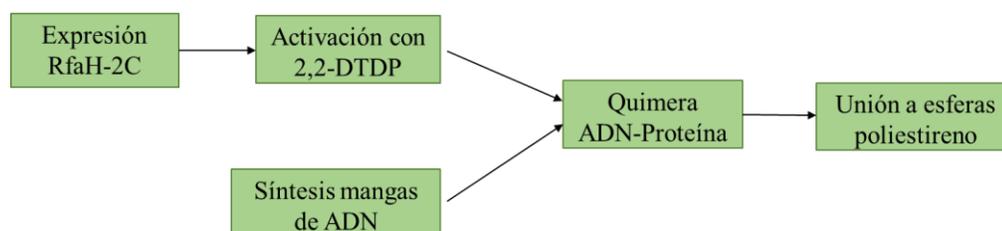
- “Direct single-molecule observation of calcium-dependent misfolding in human neuronal calcium sensor-1.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 36, pp. 13069–74, 2014.
- [97] J. Pascual, M. Pfuhl, D. Walther, M. Saraste, and M. Nilges, “Solution structure of the spectrin repeat: a left-handed antiparallel triple-helical coiled-coil.,” *J. Mol. Biol.*, vol. 273, no. 3, pp. 740–751, 1997.
- [98] P. F. Lenne, A. J. Raae, S. M. Altmann, M. Saraste, and J. K. H. Hørbær, “States and transitions during forced unfolding of a single spectrin repeat,” *FEBS Lett.*, vol. 476, no. 3, pp. 124–128, 2000.
- [99] L. J. Shimon *et al.*, “A cohesin domain from *Clostridium thermocellum*: the crystal structure provides new insights into cellulosome assembly.,” *Structure*, vol. 5, no. 3, pp. 381–390, 1997.
- [100] S. Spinelli, H. P. Fiérobe, a Belaïch, J. P. Belaïch, B. Henrissat, and C. Cambillau, “Crystal structure of a cohesin module from *Clostridium cellulolyticum*: implications for dockerin recognition.,” *J. Mol. Biol.*, vol. 304, no. 2, pp. 189–200, 2000.
- [101] G. Yang *et al.*, “Solid-state synthesis and mechanical unfolding of polymers of T4 lysozyme.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 97, no. 1, pp. 139–44, 2000.
- [102] M. Bycroft, S. Ludvigsen, a R. Fersht, and F. M. Poulsen, “Determination of the three-dimensional solution structure of barnase using nuclear magnetic resonance spectroscopy.,” *Biochemistry*, vol. 30, no. 35, pp. 8697–701, 1991.
- [103] R. B. Best, B. Li, a Steward, V. Daggett, and J. Clarke, “Can non-mechanical proteins withstand force? Stretching barnase by atomic force microscopy and molecular dynamics simulation.,” *Biophys. J.*, vol. 81, no. 4, pp. 2344–2356, 2001.
- [104] X. Shao, I. Fernandez, T. C. Südhof, and J. Rizo, “Solution structures of the Ca²⁺-free and Ca²⁺-bound C2A domain of synaptotagmin I: does Ca²⁺ induce a conformational change?,” *Biochemistry*, vol. 37, no. 46, pp. 16106–16115, 1998.
- [105] M. Carrion-Vazquez, A. F. Oberhauser, T. E. Fisher, P. E. Marszalek, H. Li, and J. M. Fernandez, “Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering,” *Progress in Biophysics and Molecular Biology*, vol. 74, no. 1–2, pp. 63–91, 2000.
- [106] H. Kuboniwa, N. Tjandra, S. Grzesiek, H. Ren, C. B. Klee, and A. Bax, “Solution structure of calcium-free calmodulin,” *Nat. Struct. Biol.*, vol. 2, no. 9, pp. 768–776, 1995.
- [107] R. B. Sutton, J. A. Ernst, and A. T. Brunger, “Crystal structure of the cytosolic C2A-C2B domains of synaptotagmin III: Implications for Ca²⁺-independent SNARE complex interaction,” *J. Cell Biol.*, vol. 147, no. 3, pp. 589–598, 1999.
- [108] K. Brejc *et al.*, “Structural basis for dual excitation and photoisomerization of the *Aequorea victoria* green fluorescent protein,” *Proc. Natl. Acad. Sci.*, vol. 94, no. 6, pp. 2306–2311, 1997.
- [109] H. Dietz and M. Rief, “Exploring the energy landscape of GFP by single-molecule mechanical experiments.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 46, pp. 16192–16197, 2004.
- [110] D. J. Leahy, I. Aukhil, and H. P. Erickson, “2.0 Å crystal structure of a four-domain segment of human fibronectin encompassing the RGD loop and synergy region,” *Cell*, vol. 84, no. 1, pp. 155–164, 1996.
- [111] L. Li, H. H. L. Huang, C. L. Badilla, and J. M. Fernandez, “Mechanical unfolding intermediates observed by single-molecule force spectroscopy in a fibronectin type III module,” *J. Mol. Biol.*, vol. 345, no. 4, pp. 817–826, 2005.
- [112] A. F. Oberhauser, C. Badilla-Fernandez, M. Carrion-Vazquez, and J. M. Fernandez, “The mechanical hierarchies of fibronectin observed with single-molecule AFM,” *J. Mol. Biol.*, vol. 319, no. 2, pp. 433–447, 2002.
- [113] A. Sharma, J. A. Askari, M. J. Humphries, E. Y. Jones, and D. I. Stuart, “Crystal structure of a heparin- and integrin-binding segment of human fibronectin.,” *EMBO J.*, vol. 18, no. 6, pp. 1468–79, 1999.
- [114] O. Mayans, J. Wuerges, S. Canela, M. Gautel, and M. Wilmanns, “Structural evidence for a possible role of reversible disulphide bridge formation in the elasticity of the muscle protein titin,” *Structure*, vol. 9, no. 4, pp. 331–340, 2001.
- [115] H. Li and J. M. Fernandez, “Mechanical design of the first proximal Ig domain of human cardiac titin revealed by single molecule force spectroscopy,” *J. Mol. Biol.*, vol. 334, no. 1, pp. 75–86, 2003.
- [116] Y. Cao and H. Li, “Polyprotein of GB1 is an ideal artificial elastomeric protein.,” *Nat. Mater.*, vol. 6, no. 2, pp. 109–114, 2007.
- [117] J. W. O’Neill, D. E. Kim, D. Baker, and K. Y. J. Zhang, “Structures of the B1 domain of protein L from *Peptostreptococcus magnus* with a tyrosine to tryptophan substitution,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 57, no. 4, pp. 480–487, 2001.
- [118] D. J. Brockwell *et al.*, “Mechanically Unfolding the Small, Topologically Simple Protein L,” *Biophys. J.*, vol. 89, no. 1, pp. 506–519, 2005.
- [119] P. Michaely, D. R. Tomchick, M. Machius, and R. G. W. Anderson, “Crystal structure of

- a 12 ANK repeat stack from human ankyrinR,” *EMBO J.*, vol. 21, no. 23, pp. 6387–6396, 2002.
- [120] M. Gao, D. Craig, O. Lequin, I. D. Campbell, V. Vogel, and K. Schulten, “Structure and functional significance of mechanically unfolded fibronectin type III intermediates,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 25, pp. 14784–14789, 2003.
- [121] T. Gallagher, P. Alexander, P. Bryan, and G. L. Gilliland, “Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR,” *Biochemistry*, vol. 33, no. 15, pp. 4721–4729, 1994.
- [122] R. B. Sutton, B. A. Davletov, A. M. Berghuis, T. C. Sudhof, and S. R. Sprang, “Structure of the first C2 domain of synaptotagmin I: A novel Ca²⁺/phospholipid-binding fold,” *Cell*, vol. 80, no. 6, pp. 929–938, 1995.
- [123] D. J. Leahy, W. a Hendrickson, I. Aukhil, and H. P. Erickson, “Structure of a fibronectin type III domain from tenascin phased by MAD analysis of the selenomethionyl protein,” *Science*, vol. 258, no. 5084, pp. 987–991, 1992.
- [124] A. F. Oberhauser, P. E. Marszalek, H. P. Erickson, and J. M. Fernandez, “The molecular elasticity of the extracellular matrix protein tenascin,” *Nature*, vol. 393, no. 6681, pp. 181–5, 1998.
- [125] S. Improta, A. S. Politou, and A. Pastore, “Immunoglobulin-like modules from titin I-band: extensible components of muscle elasticity,” *Structure*, vol. 4, no. 3, pp. 323–337, 1996.
- [126] A. L. Main, T. S. Harvey, M. Baron, J. Boyd, and I. D. Campbell, “The three-dimensional structure of the tenth type III module of fibronectin: An insight into RGD-mediated interactions,” *Cell*, vol. 71, no. 4, pp. 671–678, 1992.
- [127] Y. Oberdörfer, H. Fuchs, and A. Janshoff, “Conformational analysis of native fibronectin by means of force spectroscopy,” *Langmuir*, vol. 16, no. 26, pp. 9955–9958, 2000.
- [128] S. Vijay-Kumar, C. E. Bugg, and W. J. Cook, “Structure of ubiquitin refined at 1.8 Å resolution,” *J. Mol. Biol.*, vol. 194, no. 3, pp. 531–44, 1987.
- [129] C.-L. Chyan *et al.*, “Reversible mechanical unfolding of single ubiquitin molecules,” *Biophys. J.*, vol. 87, no. 6, pp. 3995–4006, 2004.
- [130] M. Wikstrom, T. Drakenberg, S. Forsen, U. Sjöbring, and L. Björck, “Three-dimensional solution structure of an immunoglobulin light chain-binding domain of protein L. comparison with the IgG-binding domains of protein G,” *Biochemistry*, vol. 33, no. 47, pp. 14011–14017, 1994.
- [131] G. Hummer and A. Szabo, “PNAS Szabo Free energy reconstruction from nonequilibrium single-molecule pulling experiments SMD,” *Proc. Natl. Acad. Sci.*, vol. 98, no. 7, pp. 3658–3661, 2001.
- [132] S. W. Englander, L. Mayne, Y. Bai, and T. R. Sosnick, “Hydrogen exchange: the modern legacy of Linderström-Lang,” *Protein Sci.*, vol. 6, no. 5, pp. 1101–9, 1997.
- [133] J. N. Onuchic and P. G. Wolynes, “Theory of protein folding,” *Current Opinion in Structural Biology*, vol. 14, no. 1, pp. 70–75, 2004.
- [134] S. H. Knauer, I. Artsimovitch, and P. Rosch, “Transformer proteins,” *Cell Cycle*, vol. 11, no. 23, pp. 4289–4290, 2012.
- [135] S. D. Prabhu and G. Salama, “Reactive disulfide compounds induce Ca²⁺ release from cardiac sarcoplasmic reticulum,” *Arch. Biochem. Biophys.*, vol. 282, no. 2, pp. 275–283, 1990.

Anexo

Construcción de mutantes para pinzas ópticas

Como se explicó anteriormente, una vez disociados ambos dominios de RfaH ocurre agregación del dominio N-terminal, lo que dificulta enormemente estudios en masa de esta proteína. Para evitar esto, se planificó una estrategia de experimentos de molécula individual, que permite obtener parámetros cinéticos y termodinámicos en el mismo conjunto de experimentos. Para generar un sistema que permita realizar estos experimentos a través de pinzas ópticas se requiere de las actividades ilustradas en el Esquema 3.



Esquema 3: Esquema general para el desarrollo de quimeras proteína-ADN para experimentos de espectroscopía de fuerza de molécula individual [36].

Para esto, se construyeron distintas variantes de la proteína RfaH (codificada en pIA 777) listadas a continuación:

Tabla 2: Lista de construcciones llevadas a cabo diseñadas para pinzas ópticas. Derecha se muestra si fueron (✓) o no (X) secuenciadas (Sec), si se encontraron en el sobrenadante (S) o pella (P) del test de expresión (Exp) y si fueron o no purificadas.

Nombre	Mutación	Motivo	Sec	Exp	Pur
RfaH-WT	110+ENLYQFG+111..	Sitio de corte para la proteasaTEV en bucle interdominio	✓	S	✓
RfaH-2C	Q2C, R160C	Cisteínas terminales para modificación química	✓	S	✓
RfaH-2GC	Q2GC, R160C	Adición de residuo de glicina para facilitar la modificación	✓	S	X
RfaH-9A	R160C, C9A	Eliminación de primera cisteína interna	X	P	X
RfaH-28A	Q2C, R160C, C28A	Eliminación de segunda cisteína interna	✓	P	X
RfaH-2A	C9A, C28A	Sin cisteínas internas	✓	P	X
CTD-2C	(-NTD), 113C, R 160C	CTD codificado en el plasmidio con cisteínas terminales.	✓	P	X

Los partidores C9A y el R2C generan mutaciones excluyentes, por lo que no fue posible realizar una mutante que carezca de ambas cisteínas. Debido a que RfaH-2C y RfaH-2GC tienen el mismo propósito, una vez purificada con alto rendimiento la primera no fue necesario continuar con la segunda.

Es interesante el caso de la doble mutante sin cisteínas internas RfaH-2A, ya que a pesar de que ambas cisteínas no parecen estar involucradas en la formación de un puente disulfuro, esta mutación es expresada y retenida únicamente en la pella en *E. coli* BL21 y C41. Adicionalmente, la doble mutante C9S/C28S realizada por el laboratorio comparte el mismo comportamiento. Debido a esto, se prosiguió a trabajar con la mutante RfaH-2C y confirmar en los pasos de derivatización si es que la reacción no ocurre con las cisteínas nativas de la proteína RfaH. Por otro lado, la expresión del CTD también presenta problemas durante su ejecución, puesto que sólo fue posible expresarlo en *E. coli* pLysS, ya que de otro modo el crecimiento bacteriano se veía fuertemente afectado, y aun así no se logró expresar en condiciones nativas.

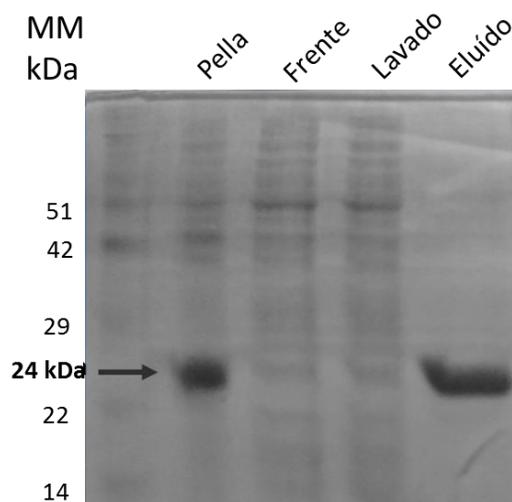


Figura 22: Purificación de RfaH-2C. SDS-PAGE al 12% A/BA. En los carriles se presentan las fracciones de purificación. Se observa que se retiene específicamente una banda de proteínas con masa molecular aproximada de 24 kDa, cercana al peso calculado de RfaH de 22 kDa.

La purificación de RfaH-2C fue realizada mediante cromatografía de afinidad a catión metálico inmovilizado, utilizando el sobrenadante del lisado de *E. coli* BL21 inducida para la expresión heteróloga de RfaH-2C codificada dentro del plasmidio pIA 777. Esta inducción se realiza una vez el cultivo en medio LB alcanzó una turbidez de 10% transmitancia (1,0 unidad de absorbancia) utilizando 0,2 mM IPTG e incubando en

agitación a 25 °C por toda la noche (≈14-16 horas). Estas condiciones de inducción se eligieron luego de un breve test de expresión tomando en cuenta las condiciones de expresión heteróloga de su análogo silvestre. La purificación fue seguida por electroforesis en gel de poliacrilamida desnaturalizante (Figura 23). En este se observa que, a pesar de encontrarse en la pella, es posible recuperar RfaH-2C desde la fracción soluble. El rendimiento de las purificaciones sucesivas con esta proteína oscila entre 5 a 8 mg/L de cultivo.

Para evaluar la pureza y eliminar el exceso de imidazol se utilizó cromatografía de exclusión molecular (Figura 23). En todo paso de purificación se utiliza una concentración de 1 mM de DTT para mantener reducidas las cisteínas expuestas, gracias a esto se observa sólo una banda para RfaH a un tamaño de 18,4 kDa.

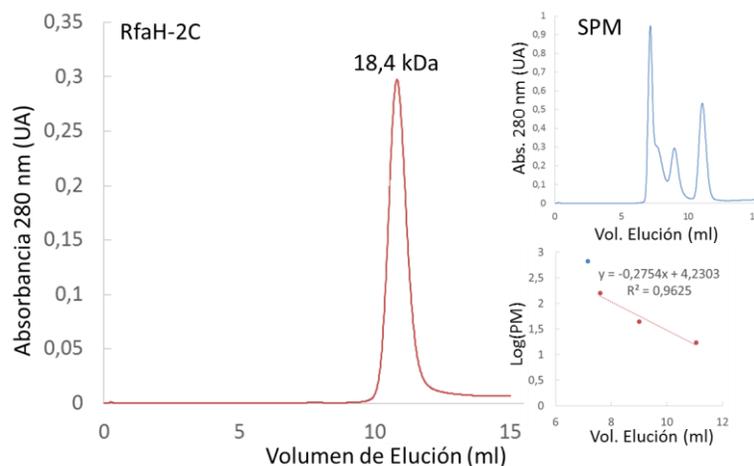


Figura 23: Cromatograma de exclusión molecular de RfaH-2C tras purificación por IMAC. A la izquierda se observa el cromatograma de RfaH tras ser incubada por 2 horas a T° ambiente con 10 mM DTT, a la derecha se observa el cromatograma del estándar de peso molecular correspondiente de menor a mayor volumen de exclusión γ -globulina bovina (158 kDa), ovoalbúmina de pollo (44 kDa), mioglobina de caballo (17 kDa) y vitamina B12 (8,5 kDa). Abajo se muestra la regresión entre el logaritmo de los pesos moleculares y el volumen de elución. En base a tal curva se estima un peso aproximado de 18,4 kDa para RfaH.

Modificación de construcciones para pinzas ópticas

Para evaluar la reactividad de las cisteínas expuestas y descartar la reactividad de las cisteínas internas se utilizó el ensayo de Ellman con la sonda ácido ditionitrobenzoico (DTNB), que se reduce a nitrotiobenzoato en presencia de sulfhidrilos, dando un característico color amarillo. Para esto, 2 μ M de RfaH-2C eluída desde cromatografía de exclusión fueron ensayados con 400 μ M de DTNB en 200 mM de

amortiguador de fosfato a pH 8. La absorbancia a 412 nm de esta mezcla fue medida e interpolada en una curva equivalente con concentraciones crecientes de DTT desde 1 a 10 μM (Figura 25). En esta se observa que RfaH-2C tiene una absorbancia reactividad equivalente a 1,21 y 0,96 veces la del DTT. Y debido a que la reactividad DTT:Cisteína es 1:2, esto sugiere que sólo las cisteínas mutantes están expuestas.

Figura 24 Reactividad de cisteínas en RfaH

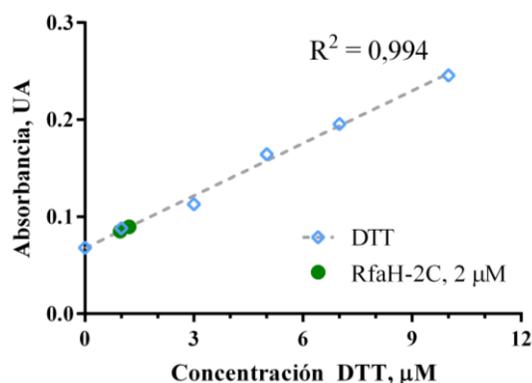


Figura 24: Reactividad de cisteínas evaluadas por ensayo de Ellman. Concentración de proteína de 2 μM tienen una reactividad $\approx 1,1 \mu\text{M}$ DTT.

Al proseguir la marcación con DTDP, se observó que tanto el acetonitrilo como dimetil sulfóxido, solventes comunes para solubilizar esta sonda, inducen la agregación de RfaH (datos no mostrados). Por ello, la solución madre de DTDP fue disuelta en 66% v/v EtOH. Sin embargo, incluso la concentración requerida para llevar a cabo la marcación de 100 μM RfaH-2C y 2,5 mM DTDP ($\approx 0,1$ % v/v EtOH final) inducía la violenta agregación de RfaH. Contra lo esperado, concentraciones de hasta 3% v/v EtOH por sí solos no inducen la agregación de RfaH-2C, y se observó que efectivamente existe una relación directa entre la concentración de DTDP en solución y el porcentaje de agregación (Tabla 3).

La reacción de modificación por DTDP es comúnmente seguida espectrofotométricamente [36], ya que esta molécula reacciona con cisteína y queda unida covalentemente a la proteína, liberando *orto*-tiopiridona que presenta una banda de absorción en agua a 343 nm ausente en DTDP [135]. Debido a que en todas las reacciones se observa turbidez producto de la agregación de RfaH, no fue posible seguir espectrofotométricamente la aparición de *orto*-tiopiridona. Sin embargo, tras 20 minutos de reacción y centrifugación a $23.000G \times 10'$ se midió la transmitancia de la solución observando la aparición de la banda característica a 343 nm ausente en el control sin proteína. Lamentablemente en nuestro caso, este método de análisis espectrofotométrico no permite estimar la estequiometría por los cambios en concentración de proteína en el tiempo.

Tabla 3: Recuperación de proteína en el sobrenadante tras centrifugación de la reacción entre 60 μ M RfaH-2C y la concentración indicada de DTDP.

[DTDP] mM	% Recuperación
1,5	N.S.
1,0	7%
0,8	17%
0,4	43%
0,2	74%
0,1	88%

A pesar de observar la presencia de proteína tras la reacción de marcación, no se obtuvo recuperación tras el paso de intercambio de amortiguador mediante ya sea columnas de exclusión molecular pequeñas (Bio-Spin, BioRad) o columnas de desalinización (HiTrap Desalting Columns, GE-Healthcare Life Sciences), probablemente debido a que corresponde a agregados solubles ya que se observó que la proteína quedaba retenida en el filtro previo a la columna. Por esto, se utilizó cromatografía de exclusión molecular, entregando una concentración de 7 μ M de RfaH-2C. Con esto se realizó espectrometría de masas MALDI-TOF en el CEPEDeq, perteneciente a la Facultad de Ciencias Químicas y Farmacéuticas de la Universidad de Chile (Figura 25). En este espectro se observa que la proteína RfaH-2C es capaz de reaccionar con 2 moléculas de

DTDP, probablemente mediante sus cisteínas terminales debido a que fueron diseñadas para estar altamente expuestas.

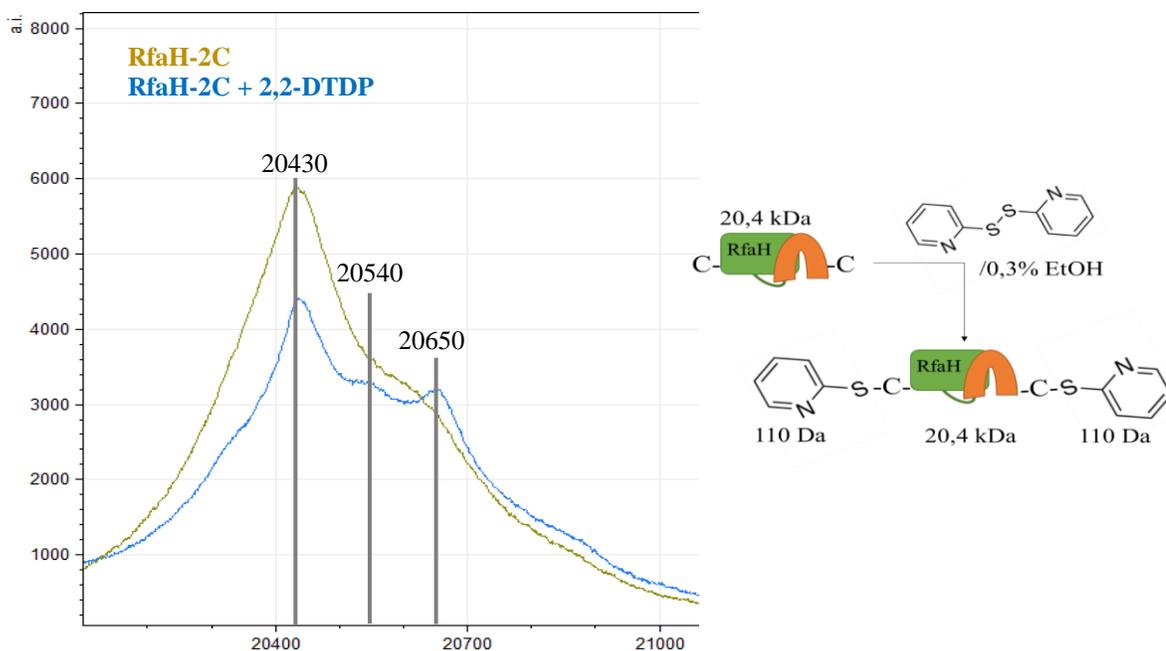


Figura 25: (izq) Espectro de masas de RfaH-2C por MALDI-TOF. Se observa la intensidad (U.A.) de bandas correspondientes al peso calculado por secuencia de RfaH (20,4 kDa) sin marcar (verde) y tras reacción con DTDP y posterior separación por exclusión molecular (azul). En esta banda se observan tres poblaciones con una diferencia de masa de 110 Da entre cada una, que corresponde a la unión de dos moléculas de o-tiopyridona.

Por otro lado, se realizó la síntesis de las mangas de ADN necesarias para dar lugar a la quimera proteína-ADN (Figura 25). Estas fueron precipitadas y concentradas por centrifugación, logrando concentraciones finales de 22 μM para la manga Bio-SH y 27 μM para la Dig-SH. Inspección visual de la electroforesis de estas indica que no hay degradación de las mismas; sin embargo, parecen presentar distintas formas multiméricas, probablemente por asociación inespecífica de las distintas especies lineales. A pesar de estar en presencia de baja concentración de agente reductor se observa una tenue banda de especie dimérica para ambas construcciones, marcada en rojo. La presencia de dímero en condiciones oxidantes sugiere fuertemente que el extremo sulfhidrilo de cada manga está reactivo.

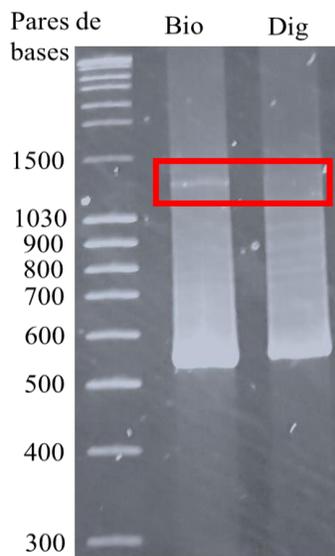
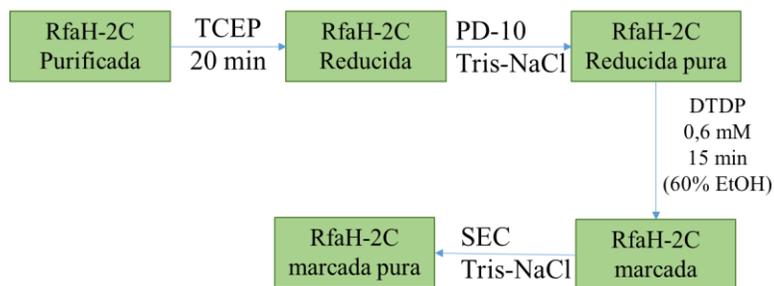


Figura 26: PAGE 6% con tinción RedGel para inspección de mangas de ADN. 1 μ g de ADN por pocillo. Debido a que ambas mangas, modificadas con biotina y digoxigenina tienen en su otro extremo una modificación de sulfhidrilo, pueden reaccionar entre sí y formar dímero, como se observa marcado en cuadro rojo.

Para optimizar el rendimiento de marcación se utilizó un protocolo de marcación y separación de RfaH-2C marcada distinto al propuesto inicialmente. La idea del protocolo es aumentar el volumen total de marcación para mantener una masa soluble de proteínas marcadas, a pesar de la agregación (esquema 4).



Esquema 4: Protocolo de reducción y marcación de cisteínas expuestas junto con separación de la especie monomérica.

El volumen inicial utilizado para este protocolo corresponde a RfaH-2C 2,5 ml a concentraciones entre 60-90 μ M. Este fue reducido con TCEP que es un reductor más estable a T° ambiente y con una cinética de reducción mucho más rápida que DTT. El exceso de reductor fue removido utilizando columna de exclusión PD10, permitiendo intercambiar la totalidad de volumen, recuperando \approx 80-90% de la concentración de

proteína inicial. La reacción de marcación se prosiguió inmediatamente con 500 μl de solución de proteína y DTDP 0,6 mM disuelto en 6% v/v etanol por 15 minutos en oscuridad, dando una relación de proteína:DTDP cercana a 1:15. La relación recomendada corresponde a 1:25, sin embargo eso resulta en concentraciones inestabilizantes de DTDP.

Luego de la marcación, se centrifugó a 25.000g x 15 minutos. La concentración final en este paso fue la limitante del protocolo, pues la recuperación de proteína oscila entre 5 y 10 μM . Intentar concentrar la proteína por filtración por centrifugación termina en agregación de la totalidad de proteína, por lo que remoción rápida del DTDP restante es clave para asegurar rendimiento positivo. En el paso de filtración en gel, a pesar de que en cada ocasión se observó la presencia de proteína y DTDP en el cromatograma, sólo una vez fue posible recolectar una fracción con concentraciones detectables ya sea por medición de absorbancia a 280 nm (ϵ_{280} 13.370 M^{-1}) o ensayo de Bradford, con una concentración de 7 μM y volumen de 100 μl .