

UNIVERSIDAD DE GRANADA

FACULTAD DE CIENCIAS



Universidad de Granada

DEPARTAMENTO DE QUÍMICA-FÍSICA

“ENERGÉTICA, CINÉTICA Y EVOLUCIÓN DE  
PROTEINAS: IMPLICACIONES EN DISEÑO  
MOLECULAR”

David Rodríguez Larrea

TESIS DOCTORAL

GRANADA

2008

Editor: Editorial de la Universidad de Granada  
Autor: David Rodríguez Larrea  
D.L.: GR. 2094-2008  
ISBN: 978-84-691-6325-2

UNIVERSIDAD DE GRANADA

FACULTAD DE CIENCIAS

DEPARTAMENTO DE QUÍMICA-FÍSICA

“ENERGÉTICA, CINÉTICA Y EVOLUCIÓN DE  
PROTEINAS: IMPLICACIONES EN DISEÑO  
MOLECULAR”

David Rodríguez Larrea

TESIS DOCTORAL

GRANADA

2008

UNIVERSIDAD DE GRANADA  
FACULTAD DE CIENCIAS  
DEPARTAMENTO DE QUÍMICA-FÍSICA

Memoria presentada para optar al Grado de Doctor  
en Ciencias

Fdo.: David Rodríguez Larrea  
Licenciado en Biología y Licenciado en Bioquímica  
por la Universidad de Salamanca

Granada a 31 de Julio de 2008

Directores de la Tesis:

Fdo.: José Manuel Sánchez Ruiz  
Catedrático de Química-Física  
Departamento de Química-Física  
Facultad de Ciencias  
Universidad de Granada

Fdo.: Beatriz Ibarra Molero  
Profesora Titular  
Departamento de Química-Física  
Facultad de Ciencias  
Universidad de Granada

## DEDICATORIA

Puedo escribir los versos más tristes esta noche...

Escribir, por ejemplo: "La noche está estrellada,  
Y tiritan, azules los astros, a lo lejos"

El viento de la noche gira en el cielo y canta.

Puedo escribir los versos más tristes esta noche.  
Yo la quise, y a veces ella también me quiso.

En las noches como ésta la tuve entre mis brazos.  
La besé tantas veces bajo el cielo infinito.

Ella me quiso, a veces yo también la quería.  
Cómo no haber amado sus grandes ojos infinitos.

Puedo escribir los versos más tristes esta noche.  
Pensar que no la tengo. Sentir que la he perdido.

Oír la noche inmensa, más inmensa sin ella.  
Y el verso cae al alma como al pasto el rocío...

...

Porque en noches como ésta la tuve entre mis brazos,  
Mi alma no se contenta con haberla perdido.

Aunque éste sea el último dolor que ella me causa,  
Y éstos los últimos versos que yo le escribo.

P.Neruda

## AGRADECIMIENTOS

Quiero agradecer sinceramente el apoyo y la enseñanza de mis directores de tesis, J.M. Sánchez-Ruiz y Beatriz Ibarra-Molero. Obviamente, sin ellos esto no hubiera salido adelante... o no lo hubiera hecho tan bien.

Quiero agradecer a los amigos que han estado apoyandome/soportándome estos casi cuatro años.

Quiero agradecer a mi familia el apoyo y el cariño.

Quiero agradecer las tapas de Granada y las cañitas de después del curro. En ocasiones fueron el motivo para acudir un día más al trabajo.

## INDICE GENERAL

	PAG
<b><u>1.-INTRODUCCIÓN</u></b>	1
1.1- CONTEXTO Y GENERALIDADES	2
1.2- ESTRUCTURA-ENERGÉTICA DE PROTEÍNAS	6
1.2.1-Factores que determinan la estabilidad	6
1.2.2-Diagrama de reacción: Estabilidad cinética y estabilidad termodinámica	10
1.2.3"Ensembles" y "Protein Landscape"	13
1.3.- EVOLUCIÓN DE PROTEINAS	15
1.3.1 Sequence space, límites selectivos y la hipótesis de Boltzmann	18
1.3.2 Algunos factores que determinan la "evolutivilidad" de una proteína	22
1.3.3 Estadística y alineamiento de secuencias	24
1.3.4 Relación de la evolución de proteínas con el presente proyecto de Tesis	27
<b><u>2.-ARTÍCULOS PUBLICADOS</u></b>	29
2.1-ARTÍCULO PRIMERO	30
2.1.1 Introducción artículo primero	31
2.1.2 "Role of solvation barriers in protein kinetic stability"	36
2.2- ARTÍCULO SEGUNDO	46
2.2.1 Introducción artículo Segundo	47
2.2.2 "Energetic and structural consequences of desolvation/solvation barriers to protein folding/unfolding assessed from experimental unfolding	

rates"	50
2.3- ARTICULO TERCERO	53
2.3.1 Introducción artículo tercero	54
2.3.2 "Beyond Lumry-Eyring: An unexpected pattern of operational reversibility/irreversibility in protein denaturation"	57
<b><u>3.- RESUMEN Y CONCLUSIONES</u></b>	63
<b><u>4.- APENDICE 1</u></b>	66
4.1 INTRODUCCIÓN	67
4.2 "SUPER-PROTEINS FROM FITNESS THRESHOLD SELECTION STATISTICS"	68
<b><u>5.- APENDICE 2</u></b>	86
5.1 CALORIMETRÍA DIFERENCIAL DE BARRIDO (DSC)	87
5.1.1 Características técnicas	87
5.1.2 DSC de proteínas	90
5.1.3 Modelos para el análisis de datos de DSC de proteínas	93
5.1.3.1 Modelo de dos estados reversible	95
5.1.3.2 Desnaturalización multiestado según el formalismo de la función de partición	98
5.1.3.3 Modelo de dos estados irreversible	99
5.2 ESTUDIO DEL DESPLEGAMIENTO/REPLEGAMIENTO DE PROTEÍNAS POR EMISIÓN DE FLUORESCENCIA	106
5.2.1 Generalidades de la técnica y su uso en proteínas	106

5.2.2 Energética del equilibrio en plegamiento/desplegamiento de proteínas seguido por fluorescencia	109
5.2.3 Cinéticas de desplegado/replegado seguidas por fluorescencia	112
5.2.4 Ensayos para la búsqueda de intermedios de equilibrio y cinéticos	118
5.2.4.1 El experimento de doble salto para la detección de intermedios de equilibrio	119
5.2.4.2 El experimento de doble salto cinético	121
<b><u>6.-APENDICE 3: ENSAYOS DE ACTIVIDAD DE TIORREDOXINA Y LIPASA</u></b>	124
6.1 ENSAYO DE ACTIVIDAD DE TRX-INSULINA	125
6.1.1 Generalidades	125
6.1.2 Metodología experimental	126
6.2 ENSAYO DE ACTIVIDAD DE TRX-TRX REDUCTASA	129
6.2.1 Generalidades	129
6.2.2 Metodología experimental	130
6.3 ENSAYO DE ACTIVIDAD DE LIPASA	133
6.3.1 Generalidades	133
6.3.2 Metodología experimental	134
<b><u>7.-APENDICE 4: PURIFICACIONES DE PROTEÍNAS</u></b>	137
7.1 PURIFICACIÓN DE TIORREDOXINA DE E.COLI	138
7.2 PURIFICACIÓN DE BIBLIOTECA DE MUTANTES DE TRX-HISTAG	143
7.3 PURIFICACIÓN DE $\gamma$ -14-3-3	149
<b><u>8.-BIBLIOGRAFÍA</u></b>	155

## INDICE FIGURAS

Figura 1	pag. 11
Figura 2	pag. 14
Figura 3	pag. 19
Figura AP1.1	pag. 83
Figura AP1.2	pag. 84
Figura AP1.3	pag. 85
Figura DSC1	pag. 89
Figura DSC2	pag. 92
Figura DSC3	pag. 95
Figura DSC4	pag. 105
Figura FLU1	pag. 108
Figura FLU2	pag. 112
Figura FLU3	pag. 114
Figura FLU4	pag. 117
Figura FLU5	pag. 120
Figura FLU6	pag. 122
Figura INS1	pag. 125
Figura INS2	pag. 126
Figura INS3	pag. 129
Figura RED1	pag. 130
Figura RED2	pag. 133
Figura LIP1	pag. 134
Figura LIP2	pag. 136
Figura PUR1	pag. 139
Figura PUR2	pag. 145
Figura PUR3	pag. 151

## ABREVIATURAS Y SÍMBOLOS

DSC	Calorimetría Diferencial de Barrido
$\Delta C_p$	Diferencia de Capacidad Calorífica
$\Delta G$	Diferencia de Energía Libre
$\Delta H$	Diferencia de Entalpia
$\Delta S$	Diferencia de Entropia
T	Temperatura (grados Kelvin)
T <sub>m</sub>	Temperatura de desnaturalización
TRX	Tiorredoxina
WT	Forma Silvestre
IPTG	Isopropil 1-tio-beta-D-galactopiranosido
EDTA	Etildimetiltetraacético
DTT	Ditrioteitol
DTNB	5-5'-dinitrobenzoico
K	Constante de equilibrio
k	Constante cinética aparente
k <sub>u</sub>	Constante cinética de desplegamiento
k <sub>f</sub>	constante cinética de plegamiento
C	Concentración de desnaturalizante
C <sub>1/2</sub>	Concentración de desnaturalizante para $\Delta G=0$
m <sub>1/2</sub>	Efecto de desnaturalizante en $\Delta G$
$\mu$	Efecto de desnaturalizante en $\Delta G$ de desplegamiento
m <sub>f</sub>	Efecto de desnaturalizante en $\Delta G$ de plegamiento
N	Estado nativo
U O D	Estado desplegado o desnaturalizado

# **1 . - INTRODUCCIÓN**

## **1.1- CONTEXTO Y GENERALIDADES**

Discutía Bertrand Russell en su libro "El método científico" que cualquier trabajo científico tiene por objeto resolver una serie de preguntas de carácter más o menos general. Las preguntas que nos realizamos van siendo cada día más y más específicas y en consecuencia el objeto inherente a la ciencia, la generalidad, se va difuminando. Conviene, por tanto, realizar un esfuerzo preliminar que nos permita situar nuestro trabajo y facilite la comprensión del mismo. A la par, este esfuerzo permitirá vislumbrar el alcance de los conocimientos generados.

Nuestro objetivo es comprender cómo y por qué funcionan las proteínas a fin de ser capaces de diseñarlas. Ser capaces de diseñar proteínas supone determinar no sólo la función que llevará a cabo la proteína (unión a ligandos, catalizadora,...), sino también sus propiedades biofísicas (estabilidad termodinámica, cinética,...). Hoy en día entendemos diseño de proteínas como la obtención de proteínas modificadas con características mejoradas, proteínas con nuevas actividades o con nuevas estructuras. Trabajar en diseño de proteínas, entendido de esta manera, tiene mucho interés por dos razones. En primer lugar por las potenciales aplicaciones de proteínas tanto en biomedicina/terapéutica como tecnológicas/industriales y en segundo lugar por intereses puramente académicos relacionados con la comprensión de principios fundamentales.

El objetivo no es trivial y mucho menos sencillo de conseguir. Las proteínas se implican en la práctica totalidad de procesos biológicos y ello supone que hay que "pasar por ellas" para comprender cualquier faceta de la vida que queramos estudiar y también suelen ser siempre la diana de fármacos contra cualquier enfermedad. Desde un punto de vista tecnológico las podemos considerar como las mejores y más eficientes nanomáquinas. También las podemos considerar como los más eficientes catalizadores químicos conocidos. Por su papel fisiológico, son utilizadas como medicamentos, y si bien en 2006 había sólo 16 anticuerpos monoclonales en el mercado para uso terapéutico, hay muchos más en lista de espera para ser testados (Vitetta ES 2006). Desde el punto de vista industrial su uso ha tenido mayor difusión, no sólo por su extremadamente alto poder catalítico, sino también por su mayor eficiencia (ausencia de productos no deseados, menor gasto energético,...) y por no ser contaminantes. Entre estas, una de las de

mayor uso son las lipasas, con aplicaciones como sustitutas de detergentes químicos (mucho más contaminantes), y por su uso en la industria papelera, maderera,... Y de momento sólo hablamos de aquellas labores que la vida ha considerado como importantes, su potencial real es posiblemente mayor ((Rothlisberger, Khersonsky et al. 2008)). Resulta por tanto difícil avanzar las implicaciones inherentes al completo conocimiento de estas moléculas.

Si bien no está claro si la naturaleza seleccionó el uso de las proteínas o todo fue un accidente, puesto que poco sabemos del origen de la vida en su formato actual, su versatilidad, su eficiencia y su sencillez han hecho que la vida se base en ellas (hasta la fecha en todos los organismos conocidos). Actúan formando parte de estructuras de la célula o el exterior celular, como membranas, citoesqueleto y matriz extracelular. Llevan a cabo procesos mecánicos, como el movimiento flagelar o la contracción muscular. Se encargan de la obtención de energía, bien sea a través de la fotosíntesis (complejos I y II) como por degradación de moléculas (cadenas respiratorias, ciclo de Krebs, enzimas de la glicolisis...). Posteriormente son las responsables de la síntesis de los compuestos celulares (lípidos, glúcidos, nucleótidos, aminoácidos,...) y de ensamblarlos correctamente (replicación, transcripción, traducción, síntesis de membranas,...). Se encuentran implicadas en la transmisión de señales y ejecución de respuestas a las mismas. En definitiva, se implican en todas las facetas de la vida. Son, digamos, la información genética en movimiento.

Es una buena aproximación, al menos a nivel conceptual, decir que cada función requiere de una o varias proteínas distintas. Esto supone que las proteínas deben manifestar diferencias significativas unas con respecto a otras y sin embargo toda esta diversidad debe ser accesible. Dicho en otras palabras, el número de proteínas posibles debe ser enorme, casi infinito diríamos, y, al mismo tiempo, debe ser fácil para un organismo trabajar con toda esa diversidad: Si cada proteína requiriese de un mecanismo especial para ser sintetizada, posiblemente la vida no existiría en su formato actual.

Este problema, que enfrenta diversidad y sencillez, se solventa gracias a que las proteínas son heteropolímeros. Las proteínas se componen de pequeñas unidades, denominadas aminoácidos de los cuales existen 20 tipos diferentes. Combinando estos 20 aminoácidos de

distintos modos podemos obtener proteínas diferentes. Por tanto, cada proteína es una secuencia ordenada de aminoácidos distintos. Cabe puntualizar que dos proteínas con los mismos aminoácidos pero en orden distinto dan lugar a dos proteínas diferentes. De este modo la vida dispone de la diversidad, de una diversidad cuyos límites parece que aún no han sido acotados (Dokholyan, Shakhnovich et al. 2002).

La información que determina la secuencia de cada una de las proteínas que una célula puede llegar a necesitar está codificada en los ácidos nucleicos de un modo elegante. El descubrimiento de dicho código, que implicó entre otros al científico americano-español Severo Ochoa, llevó a la biología a una auténtica revolución conceptual que desembocó en el denominado Dogma de la Biología Molecular: Un gen codifica una proteína con el uso del código genético universal. Obviamente el campo de los conceptos no responde a la organización de la naturaleza, sino a la organización del conocimiento y este dogma está hoy en día matizado en todos sus aspectos.

Hoy en día las técnicas de secuenciación del ADN (ácido desoxirribonucleico) han permitido disponer de inmensas bases de datos que, con el uso del Código Genético, dan lugar a la secuencia de las proteínas que usa la vida.

Otras bases de datos, como aquéllas que almacenan estructuras de proteínas, poseen también un crecimiento exponencial y tienen un inmenso valor en el trabajo con proteínas. Por tanto, dado que nuestro objetivo final es llegar a ser capaces de diseñar proteínas, enfocamos el presente trabajo haciéndonos dos preguntas: ¿Qué factores contribuyen a la estabilidad cinética en la desnaturalización irreversible de proteínas? El concepto de estabilidad cinética tiene una relevancia fundamental en el uso de proteínas a nivel biotecnológico, pues en definitiva viene a marcar el tiempo de vida útil de nuestra proteína. Y por otro lado, nos preguntamos cómo hacer uso en diseño de proteínas la información evolutiva contenida en las bases de datos de secuencias con objeto de obtener proteínas modificadas con propiedades especiales. De hecho, trabajos previos en nuestro grupo de investigación han puesto de manifiesto que existe una correlación robusta entre frecuencia de aparición de residuos en alineamientos de secuencias y estabilidad.

La presente tesis doctoral se centra en estas dos cuestiones. En primer lugar mostramos tres trabajos recientemente publicados en revistas internacionales de prestigio reconocido:

"Role of solvation barriers in protein kinetic stability"

Rodríguez-Larrea D, Minning S, Borchert TV, Sanchez-Ruiz JM.

*Journal of Molecular Biology.* 2006 Jul 14;360(3):715-24.

"Energetic and structural consequences of desolvation/solvation barriers to protein folding/unfolding assessed from experimental unfolding rates"

Rodríguez-Larrea D, Ibarra-Molero B, Sanchez-Ruiz JM.

*Biophysical Journal.* 2006 Sep 1;91(5):L48-50.

"Beyond Lumry-Eyring: an unexpected pattern of operational reversibility/irreversibility in protein denaturation"

Rodríguez-Larrea D, Ibarra-Molero B, de Maria L, Borchert TV, Sanchez-Ruiz JM.

*Proteins.* 2008 Jan 1;70(1):19-24.

En ellos se abordan cuestiones que se podrían englobar en la primera de las cuestiones antes mencionadas, véase, cuales son los factores que determinan la estabilidad cinética de las proteínas. Ello implica la caracterización de los estados de transición cinéticamente relevantes en el proceso de desnaturalización irreversible (Rodríguez-Larrea, Minning et al. 2006)(Rodríguez-Larrea, Ibarra-Molero et al. 2006). Así mismo, realizamos una extensión de los modelos de Lumry-Eyring (Sanchez-Ruiz 1992), modelos tradicionalmente propuestos para entender los fenómenos de desnaturalización reversible e irreversible de proteínas (Rodríguez-Larrea, Ibarra-Molero et al. 2008).

Finalmente, y expuesto en el apéndice 1 por razones de forma inherentes al modelo de tesis presentado, proponemos un trabajo que aborda la segunda de las cuestiones, véase, desarrollar una metodología que permita mejorar propiedades de las proteínas de interés basada en la información evolutiva contenida en los alineamientos de secuencias. Este trabajo supuso gran parte del periodo de formación e investigación y por tanto es oportuno incluirlo. En él tratamos de desarrollar una metodología que permita la obtención de proteínas con distintas propiedades simultáneamente mejoradas (y en gran medida), para ello nos basamos en teorías de

evolución biológica y las aplicamos a los alineamientos de secuencias (haciendo uso de las extensas bases de datos). Dicho trabajo, en el momento del depósito de la tesis, se encuentra remitido a una revista de reconocido prestigio y está siendo evaluado por "referees". Sin embargo, hasta la aceptación del mismo no puede ser incluido dentro del cuerpo formal de la tesis doctoral.

A continuación expondremos dos apartados introductorios para englobar los trabajos presentados.

## **1.2- ESTRUCTURA-ENERGÉTICA DE PROTEÍNAS**

El grupo en el que se ha desarrollado el presente trabajo está interesado en el diseño de proteínas. Presenta una gran trayectoria en el diseño de la estabilidad y actualmente posee varias líneas de gran interés en nuevos campos (diseño de actividades de novo, mejora de capacidad catalítica,...). Con lo que respecta al campo de la estabilidad, su importancia radica en los potenciales usos de las proteínas, tanto da si nuestro campo es la industria o la biomedicina. Es por tanto que merece la pena introducir una serie de conceptos sobre energética. Sin embargo, esta no se puede desligar de la estructura y la dinámica de la misma. Por tanto, y en la medida de lo posible, trataremos de dar una explicación estructural a cuantos fenómenos energéticos describamos.

En primer lugar, cabe distinguir entre las proteínas globulares y las fibrilares. Las primeras son las más comúnmente estudiadas porque suelen ser las que revisten mayor interés biomédico y biotecnológico. Por otro lado cabe distinguir, dentro de las proteínas globulares, entre aquéllas que son de membrana y aquéllas solubles en agua.

### **1.2.1-FACTORES QUE DETERMINAN LA ESTABILIDAD**

La cadena de aminoácidos que compone una proteína se encuentra, en su estado fisiológico o nativo, adquiriendo una conformación determinada en el espacio (salvo excepciones como las Natively Unfolded Proteins, (Uversky 2002)). Esta estructura es la que permite a la proteína llevar a cabo las tareas que le son propias y, junto a las interacciones que mantenga con el solvente y distintos ligandos, la que explica también la estabilidad de dicha conformación.

Entendemos por estabilidad una medida de la tendencia de la cadena polipeptídica a mantenerse en una conformación tridimensional determinada. No obstante, según el campo en el que nos movamos observaremos que el uso del concepto de estabilidad se acomoda a las necesidades. Así quizás en la industria farmacéutica lo entenderán como "shelf life", que es el tiempo útil del medicamento y quizás un biólogo molecular argumente que es el tiempo de vida media en el organismo. Nosotros trabajaremos con el concepto descrito al inicio del párrafo, haciendo una clara diferencia: la estabilidad puede ser termodinámica o cinética, cuya diferencia se discutirá brevemente a lo largo del texto.

Pese al vasto conocimiento acumulado, hoy en día resulta extremadamente difícil determinar si una secuencia de aminoácidos adquirirá una conformación tridimensional (Uversky, Gillespie et al. 2000). Más difícil aún resulta determinar cuál será la estructura concreta que adoptará la cadena de aminoácidos en el espacio, problema normalmente considerado como problema del "folding" o "Segundo Código Genético" (Qian, Raman et al. 2007). Y casi imposible determinar cuál será la estabilidad de la estructura tridimensional que adopte.

Gracias al gran desarrollo de las técnicas de resolución de estructura de proteínas, como cristalografía de rayos X, resonancia magnética nuclear y cada vez más la microscopía electrónica, junto con el avance de la dinámica molecular (y su uso en predicción de estructura), suele ser frecuente que nos enfrentemos al problema de la estabilidad a partir de una estructura dada.

Sin embargo, para afrontar el diseño de la estabilidad en proteínas es necesario conocer primero cuáles son las fuerzas que determinan que la cadena aminoacídica adquiera una conformación tridimensional.

Desde el punto de vista termodinámico, la cadena polipeptídica se pliega cuando la energía libre de la cadena plegada es menor a la de la cadena desplegada. La energía libre posee dos componentes, el entálpico y el entrópico. En el estado plegado, las interacciones entre residuos están maximizadas con respecto al estado desplegado, mientras que cuando la cadena de aminoácidos no está plegada presenta una mayor entropía, esto es, tiene un mayor número de grados de

libertad (Dill 1990), (Sanchez-Ruiz 1995). La estabilidad termodinámica, que es la diferencia de energía libre entre los estados desplegado y plegado, surge de un fino balance de los términos entálpicos y entrópicos.

De entre todas las contribuciones a estos términos, se ha considerado que el efecto hidrofóbico (a veces también descrito como colapso hidrofóbico) es uno de los principales directores del plegamiento de la cadena polipeptídica. De los 20 aminoácidos que suelen componer las proteínas, una buena fracción de ellos son de naturaleza hidrofóbica. Por tanto, la cadena polipeptídica tiende a adquirir una forma globular que localice en el núcleo, alejados del solvente, a los residuos hidrofóbicos, y mantenga en la superficie, en contacto con las moléculas de agua (normalmente el solvente en el que se encuentran), a los residuos hidrofílicos. Sin embargo, esta norma que parece sencilla presenta dos características que nos devuelven a la naturaleza del problema (Dill 1990):

-Por un lado un exceso de residuos hidrofóbicos produce un colapso inespecífico. Es decir, la cadena polipeptídica adquiere una forma globular no definida (dos mismas cadenas pueden adoptar configuraciones diferentes). En tal caso, la proteína no puede ejercer función alguna.

-Por otro lado el orden de los residuos debe permitir la formación de estructuras secundarias, dado que estas son disposiciones regulares en las que se maximizan las interacciones entre residuos. Hay que tener en cuenta que cualquier secuencia no tiene por qué dar lugar a estructuras secundarias.

Es una posible hipótesis que sea el balance entre el colapso hidrofóbico y la formación de estructura secundaria el que haga que las proteínas adquieran una estructura unívoca (Onuchic, Luthey-Schulten et al. 1997) (Plotkin and Onuchic 2002).

Una vez vista la importancia del colapso hidrofóbico y las estructuras secundarias, cabe preguntarse el tipo de interacciones que mantienen entre sí los residuos para llegar a estabilizar el estado plegado. Esencialmente todas las interacciones son de naturaleza electrostática, si bien por claras diferencias en su naturaleza cabe clasificarlas en (Nölting 2006) (Dill 1990):

- Interacciones de Van der Waals
- Puentes de Hidrógeno
- Interacciones carga-carga

Una vez plegada la cadena polipeptídica, los residuos mantienen entre sí numerosas interacciones de Van der Waals. Su importancia en el plegamiento está fuera de lugar a dudas y es bien sabido que una mutación disruptiva en el núcleo hidrofóbico tiene efectos dramáticos en la estabilidad de las proteínas. Sin embargo, resulta curioso que estas interacciones no se encuentren maximizadas (Liang and Dill 2001). La razón por la cual las proteínas encontradas en los seres vivos no forman un "puzzle" perfecto no está perfectamente comprendida, pero se ha argumentado que la necesidad de una cierta flexibilidad en la estructura tridimensional es necesaria para llevar a cabo la función. Cabe también proponer, como se justificará a lo largo de la introducción, que un núcleo hidrofóbico flexible permite "acomodar" mejor posibles eventos mutacionales que tienen lugar a lo largo de la evolución.

La importancia de los puentes de hidrógeno en la estabilidad de las proteínas es un tanto controvertida (Dill 1990) (Nölting 2006). La razón es que cualquier puente de hidrógeno entre residuos de la proteína puede ser reemplazado en el estado desnaturalizado por un puente de hidrógeno con el solvente (normalmente agua). Se ha argumentado que su papel si debe ser especialmente relevante en proteínas de membrana, cuyo solvente sería de naturaleza hidrofóbica. Sin embargo, un estudio reciente que implica ciclos de doble mutante, que permiten deducir el valor de una interacción entre residuos gracias a la formación de un ciclo termodinámico, parecen rebajar también la importancia de este tipo de interacción entre residuos incluso para ambientes hidrofóbicos (Joh, Min et al. 2008). Por otro lado, se ha argumentado que los puentes de hidrógeno pueden tener un papel relevante en el origen de la cooperatividad en el plegamiento/desplegamiento de proteínas (Makhatadze 2005).

Finalmente, las interacciones carga-carga han sido frecuentemente infravaloradas por su baja energía de interacción. Sin embargo, y dado que la estabilidad de las proteínas es marginal, su importancia puede ser más relevante de lo pensado (Makhatadze, Loladze et al. 2003) (Loladze, Ibarra-Molero et al. 1999). Por otro lado,

presentan una ventaja sobre los anteriores tipos de interacción: su modelización es más robusta en el sentido de que pequeñas imprecisiones en el método de cálculo no tienen consecuencias dramáticas (como podría suceder por ejemplo al modelizar interacciones de Van der Waals) (Ibarra-Molero and Sanchez-Ruiz 2002).

Hemos dejado de lado otro tipo de fuerzas como la formación de puentes disulfuro, que son enlaces de tipo covalente y pese a tener un papel muy importante en la estabilidad de las proteínas que los presentan, no son generales a todas las proteínas.

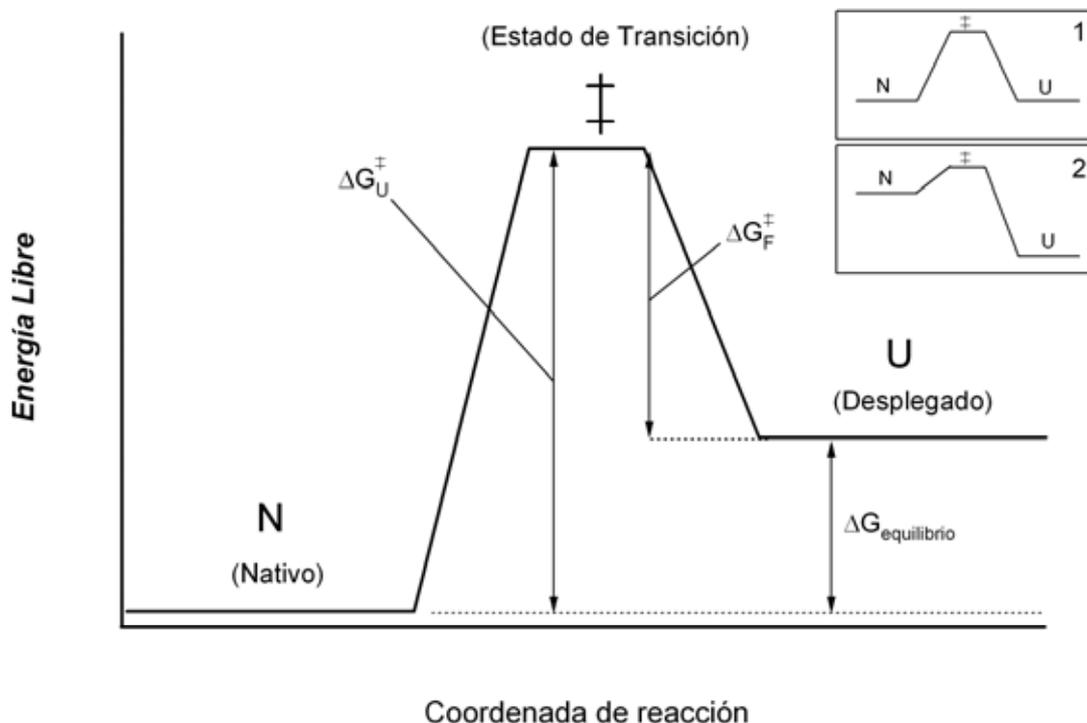
### **1.2.2-DIAGRAMA DE REACCIÓN: ESTABILIDAD CINÉTICA Y ESTABILIDAD TERMODINÁMICA.**

Una de las maneras más ilustrativas de explicar los estados implicados en el plegamiento/desplegamiento de proteínas y su estabilidad es a través de un diagrama de reacción (Figura 1). En él se representa la energía libre en función de alguna medida del avance del proceso (grado de desplegamiento). El caso tipo es aquél que implica a dos macroestados, el nativo y el desplegado, separados por una barrera de energía libre que viene a ser el estado de transición. Por supuesto, este es sólo un caso tipo y el diagrama puede implicar muchos más estados.

La diferencia de energías entre los macroestados (los estados nativo y desplegado en nuestro caso) determina la estabilidad termodinámica. Sin embargo, para extraer dicha información es preciso que ambos estados se encuentren en equilibrio. Ello implica que las poblaciones varían como función de la constante de equilibrio, que, en último término, depende de la diferencia de energía entre los estados implicados.

Sin embargo, como se observa en la Figura 1, para que esto ocurra la proteína tiene que atravesar un estado de alta energía, el estado de transición. Si bien para un químico puede ser fácilmente asumible y lógico la presencia de un estado de máxima energía, para el caso de proteínas no resulta tan fácilmente aceptable. Si recordamos lo mencionado en el apartado anterior, el principal componente de la proteína que dirige el plegamiento son las interacciones en el estado nativo, mientras que la ganancia entrópica del estado desnaturalizado se opone a él. Algunos estudios sugieren que las barreras de energía

libre se originan por una asincronía en la compensación entálpico-entrópica, de modo que existe un momento donde las interacciones entre residuos se han roto y la cadena polipeptídica aún no ha ganado en entropía conformacional (Munoz 2007).



**Figura 1:** Diagrama de reacción donde se representan los estados (caracterizados por una coordenada de reacción) como función de la energía. La figura central representa el perfil de energía cuando la proteína se encuentra en condiciones nativas. El panel inserto 1 representa el perfil de energía cuando los estados nativo y desplegado son isoenergéticos. El panel inserto 2 representa el perfil cuando las condiciones son desnaturalizantes.

Por otro lado y como se mencionó en el apartado anterior, los residuos forman interacciones con el solvente cuando la proteína está desplegada y entre sí cuando se encuentra plegada. Algunos estudios teóricos recientes apuntan a que en el estado de transición existe una asincronía en esta compensación (que sería entálpico-entálpica) y que ello da lugar a uno de los componentes de la barrera de energía libre (Liu and Chan 2005). Dado que bajo numerosas condiciones de aplicación de proteínas la estabilidad cinética es la relevante, iniciamos una caracterización de los estados de transición relevantes cinéticamente en la desnaturalización térmica de la lipasa de *Thermomyces*

lanuginosa. Tras un exhaustivo estudio del efecto de la urea en la estabilidad cinética, concluimos que existía una discrepancia en la información suministrada por el efecto de la urea y la energía de activación obtenida experimentalmente (como se verá más adelante). Ello implica que el balance energético en estados de transición es diferente que en estados nativos. A este componente lo denominamos barrera de solvatación y consiste en que después de la rotura de una interacción hay un momento en el que la molécula de agua aún no puede entrar debido a su tamaño, por lo que la situación es aquella donde se han roto interacciones y aún no se han formado otras nuevas con el solvente. En los trabajos denominados "Artículo Primero" (Rodríguez-Larrea, Minning et al. 2006) y "Segundo" (Rodríguez-Larrea, Ibarra-Molero et al. 2006) aportamos la primera evidencia experimental de dicho componente y, haciendo uso de ciertas correlaciones existentes entre entalpía y áreas de interacción, proponemos una imagen estructural del estado de transición.

La existencia de barreras de energía libre al plegamiento/desplegamiento, hace que también tengamos que hablar de la estabilidad cinética de una proteína. La estabilidad cinética es una medida de la escala temporal que implica la desnaturalización de una fracción de nuestra proteína. Su importancia radica en que en numerosos sistemas de proteínas sólo tiene sentido hablar de estabilidad cinética, puesto que la fracción nativa va desapareciendo con el tiempo, bien sea porque no son estables termodinámicamente o bien porque ocurran fenómenos en el estado desnaturalizado que provoquen irreversibilidad en el proceso de desnaturalización. Por otro lado, numerosas aproximaciones a la inhibición de la amiloidogénesis se basan en una estabilización cinética del estado nativo (Hammarstrom, Wiseman et al. 2003). La magnitud de la barrera de energía libre será la que determine la velocidad a la que ocurre la reacción de plegamiento/deplegamiento y, en sistemas irreversibles, la que determine la velocidad a la cual perdemos fracción nativa.

En el caso de que no exista reversibilidad en el proceso de desplegado (o bien que ésta sólo exista en las condiciones de tubo de ensayo) la estabilidad cinética de la proteína marca el tiempo "útil" de la misma (Jaswal, Sohl et al. 2002) (Jaswal, Truhlar et al. 2005). Por ello, la determinación de la estabilidad cinética de una proteína, así como de los factores que contribuyen a ella, reviste una especial importancia. Es en este sentido donde se han centrado gran

parte de nuestros esfuerzos, apoyándonos para ello en el estudio de la lipasa de *Thermomyces lanuginosa*, una proteína de enorme interés industrial. Como más adelante se señalará, los estudios a través de calorimetría diferencial de barrido en presencia de cosolventes nos han permitido realizar una demostración experimental de la relevancia del agua en el estado de transición del proceso de desplegamiento.

### **1.2.3 "ENSEMBLES" Y "PROTEIN LANDSCAPE"**

Durante los apartados anteriores hemos trabajado asumiendo una estructura nativa constante y definida. Si bien esta imagen es válida para expresar claramente ciertos conceptos, la realidad dista de esa visión. Las proteínas mantienen su estructura gracias a una miríada de interacciones débiles que, debido a la energía térmica del sistema, están sometidas a constantes fluctuaciones (Cooper 1976).

Ello hace que el estado nativo no sea una estructura perfectamente definida, sino un conjunto de conformaciones que fluctúan normalmente en torno a un motivo común. Esta visión del estado nativo es especialmente relevante para comprender determinados procesos catalíticos y de reconocimiento y unión a ligandos (Frederick, Marlow et al. 2007).

En esta línea, se han descubierto proteínas que en condiciones nativas se encuentran fluctuando en torno a dos conformaciones tridimensionales considerablemente diferentes (Frederick, Marlow et al. 2007). Ello quiere decir que para este caso existen dos conformaciones nativas claramente diferenciadas compatibles con la misma secuencia y por tanto alguno de las ideas expuestas anteriormente son incompletas.

Por otro lado, durante la década de los 90 surgió lo que ha pasado a denominarse como Teoría del "Protein Landscape" o como la "New View" de las proteínas (Onuchic, Luthey-Schulten et al. 1997; Plotkin and Onuchic 2002). Su desarrollo vino motivado por la incapacidad de los modelos tradicionales para explicar la cinética de plegamiento y cómo la proteína al plegarse era capaz de resolver la paradoja de Levinthal (que establece que si la proteína explorara todas las posibles conformaciones al azar tardaría en encontrar la conformación nativa un tiempo excesivamente grande). Esta teoría está basada en principios bien conocidos de la química de polímeros y trata

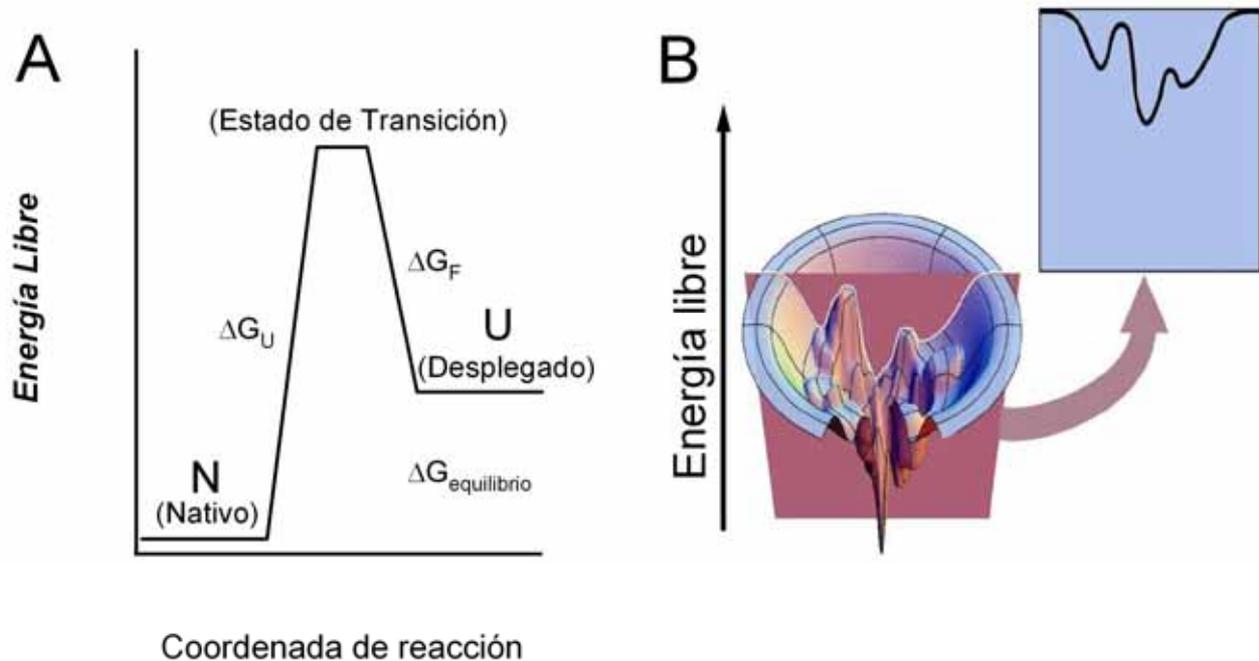
las proteínas como tales (heteropolímeros) sin asumir principios no justificados como una estructura nativa unívoca o la existencia de rutas de plegamiento.

Para resaltar la diferencia con lo expuesto en el apartado anterior, apuntaremos que nuestro diagrama anterior expresaba la energía como función de un solo parámetro de orden. Sin embargo, actualmente se consideran todos los grados de libertad del sistema de modo que se expresa la energía de cada una de las conformaciones que puede adoptar una cadena polipeptídica. Es decir, pasamos de un diagrama bidimensional a un diagrama multidimensional (Figura 2). Por lo que el plegamiento pasa a ser considerado como un proceso de disolución en una hipersuperficie que se define al representar la energía libre interna frente a todas las coordenadas que definen la conformación de la cadena polipeptídica.

Este tipo de planteamientos, carente de restricciones conceptuales, ha permitido predecir con anterioridad a su descubrimiento plegamiento de proteínas en ausencia de barreras de energía libre (García-Mira, Sadqi et al. 2002) y admite con total naturalidad la posibilidad de que existan estados alternativos al nativo (es decir, varios mínimos de energía compatibles con la misma estructura).

En este sentido, el trabajo presentado bajo la denominación de "Artículo Tercero" (Rodríguez-Larrea, Ibarra-Molero et al. 2008) supone una descripción fenomenológica inaudita en la que la proteína pasa, a través del estado desnaturalizado, a un estado nativo "alternativo". Los resultados obtenidos en él son incompatibles con los diagramas de reacción tradicionalmente propuestos (Sanchez-Ruiz 1992). Por otro lado, trabajos anteriores habían resaltado la importancia de las barreras de energía libre para preservar la estructura nativa y funcional de una proteína con respecto al estado desnaturalizado (Sohl, Jaswal et al. 1998). El trabajo presentado parece sugerir que las barreras de energía libre también manifiestan un papel relevante para preservar la estructura nativa y funcional de la proteína con respecto a otras estructuras tridimensionales compatibles con la secuencia. Finalmente, este trabajo presentado sugiere la existencia de barreras de energía libre en el estado desnaturalizado. Su existencia no ha sido documentada hasta la fecha y

actualmente supone una de las líneas de trabajo que se mantienen en el grupo.



**Figura 2:** Se representa el perfil de energía en función de una coordenada de reacción (A) o de dos coordenadas de reacción (B). La teoría del "protein landscape" no limita el número de coordenadas de reacción, de modo que los perfiles de energía que se generan son multidimensionales. En B se puede observar el perfil de energía que encontraría una molécula de proteína siguiendo una determinada trayectoria.

### 1.3.- EVOLUCIÓN DE PROTEINAS

El final del siglo XVIII y el siglo XIX presentaron una serie de peculiaridades significativas entre ellas la introducción sistemática del concepto de evolución a los distintos campos del saber. Así encontramos su aplicación a la ciencia (evolución biológica) y la historia (materialismo histórico). Esencialmente hablamos de evolución cuando un sistema varía desde una posición inicial a una posición posterior movido por una serie de fuerzas cognoscibles.

Desde el punto de vista biológico, que es el que nos atañe, se podría decir que fue el naturalista francés Lamarck (1744-1829) el primero que propuso una teoría evolutiva. Sin embargo sustentó la raíz del cambio gradual de las especies (su evolución) en el uso o desuso de determinados caracteres. Fueron C.R. Darwin (1809-1882) y A.R. Wallace (1823-1913) quienes sugirieron el mecanismo que, con un cierto grado de enriquecimiento, hoy en día se acepta como motor de la evolución de las especies: la selección natural. Esta tiene que actuar sobre un conjunto heterogéneo de individuos, de modo que unos sean seleccionados y otros no. Sin embargo los conocimientos de la época no les permitieron proponer mecanismos que generaran diversidad biológica y, por aquel entonces, sólo se disponían de pequeños conceptos intuitivos sobre la herencia biológica.

Actualmente, gracias a los trabajos de Mendel (1822-1884), al conocimiento de la estructura del ADN, al desciframiento del código genético, así como de otros muchos fenómenos, se ha desarrollado lo que se da a conocer como NeoDarwinismo o Teoría Sintética Evolutiva. De este modo los fenómenos evolutivos se explican mediante la acción conjunta de mutaciones fortuitas y recombinación génica dentro de la población, que explicarían la diversidad biológica. La selección natural y el aislamiento genético son los mecanismos propuestos para explicar la fuerza motriz de la evolución.

Sin embargo esta teoría que se desarrolló para explicar la evolución de los organismos fallaba al tratar de explicar la heterogeneidad observada en determinados elementos moleculares como DNA y proteínas (Ota and Kimura 1971). Fue Motoo Kimura (1924-1994) quien formuló la Teoría de la Evolución Neutra, que hoy con ciertas acepciones se considera válida para explicar la evolución a nivel molecular (Kimura and Ota 1974). Este autor formuló los siguientes 5 puntos:

- El ritmo al que evoluciona una proteína, considerado como número de sustituciones por año y residuo, es aproximadamente constante siempre y cuando estructura y función permanezcan esencialmente inalterados.
- Aquellas moléculas o partes de la molécula que presentan una importancia menor evolucionan más rápido.
- Aquellas mutaciones que producen una perturbación menor en la estructura y función de una molécula ocurren con mayor frecuencia en

la evolución que las mutaciones que producen disrupciones de mayor calado.

-La duplicación génica debe preceder siempre a la aparición de un gen con una nueva función.

-La eliminación selectiva de aquellos mutantes deletéreos y la fijación aleatoria de mutaciones con poco o ningún valor selectivo (mutaciones neutras) ocurren con una frecuencia mucho mayor que la selección Darwiniana positiva sobre mutantes ventajosos.

Estas conclusiones están esencialmente apoyadas en dos observaciones: La mayor parte de las mutaciones que se introducen aleatoriamente en una proteína tienen un efecto nulo o negativo (Steipe, Schiller et al. 1994). Y por otro lado la observación de que existen un gran número de polimorfismos a nivel de ADN y proteínas, sin que ellos aporten ninguna ventaja selectiva.

Posteriormente matizaremos alguno de estos puntos, pero antes conviene desarrollar ligeramente el concepto de Espacio de Secuencias (Sequence Space). Para ello, previamente aclararemos un principio evolutivo: La entidad que sufre evolución no es nunca el organismo, sino la población. Por tanto no podemos limitarnos a pensar que un organismo va sufriendo una mutación tras otra, sino que la población va experimentando muchas de las mutaciones posibles de un modo aleatorio. De este modo, y como una primera aproximación, podemos considerar que se van poblando todas las secuencias compatibles con la estructura y función de la proteína. Dado que la probabilidad de que un mismo gen sufra simultáneamente dos mutaciones es en la práctica despreciable, asumiremos que todas estas secuencias se hallan conectadas a través de un solo evento mutacional (DePristo, Weinreich et al. 2005). Este conjunto de secuencias "válidas", que se encuentran conectadas a través de una sola mutación es lo que se denomina Espacio de Secuencias. La forma y tamaño de este espacio es lo que determinará finalmente la evolución de la proteína dado que es sobre este conjunto de secuencias donde actuará bien la selección natural o bien el aislamiento genético. Resulta obvio, pues, a juicio del que redacta, que la Teoría del Neutralismo no invalida al NeoDarwinismo, sino que lo enriquece y explica.

### **1.3.1 ESPACIO DE SECUENCIAS, LÍMITES SELECTIVOS Y LA HIPÓTESIS DE BOLTZMANN.**

El paisaje dibujado en el apartado anterior mostraba un conjunto de secuencias interconectadas a través de mutaciones puntuales. Dicho conjunto de secuencias se caracterizaba por dar lugar a proteínas que mantenían la estructura y la función en el organismo. Esta característica parece lógica y plausible. Si la reexaminamos brevemente podemos reinterpretar el Espacio de Secuencias como todas aquellas secuencias que estando conectadas a través de mutaciones simples satisfacen un mínimo. Si bien la selección opera sobre el organismo, cabe presuponer una conexión entre la secuencia (que da lugar a las propiedades de la proteína) y la selección del organismo, dando lugar a que en apariencia haya una selección sobre la secuencia. Por ello, a este mínimo es lo que se denomina "Fitness threshold" (límite selectivo), y marca la selección negativa de un organismo (que se relaciona con una secuencia dada) cuando este tiene un valor de "Fitness" menor al límite.

La secuencia de la proteína evoluciona a través de mutaciones puntuales que mantienen al organismo siempre por encima de este valor. Asumiendo que la probabilidad de mutación es la misma para cada posición y teniendo en cuenta que existe un límite que marca si las mutaciones se fijan o se rechazan, llegamos a un escenario de extraordinario valor estadístico: Si ha pasado suficiente tiempo se habrá establecido un equilibrio mutacional para cada posición (Shortle 2003). En dicha situación las mutaciones van siendo más frecuentes cuanto más positivo sea el valor que le confiere a la secuencia (siempre con respecto al límite o "threshold") y tanto menos frecuentes cuanto más pernicioso (negativo) es el efecto. La existencia de este equilibrio supone que la hipótesis de Boltzmann aplica a la evolución de la secuencia. De este modo, la probabilidad de encontrar una mutación determinada en una posición dada viene dada por:

$$p(x) = \exp\left(\frac{-\Delta G}{RT}\right) \quad (\text{Ec. 1})$$

Donde la energía es una medida del efecto de la mutación en el "fitness" del organismo (ver Figura 3). Por tanto, a priori es

posible extraer información útil a partir de un simple análisis estadístico de una base de datos de secuencias.

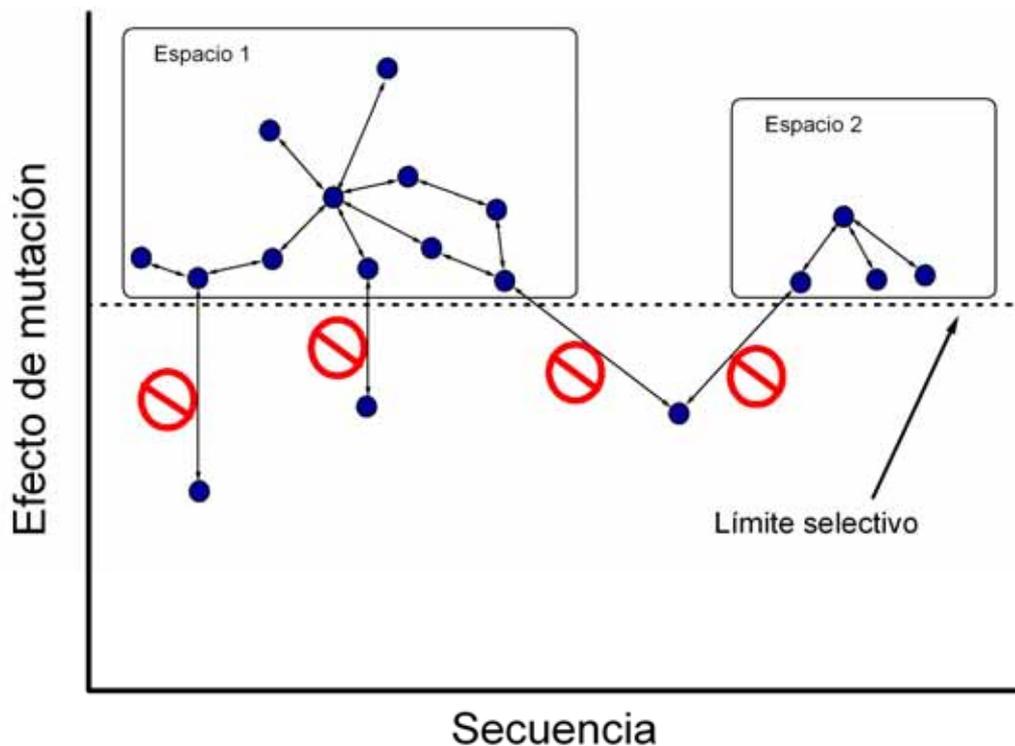
Una de las ventajas del modelo evolutivo que se propone es la "extremada sencillez" para modelarlo matemáticamente y para testarlo experimentalmente. Ésta es una de las razones de que haya sido exhaustivamente analizado bajo distintas premisas que sofistican el modelo (Xia and Levitt 2004). En términos generales estos modelos atribuyen a cada mutación una energía y establecen un límite de selección (Ecuación 1). Posteriormente realizan un estudio exhaustivo del espacio de secuencias o bien lo muestrean, siempre como función de la accesibilidad computacional de dicho cálculo.

Una vez la simulación ha alcanzado el estado estacionario (hay que recordar que en estas aproximaciones se suele imponer una tasa constante de mutación) dos propiedades emergen(Xia and Levitt 2004):

-Las proteínas son mutacionalmente robustas. Esto quiere decir que las secuencias que suelen poblarse dentro de una población que está evolucionando son aquellas que poseen mayor número de conexiones viables dentro del espacio de secuencias, siempre y cuando la población tenga un tamaño suficientemente grande. (Ver Figura 3) (Bloom, Lu et al. 2007).

-Las proteínas son marginalmente estables (con respecto al límite). Esto resulta del gran número de secuencias accesibles cercanas al límite de selección, un efecto en ocasiones denominado en la bibliografía como entropía del espacio de secuencias.

Como prueba de la validez de que la hipótesis de Boltzmann aplica a la evolución de las proteínas, podemos citar un trabajo del grupo de investigación en el que se ha realizado esta tesis doctoral (Godoy-Ruiz, Ariza et al. 2006). En él se observa una clara correlación entre el efecto en la estabilidad de mutaciones conservativas (aquellas que sustituyen residuos determinados por otros residuos de parecidas características físico-químicas) y la frecuencia de aparición de uno u otro en un alineamiento de secuencias. Efectivamente resulta que la estabilidad de la proteína es uno de los factores que más influyen en que se acepte/rechace una mutación dada. La importancia de la estabilidad en la evolución de una proteína se discutirá en un apartado posterior de la presente introducción.



**Figura 3:** Se representa el efecto de las mutaciones sobre una proteína. Cuando la mutación rebaja el valor de la secuencia por debajo del límite selectivo, es rechazada. Ello hace que se vayan poblando más aquellas secuencias con mutaciones ventajosas. Nótese además que el número de secuencias cercanas al límite es mayor (entropía del espacio de secuencias). Se representan dos espacios de secuencia separados, dado que para llegar de uno a otro hay que pasar por una secuencia que es seleccionada negativamente. Si el límite se rebaja dicho camino se hace posible. Dentro del espacio de secuencias no hay selección negativa, por lo que rige la evolución neutra.

Pese a lo dicho anteriormente caben dos situaciones extremas que completan la visión de la evolución de proteínas. La primera de ellas supone la presencia de límites selectivos extremadamente exigentes como sería el caso, por ejemplo, del desarrollo de resistencia a un antibiótico de nueva generación (Weinreich, Delaney et al. 2006). En este caso las teorías NeoDarwinianas se aproximan a la realidad. Ello se debe a que el espacio de secuencias se ve extremadamente reducido y el conjunto de mutaciones que ha de acumular la proteína para llegar a desarrollar la resistencia contra el nuevo antibiótico debe seguir un

orden secuencial determinado (la introducción de las mutaciones en otro orden suponen rebajas en el "fitness" y son selectivamente rechazadas). Por tanto, este modo de evolución se aproxima al modo de evolución gradual propuesto por Darwin.

El otro caso extremo trata sobre las actividades promiscuas en proteínas (Khersonsky, Roodveldt et al. 2006). Por promiscuidad entendemos la capacidad de algunas enzimas para llevar a cabo reacciones químicas distintas de aquellas para las cuales fueron evolucionadas. Este tipo de evolución puede presentar una especial importancia en la evolución de resistencias contra drogas, en la evolución de nuevos receptores,... y, desde un punto de vista fundamental, en la evolución de las distintas rutas metabólicas que podemos encontrar hoy en día (Aharoni, Gaidukov et al. 2005).

Este caso lleva la teoría de la evolución neutra a su máximo, puesto que supone que una proteína posee una nueva función (aunque sea de modo residual) sin perder la primigenia. Esto se explica, conforme a la teoría de la evolución neutra, porque la proteína explora todo el espacio de secuencias que le es accesible (aquel donde las secuencias están interconectadas por mutaciones simples y siempre por encima de los límites de "fitness"). Algunas de estas secuencias dan lugar a proteínas con la capacidad de catalizar reacciones químicas sobre nuevos sustratos de un modo residual (en ocasiones reacciones de una naturaleza considerablemente distinta a la "natural"). La clave para entender la importancia de este fenómeno es que estas secuencias son capaces de evolucionar extremadamente rápido hacia proteínas con una capacidad catalítica excelente hacia el nuevo sustrato. Dado que una vez llegados a este punto es de suponer que la actividad primigenia quedase relegada a una actividad marginal o nula, para que se diera este último paso evolutivo deberíamos encontrarnos en un escenario donde la presión selectiva por la vieja función desapareciera (escenario que se daría, por ejemplo, después de un proceso de duplicación génica) (Miller and Raines 2004).

Para resaltar la diferencia entre las teorías del NeoDarwinismo y este caso extremo de Evolución Neutra indicaremos cómo el NeoDarwinismo explica la aparición de una nueva función: En un primer momento ocurre un evento de duplicación génica. Posteriormente la secuencia evoluciona hacia la nueva función. Esta explicación se ve fuertemente rebatida por el hecho de que la mayor parte de las

mutaciones son deletéreas (Guo, Choe et al. 2004). Ello, junto con el descubrimiento de enzimas promiscuas, hace más factible que la existencia de individuos en la población con proteínas "multifunción" sea el punto de origen hacia la adquisición de una nueva función.

### **1.3.2 ALGUNOS FACTORES QUE DETERMINAN LA "EVOLUTIVILIDAD" (EVOLVABILITY) DE UNA PROTEÍNA.**

Cabe preguntarse cuáles son los factores, más allá de la función, que determinan la evolución de la secuencia de una proteína.

Uno de los factores que determina la evolución es el tamaño de la población (la unidad de evolución en biología) y la tasa de mutación (número de mutaciones por generación) (Bloom, Raval et al. 2007). Si bien estos factores pertenecen más al campo de la genética de poblaciones, mencionaremos que tanto en simulaciones (Bloom, Raval et al. 2007) como a nivel experimental (Bloom, Lu et al. 2007) se ha verificado que en grandes poblaciones con tasas de mutación elevadas las proteínas evolucionan hacia una "robustez mutacional" ligada a un "exceso de estabilidad".

Y es que, como ya se ha mencionado, uno de los factores que parece tener una mayor relevancia es la estabilidad de la proteína. Ello se debe posiblemente a que la estabilidad de la proteína determina si la secuencia de aminoácidos será capaz de plegarse y hacerlo del modo correcto, lo cual es esencial para la función. Además, la estabilidad está relacionada también con la susceptibilidad a que la proteína sea degradada por el sistema de proteasas de la célula y la tendencia a la agregación (DePristo, Weinreich et al. 2005). Del mismo modo que aquellos residuos importantes para la catálisis se muestran conservados en la evolución, también cabe esperar que aquellos residuos implicados significativamente en la estabilidad muestren algún grado de conservación estadística, como veremos en el siguiente apartado y en el trabajo del Apéndice 1.

Por otro lado, la estabilidad confiere una propiedad extremadamente importante para la evolución: La robustez ante mutaciones (Bloom, Labthavikul et al. 2006). Esta propiedad es importante dado que la generación de un sitio activo generalmente conlleva la desestabilización de la estructura plegada. Por otro lado cabe mencionar que, como se indicó en el apartado anterior, no tiene

por qué ser posible pasar de una secuencia a otra a través de mutaciones puntuales, dado que alguno de los intermedios puede presentar una estabilidad excesivamente baja (en cuyo caso siempre sería rechazado por la evolución). Por tanto, cuando el punto de partida es una secuencia que genera una proteína hiperestable aumentan las posibilidades de evolución.

Este último punto viene ejemplarizado en un reciente trabajo de E.I. Shakhnovich (Berezovsky and Shakhnovich 2005). En él, tras realizar un amplio estudio comparativo entre proteínas de organismos hipertermófilos, se observaron dos mecanismos de estabilización. Por un lado un incremento en el contenido hidrofóbico y en el grado de empaquetamiento del núcleo de la proteína. Por el otro, una distribución óptima de cargas en la superficie. Al clasificar los organismos por su filogenia y comprobar qué tipo de organismo tenía qué tipo de estabilización se observó un dato revelador: Aquellos organismos cuya filogenia proviene de hipertermófilos tenían optimizado el núcleo hidrofóbico. Los que provenían de mesófilos optimizaban las cargas de la superficie. Esta diferencia queda explicada por la "imposibilidad" de generar un núcleo hidrofóbico compacto a base de mutaciones puntuales: una mutación puntual en el núcleo difícilmente estabilizará si los residuos de su ambiente se mantienen constantes. De este modo se ejemplifica el devenir de una secuencia de aminoácidos ante los distintos límites evolutivos y se resalta la importancia de un "extra" de estabilidad para permitir la adquisición de nuevas funciones o propiedades.

Otro de los puntos que se ha propuesto que determinan la evolución de una proteína es su nivel de expresión. Dos mecanismos han sido propuestos para explicar este fenómeno (Drummond, Bloom et al. 2005). Por un lado se trataría de optimizar la lectura de nucleótidos que codifican la proteína (no todos los tripletes tienen la misma eficiencia en la bacteria). Por otro lado, se trataría de obtener secuencias de proteínas más robustas ante fallos en el proceso de traducción. De este modo se minimizarían los posibles plegamientos incorrectos (de nuevo, robustez ante mutaciones... aunque en este caso la mutación es un fallo de traducción). Por ello, esta hipótesis sugiere que las proteínas con altos niveles de expresión están diseñadas para tener un "extra" de estabilidad, lo que supone una restricción al espacio de secuencias que pueden experimentar y, como consecuencia, evolucionan más lentamente.

Existen otros factores que determinan la evolución de las proteínas, como la accesibilidad de la mutación a nivel de tripletes de nucleótidos en el ADN), lo "cara" que le salga a la célula sintetizar la proteína (pues no todos los aminoácidos son igual de fáciles de sintetizar),... Sin embargo consideraremos que las bases biofísicas de la evolución de proteínas están suficientemente explicadas para introducir el trabajo presentado en el Apéndice 1.

### **1.3.3 ESTADÍSTICA Y ALINEAMIENTO DE SECUENCIAS.**

Un alineamiento de secuencias consiste en una disposición ordenada de secuencias. El orden al que nos referimos trata de maximizar las coincidencias de aminoácidos para cada posición conforme a una secuencia diana de interés. Ello, a priori, nos permite examinar visualmente el parecido entre secuencias relacionadas, o que creemos relacionadas.

Una vez obtenido este orden, resulta factible realizar sobre el alineamiento un análisis estadístico. Los análisis estadísticos pueden comprender una considerable complejidad, pero no es objeto de esta introducción realizar un estudio sobre ellos sino contextualizar el trabajo presentado en el Apéndice 1. Por ello nos limitaremos a describir los métodos más sencillos de análisis (alguno de los cuales se ha utilizado en el mencionado Apéndice 1, y otro de los cuales se ha utilizado pero no forman parte del proyecto de tesis doctoral que se presenta). Estos métodos serán Consenso, Conservación, Covarianza y Reconstrucción Filogenética. La validez de estos sencillos métodos se ejemplificará con un objetivo final: Demostrar que la información que nos suministran las bases de datos de secuencias, bien de proteínas bien de ácidos nucleicos, poseen información extremadamente útil en la comprensión y diseño de proteínas.

El análisis más sencillo que podemos realizar sobre un alineamiento de secuencia es un conteo de frecuencias aminoacídicas para cada posición. De este modo, podemos obtener lo que se denomina "secuencia consenso" de la proteína. Esta secuencia es aquella que comprende los aminoácidos más frecuentes para cada posición. No hay que obviar, ni en este análisis ni en los posteriores, que el alineamiento de secuencias se ha de realizar bajo unos criterios de similaridad "estrictos", como por ejemplo una similaridad de secuencia

mínima del 25%. Este valor se suele tomar dado que dos proteínas que compartan dicha similitud suelen adquirir un mismo plegamiento, salvo excepciones. La hipótesis detrás de este análisis es que el aminoácido más frecuente será el más beneficioso, pues ha sido seleccionado más frecuentemente que los demás.

Este tipo de análisis ha sido efectivo en la predicción de mutaciones que estabilizarían la proteína diana en numerosas ocasiones (si bien es cierto que sigue habiendo un número significativo de fallos) (Steipe, Schiller et al. 1994). Los cambios en energía libre de plegamiento y en energía libre de activación experimentales muestran una excelente correlación con la energía estadística obtenida a partir de la hipótesis de Boltzmann mencionada anteriormente (Godoy-Ruiz, Ariza et al. 2006). El efecto que tendrá la mutación en la estabilidad puede ser estimado a través de:

$$\Delta\Delta G(a \rightarrow b) = RT \ln N_a / N_b \quad (\text{Ec. 2})$$

Donde  $N_a$  es el número de residuos de tipo "a" y  $N_b$  para la posición "b" que estamos evaluando en el alineamiento

Uno de los problemas que puede originar este análisis es que podría estar sesgado en caso de que nuestro alineamiento de secuencias no represente estadísticamente bien el equilibrio mutacional-selectivo sobre cada posición.

La alternativa reside en referenciar la frecuencia en el alineamiento de cada uno de los aminoácidos de nuestra proteína diana con la frecuencia de aparición de cada uno de los veinte aminoácidos entre todas las secuencias de proteínas conocidas. Esta aproximación es la que se conoce por Conservación y solucionaría, sólo en parte, el problema del análisis anterior.

Para aclarar las diferencias entre las aproximaciones de Consenso y Conservación, resaltaremos que la primera presupone que dentro de una misma familia de proteínas (las secuencias del alineamiento) se ha dado un equilibrio mutacional, de modo que son aceptadas o rechazadas como función del efecto que producen, en último término, sobre el "fitness" y que por tanto el aminoácido más frecuente será el selectivamente más beneficioso. La segunda presupone

que es la desviación con respecto a la frecuencia de aparición del aminoácido en el conjunto de proteínas la que informa sobre el efecto que produce en el "fitness".

Otro análisis posible reside en realizar un estudio de covarianzas. Está también basado en energías estadísticas (basado pues en la validez de la hipótesis de Boltzmann gracias a la existencia de un equilibrio mutacional para cada posición), es decir, en probabilidades (o frecuencias relativas). Sin embargo aquí la frecuencia que se examina es la de que dos aminoácidos se encuentren simultáneamente mutados en una misma secuencia dentro del alineamiento. Y, de nuevo, nuestra frecuencia de referencia para obtener la energía estadística podrá provenir del propio alineamiento o bien de las probabilidades obtenidas en toda la base de datos de secuencias. En este caso, precisaríamos de un alineamiento de secuencias considerablemente mayor para obtener probabilidades de covarianza "creíbles". Este tipo de análisis ha servido para detectar tanto residuos de alta importancia en el plegamiento de proteínas (Jain and Ranganathan 2004) como residuos energéticamente acoplados (Lockless and Ranganathan 1999) (Perez-Jimenez, Godoy-Ruiz et al. 2006). La importancia de este tipo de acoplamiento queda de manifiesto, por ejemplo, cuando la unión de un ligando a una región de la proteína produce modificaciones estructurales en otra región, explicando así fenómenos como el alosterismo.

Por último, el filogenético es uno de los análisis más frecuentemente realizados sobre alineamiento de secuencias. Las premisas bajo las cuales se realizan dichos análisis comprenden los métodos de máxima parsimonia,... Pero sin entrar en demasiado detalle mencionaremos que se basan principalmente en matrices de sustitución (la probabilidad de que un residuo mute a otro) y en tasas de mutación. En cualquier caso, este tipo de análisis manifiesta un serio problema: la dificultad en la verificación de su validez. No obstante, la conjunción de dos trabajos recientemente publicados permiten "creer en su validez". El primero de ellos realiza la construcción del árbol evolutivo gracias al ingente número de genomas secuenciados (Ciccarelli, Doerks et al. 2006). Otro grupo, basándose en este árbol filogenético, reconstruye la evolución de una proteína, purificando y obteniendo la proteína pertinente de cada nodo del árbol. Tras su caracterización biofísica, el resultado, sencillamente, asombra: Las proteínas reconstruidas existieron según el modelo en determinados

momentos de la historia. Sus temperaturas de desnaturalización siguen las distintas tendencias supuestas para la evolución de la temperatura en la historia de la Tierra (Gaucher, Govindarajan et al. 2008). Estos trabajos están sustentados en muchas de las ideas expuestas anteriormente y si bien no es prueba de validez, sí es indicativo de la veracidad de muchas de las hipótesis propuestas.

#### **1.3.4 RELACIÓN DE LA EVOLUCIÓN DE PROTEÍNAS CON EL PRESENTE PROYECTO DE TESIS**

Uno de los mayores problemas que plantean las proteínas a la hora de su uso en biotecnología es la falta de estabilidad. Dado que su obtención no es sencilla ni barata, la aplicación de las proteínas a los distintos usos industriales o biomédicos requiere de una optimización de la longevidad operativa de la proteína. Esto es, se suele requerir un incremento de estabilidad.

Sin embargo la optimización en la estabilidad de una proteína suele estar anticorrelacionada con la actividad catalítica de la (Wang, Minasov et al. 2002), lo cual representa un serio problema si nuestro objetivo final es aprovecharnos de sus propiedades. Hasta donde llega nuestro conocimiento, sólo en un caso se ha conseguido optimizar simultáneamente la actividad y la estabilidad (Miyazaki, Wintrode et al. 2000). Sin embargo, dicho trabajo está realizado sobre una proteína (y sus formas mutantes) que ha sufrido una extensa caracterización en la bibliografía dado su alto interés industrial.

Actualmente y a grandes rasgos, existen 3 alternativas para la obtención de proteínas mutantes con alguna propiedad mejorada:

-El primero de ellos es el diseño racional. Bien a través de cálculos electrostáticos, modelizaciones o a través de otro tipo de aproximaciones, se sugieren una serie de mutaciones que presuntamente mejorarán la proteína. Cualquier aproximación de este tipo tiene el inconveniente de precisar de una estructura tridimensional de la que partir, lo cual no es siempre posible.

-Por otro lado tenemos los métodos de evolución dirigida. Ellos consisten en la generación de una amplia librería de mutantes aleatorios, los cuales son evaluados (screening) por la propiedad buscada. El inconveniente general es que la obtención de proteínas

mejoradas depende mucho de la capacidad de screening (hay que considerar que frecuentemente se requiere evaluar miles de mutantes distintos). Por otro lado, la aproximación en ocasiones implica la pérdida de alguna de las propiedades de la proteína por la que no hemos hecho selección.

-Finalmente encontramos los métodos basados en información evolutiva. Su fundamento es sencillo: asumir que la naturaleza ya ha realizado el trabajo de laboratorio y que simplemente tenemos que saber leer la información. Nuestro trabajo presentado en el Apéndice 1 se basa en dicha información para generar una pequeña librería de mutantes.

Por tanto, apoyados en las teorías evolutivas mencionadas anteriormente y haciendo uso de las cada vez más desmesuradas grandes bases de datos de secuencias, diseñamos una aproximación experimental que permite la optimización simultánea de distintas propiedades de la proteína. Hasta donde llega nuestro conocimiento, se trata del primer trabajo que considera la coexistencia de distintas presiones selectivas cuyo efecto es visible tras un análisis estadístico de un alineamiento de secuencias. Del mismo modo que, exceptuando el trabajo antes citado (Miyazaki, Wintrobe et al. 2000) por sus particularidades, este trabajo es el primero que permite la optimización simultánea de estabilidad termodinámica, cinética y función. Y presenta, además, vista la generalidad de los principios aplicados, la posibilidad de extenderlo a proteínas de las cuales ni tan siquiera existe estructura tridimensional resuelta (siendo por tanto especialmente aplicable a proteínas de membrana de interés). Si bien este trabajo está en vías de publicación y por tanto no puede formar parte del cuerpo de este proyecto de tesis, hemos decidido incluirlo en el Apéndice 1 dada su enorme relevancia dentro de las investigaciones llevadas a cabo.

## **2.-ARTICULOS PUBLICADOS**

**PRIMER ARTÍCULO**

"Role of solvation barriers in protein kinetic  
stability"

Rodriguez-Larrea D, Minning S, Borchert TV,  
Sanchez-Ruiz JM.

**Journal of Molecular Biology.** 2006 Jul  
14;360(3):715-24.

INTRODUCCIÓN ARTÍCULO 1º

La estabilidad de muchas proteínas en ocasiones, viene determinada por aspectos cinéticos. A efectos prácticos esto tiene un gran interés no sólo por las implicaciones obvias en potenciales aplicaciones biotecnológicas de las mismas, sino también por el hecho de que algunos procedimientos moleculares emergentes para la inhibición de amiloidogénesis se centran en un aumento de la estabilidad cinética de los estados nativos de proteínas (Hammarstrom, Wiseman et al. 2003).

Ciertamente, la literatura científica se ha centrado en mayor medida en el estudio de la estabilidad termodinámica. Ello se debe a que extraer información de sistemas donde la estabilidad cinética es la relevante resulta más complicado. La Lipasa de *Thermomyces lanuginosa* es una de las proteínas con un mayor abanico de usos en industria (fabricar papel, detergentes, industria maderera...), por tanto una mejora de su estabilidad es un objetivo deseable y muy ambicioso.

En el presente trabajo nos planteamos, en colaboración con la empresa Novozymes (líder mundial en biotecnología en cuanto a cuota de mercado), estudiar la energética de la lipasa. En los inicios de dicha colaboración los Doctores J.M. Sánchez Ruiz y Beatriz Ibarra Molero (directores, además, de esta tesis), trataron de aumentar la estabilidad de la proteína diseñando la distribución de cargas superficiales. Para ello utilizaron una metodología de búsqueda de mutaciones basada en un algoritmo genético (Ibarra-Molero and Sanchez-Ruiz 2002). Sin embargo, las mutaciones realizadas bajo estos principios no surtieron el efecto buscado. Es por ello que se inició una caracterización más detallada tanto de la forma silvestre (Wild Type) como de los distintos mutantes.

Como primer paso de este trabajo se estudió la desnaturalización térmica de las proteínas por DSC y se comprobó que todas las formas respondían a un modelo de desnaturalización de dos estados irreversible. Esto quiere decir que son las diferencias entre el estado nativo y el estado de transición las que determinan la estabilidad del estado funcional de la proteína. Dicho de otro modo, existe una barrera energética que se opone a la desnaturalización, y cuanto mayor sea esta barrera menor la facilidad con la que la proteína desnaturaliza. Es en estos casos cuando decimos que la desnaturalización de la proteína tiene un control cinético.

La desnaturalización irreversible de proteínas se comenzó a estudiar con interés a finales de los ochenta y principios de los noventa (Sanchez-Ruiz, Lopez-Lacomba et al. 1988). Contemporáneamente apareció la caracterización del estado de transición basándose en los efectos energéticos de mutaciones estratégicamente elegidas, metodología comúnmente conocida como "valores phi" (Matouschek, Kellis et al. 1989). Lamentablemente esta metodología, así como otras basadas en los mismos principios ("valores psi", (Krantz, Dothager et al. 2004)), no es aplicable al estudio de estados de transición en sistemas irreversibles, dado que no es posible caracterizar la energética del equilibrio. Por tanto no existen métodos indirectos que puedan dar información sobre la relación estructura-energética del estado de transición en desnaturalización irreversible de proteínas. Una de las posibles aproximaciones para vencer los inconvenientes es el uso de la Dinámica Molecular. Basados en datos obtenidos en simulaciones realizadas a alta temperatura, los científicos de Novozymes identificaron posibles regiones de máxima flexibilidad. Tras el uso de técnicas de generación de variantes aleatorias dentro de estas regiones, consiguieron proteínas mutantes con una estabilidad significativamente mayor. Algunas de estas variantes fueron amablemente cedidas por la empresa y formaron parte del trabajo que presentamos. El siguiente paso del estudio se inició considerando los siguientes puntos:

-La proteína desnaturaliza irreversiblemente. La estabilidad la determina la diferencia energética entre el estado nativo y el estado de transición.

-Las mutaciones de carga sugeridas no funcionaron. Hay que resaltar aquí que el método utilizado optimiza la distribución de cargas del estado nativo asumiendo que el estado con respecto al cual se estabiliza es un estado desnaturalizado totalmente desplegado.

-Cuando se realizan mutaciones en las regiones flexibles de la proteína, identificadas por Dinámica Molecular, se consigue un considerable éxito en la estabilización de la proteína.

Es razonable pensar que estructura y energética se hallan íntimamente ligados. Bajo esta premisa y con la información anterior, parece probable que la ausencia de estabilización en las mutaciones

de carga tenga su origen en una similitud estructural entre el estado nativo y el estado de transición. Por otro lado, pensamos que las simulaciones de Dinámica Molecular podrían haber identificado aquellas regiones con un cierto grado de desplegamiento en el estado de transición.

Por ello nos planteamos hacer uso de codisolventes como la urea, cuyo efecto en la estabilidad depende de la diferencia en accesibilidad al solvente entre los estados implicados (Street, Bolen et al. 2006). Este efecto en la estabilidad se parametriza como valores "m", que serán de equilibrio cuando hablemos del efecto en la estabilidad termodinámica y serán cinéticos cuando hablemos del efecto en la estabilidad cinética. Dado que las mutaciones apenas afectan al área accesible del estado nativo, los experimentos de calorimetría diferencial de barrido (DSC) en presencia de urea nos dan una idea del cambio de área accesible en el estado de transición.

Tras el estudio del efecto de la urea en la estabilidad de la proteína wild type, quedó claro que la lipasa presentaba un estado de transición bastante compacto. Dicho de otro modo, la proteína, al pasar del estado nativo al estado de transición expone tan solo unos pocos residuos. Cuando se realizó el mismo estudio sobre las formas mutantes se obtuvieron estimaciones de exposición al solvente (a partir del efecto de la urea en la estabilidad) y valores de energía de activación. Los resultados revelaron un hecho sorprendente: La exposición al solvente en el estado de transición no correlacionaba con los cambios en entalpía de activación. La situación "normal" es que un alto valor de entalpía de activación vaya acompañado de un gran cambio de área accesible al solvente (Robertson and Murphy 1997).

Previo al presente trabajo, estudios teóricos (Liu and Chan 2005) manifestaron la posible importancia de la naturaleza particulada del agua en la energética de los estados de transición. En la energética del equilibrio los enlaces internos que pierde la proteína son compensados por enlaces con las moléculas del solvente. Sin embargo en un estado de transición, que es un estado de máxima energía, podemos encontrar residuos que hayan perdido todas o parte de sus interacciones. Esto se traduciría en un gran cambio en la entalpía de la reacción que no iría acompañado de un cambio en la accesibilidad al solvente (al contrario de lo que sucedería en la energética del equilibrio).

En este trabajo sugerimos, por tanto, que es la asincronía existente entre la rotura de interacciones entre residuos y la formación de nuevas interacciones con las moléculas de agua la responsable de los grandes cambios en la energía de activación detectados para las formas mutantes estudiadas. De esta forma, este trabajo, hasta donde llega nuestro conocimiento, es el primero que identifica experimentalmente la existencia de barreras entálpicas de solvatación y su contribución a la estabilidad cinética de la proteína.

## Role of Solvation Barriers in Protein Kinetic Stability

David Rodriguez-Larrea<sup>1</sup>, Stefan Minning<sup>2</sup>, Torben V. Borchert<sup>2</sup>  
and Jose M. Sanchez-Ruiz<sup>1\*</sup>

<sup>1</sup>Departamento de Química Física, Facultad de Ciencias 18071-Granada, Spain

<sup>2</sup>Research and Development Novozymes A/S, Brudelysvej 26, 2880 Bagsvaerd, Denmark

The stability of several protein systems of interest has been shown to have a kinetic basis. Besides the obvious biotechnological implications, the general interest of understanding protein kinetic stability is emphasized by the fact that some emerging molecular approaches to the inhibition of amyloidogenesis focus on the increase of the kinetic stability of protein native states. Lipases are among the most important industrial enzymes. Here, we have studied the thermal denaturation of the wild-type form, four single-mutant variants and two highly stable, multiple-mutant variants of lipase from *Thermomyces lanuginosa*. In all cases, thermal denaturation was irreversible, kinetically controlled and conformed to the two-state irreversible model. This result supports that the novel molecular-dynamics-focused, directed-evolution approach involved in the preparation of the highly stable variants is successful likely because it addresses kinetic stability and, in particular, because heated molecular dynamics simulations possibly identify regions of disrupted native interactions in the transition state for irreversible denaturation. Furthermore, we find very large mutation effects on activation enthalpy and entropy, which were not accompanied by similarly large changes in kinetic urea *m*-value. From this we are led to conclude that these mutation effects are associated to some structural feature of the transition state for the irreversible denaturation process that is not linked to large changes in solvent accessibility. Recent computational studies have suggested the existence of solvation/desolvation barriers in at least some protein folding/unfolding processes. We thus propose that a solvation barrier (arising from the asynchrony between breaking of internal contacts and water penetration) may contribute to the kinetic stability of lipase from *T. lanuginosa* (and, possibly, to the kinetic stability of other proteins as well).

© 2006 Elsevier Ltd. All rights reserved.

\*Corresponding author

**Keywords:** protein stability; kinetic stability; irreversible denaturation; free-energy barriers; solvation/desolvation

### Introduction

Protein stability has thermodynamic, as well as kinetic aspects.<sup>1,2</sup> Early evidence for the kinetic control of protein stability arose during the late eighties and the early nineties as a result of differential scanning calorimetry (DSC) studies into calorimetrically irreversible transitions for protein denaturation, which showed a strong scanning-rate dependence of the DSC transitions (a signature of kinetic control) for many protein systems (see Plaza del Pino *et al.*<sup>1</sup> and

references quoted therein). From a more general viewpoint, these DSC studies made clear that thermodynamic stability (a positive value for the unfolding free energy at physiological temperature) does not guarantee that the protein will remain in the native (biologically functional) state during a given time scale, since irreversible protein alterations (even if they occur from lowly populated unfolded or partially unfolded states) may deplete the native state in a time-dependent manner. It appeared likely then (see Plaza del Pino *et al.*<sup>1</sup> for discussion) that many proteins (in particular, complex protein systems) must have been naturally selected to have significant kinetic stability.<sup>1,3</sup> Indeed, in recent years, the stability of many complex protein systems of interest has been shown to have a kinetic origin; examples include the Newcastle disease virus,<sup>4</sup>  $\alpha$ -lytic

Abbreviations used: DSC, differential scanning calorimetry; WT, wild-type.

E-mail address of the corresponding author: sanchezr@ugr.es

protease,<sup>5-8</sup> Cu/Zn superoxide dismutase,<sup>9,10</sup> the viral capsid protein SHP<sup>11</sup> and human low-density lipoprotein.<sup>12</sup> Also, evidence for selective evolutionary pressure on the unfolding barriers of kinetically stable proteases has been recently presented.<sup>3</sup>

The interest of understanding protein kinetic stability is emphasized by the fact that some emerging molecular approaches to the inhibition of amyloidogenesis focus on the increase of the kinetic stability of the protein native state,<sup>13-15</sup> as was in fact suggested by our analyses of simple Lumry-Eyring models of irreversible denaturation.<sup>1</sup> Furthermore, kinetic stability may be of considerable biotechnological importance. Researchers interested in fundamental aspects of protein folding may generally choose "model" proteins and solvent conditions in such a way that equilibrium folding/unfolding is observed. However, the proteins and/or solvent conditions employed in technological applications often imply irreversible denaturation and kinetic control of the stability (a clear example is actually provided by the protein studied here: see below). In fact, enhancing protein stability for biotechnological applications may in many cases mean enhancing protein kinetic stability.

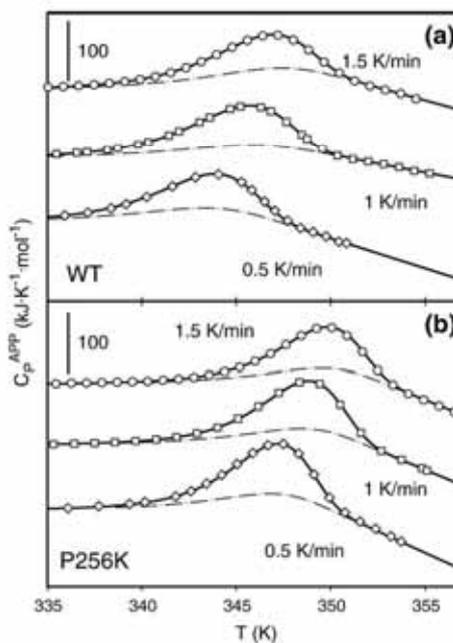
Lipases are among the most important industrial enzymes, being used in many different industrial segments<sup>16</sup> (detergent, food, baking, pulp and paper, etc.). Here, we have studied by DSC the thermal denaturation of the WT form and six variants of lipase from *Thermomyces lanuginosa* under conditions (pH 5) relevant for its practical applications. Four of the variants were single mutants (P256K, P256T, Q4E, V154L), some of them suggested by the simple electrostatic approach that we have previously used to design stabilizing mutations (see Sanchez-Ruiz & Makhatadze<sup>17</sup> and references quoted therein). The two multiple-mutant variants (with five and seven mutations: D27R/D111A/L227G/S216P/P256T and D27R/G91R/N94K/D111A/S216P/L227G/P256T) have been previously obtained using a novel molecular-dynamics-focused, directed-evolution approach, as described by Minning *et al.*,<sup>18</sup> (these two variants are referred to as Smin041 and SMin149 by Minning *et al.*<sup>18</sup>). Most of the mutations in these lipase variants studied involve surface positions in the lipase structure,<sup>19</sup> as shown by the accessible surface area (ASA) values of the WT residues at those positions. The only exceptions appear to be position 154 and position 216, although Ser216 is actually close to the protein surface in the lipase structure.

## Results and Discussion

### Thermal denaturation of the WT form and variants of lipase from *Thermomyces lanuginosa* is calorimetrically irreversible, kinetically controlled and conforms to the two-state irreversible model

The WT form, the four single-mutant variants and the two multiple-mutant variants of lipase

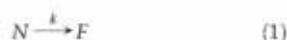
from *T. lanuginosa* have been studied in aqueous buffer (50 mM sodium acetate, pH 5) in the absence as well as in the presence of 1 M and 2 M urea. For each protein/solvent system, at least three different DSC experiments were carried out at different scan-rates. The DSC profiles for the seven protein forms were similar: a clearly defined transition with maximum within the 340–360 K range (depending on the variant, the urea concentration and the scan-rate) was followed by a drop in calorimetric signal associated to protein aggregation (see illustrative examples in Figure 1). In fact, aggregation was evident in the samples extracted from the calorimetric cells after the DSC experiments. Actually, calorimetric experiments analyzed here were carried out at low protein concentrations (about 0.1 mg/ml) in order to minimize the aggregation effects. We did check in several representative cases, nevertheless, that the temperature for the maximum of the transitions was independent of protein concentration



**Figure 1.** Illustrative examples of the experimental DSC thermograms for lipase denaturation at the indicated scan rates. The profiles shown correspond to WT lipase and the P256K variant in the absence of urea, although profiles for these and other variants in the presence and absence of urea were qualitatively similar. Open symbols are the experimental apparent heat capacity data and the continuous lines are the best fits of the two-state irreversible model. Note that the profiles have been shifted in the *y*-axis for the sake of clarity.

within the 0.1–1 mg/ml range. In all cases, the DSC transitions were calorimetrically irreversible, as shown by their absence in a reheating run performed after stopping the first run immediately after the transition. They were also strongly scan-rate-dependent (see Figure 1 for illustrative examples), indicating kinetic control of the denaturation process.

We have found that the shape of the DSC transitions as well as the scan-rate effect on them can be adequately explained by the simple two-state irreversible model<sup>20</sup> that has been previously used to describe adequately many cases of irreversible thermal denaturation (see Plaza del Pino *et al.*<sup>1</sup> and references quoted therein):



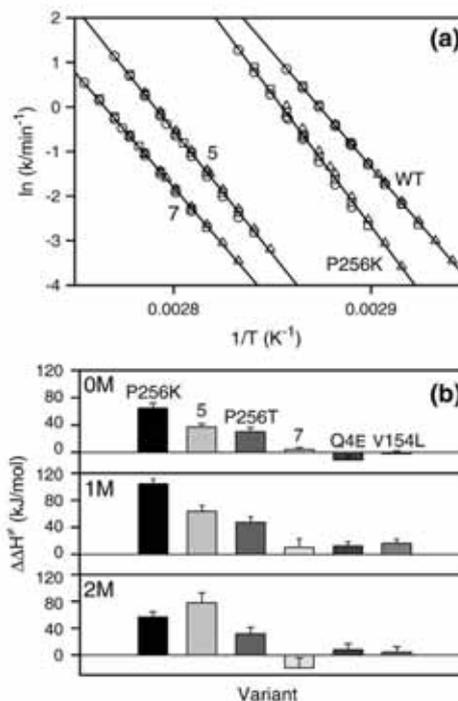
In this model, denaturation is modelled as a first-order kinetic conversion from the native state (N) to a "final" state (F), which is unable to fold back to the native protein, and the rate constant ( $k$ ) for the process is assumed to change strongly with temperature according to the Arrhenius equation:

$$k = \exp \left[ -\frac{E_A}{R} \left( \frac{1}{T} - \frac{1}{T^*} \right) \right] \quad (2)$$

where  $E_A$  is the activation energy and  $T^*$  is the temperature at which  $k=1 \text{ min}^{-1}$ . The fitting procedure is described in some detail in Materials and Methods. Fits were excellent in most cases, despite the fact that the model only uses three parameters to describe the transition (denaturation enthalpy, activation energy and  $T^*$  value; four additional fitting parameters are used to describe the pre- and post-transition baselines). However, in order for the model to be an adequate description of the thermal denaturation, there must be consistency between the fits achieved for the DSC profiles obtained at different scan rates ( $v$ ). We have carried out a very stringent consistency test on the basis of rate constant values calculated from the well-known equation:<sup>20</sup>

$$k = \frac{vC_p^{\text{ex}}}{\Delta H - \langle H \rangle} \quad (3)$$

where  $k$  is obtained at a given temperature from the excess heat capacity and the excess enthalpy ( $C_p^{\text{ex}}$  and  $\langle H \rangle$ ) at that temperature. These excess values are defined with reference to the chemical baseline and can be easily computed once the fitting process has "isolated" the denaturation transition (see Materials and Methods for details). Actually, for each transition  $k$  can be calculated for several temperatures within the transition range to generate Arrhenius plots of  $\ln k$  versus  $1/T$ . In such plots (see illustrative examples in Figure 2) we observe agreement between the rate



**Figure 2.** (a) Illustrative examples of Arrhenius plots for the thermal irreversible denaturation of the lipase forms studied here. The values of the first-order rate constant for irreversible denaturation were derived from the analysis of experimental DSC profiles according to the two-state irreversible model, as described in the text and illustrated in Figure 1. The different symbols correspond to the different scan rates used in the DSC experiments (triangles, 0.5 K/min; squares, 1 K/min; circles, 1.5 K/min). The labels "5" and "7" refer to the fivefold and sevenfold multiple-mutant variants (referred to as SMin041 and SMin149 by Minning *et al.*<sup>19</sup>). Note that the excellent agreement between the rate constant values derived from experiments at different scan rates provides clear support for the applicability of the two-state irreversible model. The data shown correspond to aqueous solution, but similar agreement was obtained for data at 1 M and 2 M urea. The lines are the best fits of the Arrhenius equation (equation (2) in the text). The values of activation energy and  $T^*$  used in subsequent calculations were derived from these Arrhenius fittings. (b) Mutation effects on activation enthalpy for aqueous buffer and the several urea concentrations used here.

constants derived from DSC transitions obtained at different scan rates. Such an agreement constitutes strong evidence for the applicability of the two-state irreversible model, as it implies that the model accounts, not only for the shape of the transitions, but also for the scan-rate effect on them.

**The molecular-dynamics-focused, directed-evolution approach involved in the preparation of the highly stable lipase variants<sup>18</sup> is successful likely because it addresses kinetic stability and, in particular, because heated molecular dynamics simulations possibly identify regions of disrupted native structure in the transition state for irreversible denaturation**

The highly stable lipase variants studied here were obtained (Minning *et al.*<sup>18</sup>) using a directed evolution procedure focused to the protein regions that appear as flexible in heated molecular dynamics simulations. This was certainly supplemented by error-prone PCR, but the stabilizing mutations obtained using this random approach occurred in fact in positions close to (and directly affecting) the regions seen as flexible in the molecular dynamics simulations. Clearly, these regions play an important role in stability. The results reported here regarding the kinetically controlled character of lipase denaturation suggest a plausible explanation for the role of these regions in stability. To the extent that transition-state theory is appropriate, protein kinetic stability can be described in terms of activation changes in free energy, i.e. the difference between the free energies for the transition and native states. In very simple terms, mutations in regions that are native-like in the transition state are not expected to affect the activation free energy. Only mutations in regions of disrupted native structure in the transition state are expected to have the potential to change significantly the activation free energy for the irreversible denaturation process and, consequently, the kinetic stability of the protein. It appears likely then that the molecular dynamics simulations identify such regions in the transition state as flexible regions in the heated simulations. Of course, a verification of this plausible hypothesis will require the determination ( $\phi$ -value analysis) of the transition-state "structure" (work in progress).

**Transition-state analysis of lipase kinetic stability reveals large mutation effects on activation enthalpy and entropy, which cancel each other out to a significant extent in the activation free-energy value**

The Arrhenius description of the kinetic stability of WT lipase and the several variants studied can be easily converted into a transition-state-theory description, as explained in Materials and Methods. This leads to the values of the mutation effects on activation enthalpy, entropy and free energy:  $\Delta\Delta H^\ddagger$ ,  $\Delta\Delta S^\ddagger$  and  $\Delta\Delta G^\ddagger$  (where  $\Delta X^\ddagger = X(\text{transition state}) - X(\text{native state})$  and  $\Delta\Delta X^\ddagger = \Delta X^\ddagger(\text{variant}) - \Delta X^\ddagger(\text{WT})$ ). The values of  $\Delta\Delta H^\ddagger$  for the six lipase variants studied here are shown in Figure 2(b). It is clear that there are very large mutation effects on activation

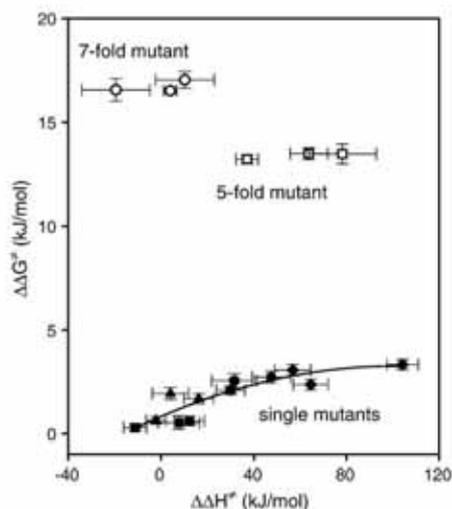
enthalpy<sup>†</sup>, although its urea dependence is rather weak. Actually, these effects are so large that they are apparent in the original DSC thermograms upon visual inspection. For a two-state irreversible model, both the width of the transitions and the scan-rate effect on them are smaller the higher the activation energy value<sup>20</sup> (note that  $E_A = \Delta H^\ddagger$ ; see Materials and Methods). Thus for instance, narrower transitions and smaller scanning-rate effect are clearly observed for the P256K variant when compared with the WT form (see Figure 1), revealing the higher activation energy value in the former case and, consequently, the very large and positive mutation  $\Delta\Delta H^\ddagger$  value. Note also that the mutation effects on  $\Delta H^\ddagger$  are too large to be explained in terms of temperature effects (related with the value of the activation heat capacity changes); thus, the ratios of  $\Delta\Delta H^\ddagger / \Delta T_m$  (with  $\Delta T_m = T_m(\text{variant}) - T_m(\text{WT})$ ) are, in most cases, much higher than even the value of the equilibrium unfolding heat capacity expected for a protein of the size of lipase (about  $16 \text{ kJ K}^{-1} \text{ mol}^{-1}$  from the regression of  $\Delta C_p$  with the number of residues reported by Robertson and Murphy<sup>21</sup>).

It is very interesting that the large mutation effects on activation enthalpy are to a large extent cancelled by the corresponding effects on activation entropy ( $\Delta\Delta G^\ddagger = \Delta\Delta H^\ddagger - T \cdot \Delta\Delta S^\ddagger$ ) and consequently the mutation effects on kinetic stability (as measured by the mutation effect on activation free energy) reflect only weakly the mutation effects on activation energy (enthalpy): see the plot of  $\Delta\Delta G^\ddagger$  versus  $\Delta\Delta H^\ddagger$  of Figure 3. Furthermore, in this plot, the WT and single variant forms show a common behaviour, while different and more favourable enthalpy-entropy compensation is related to the enhanced stability of the multiple-mutant variants; that is, the  $\Delta\Delta H^\ddagger$  values for the multiple-mutant variants are in fact smaller than those for some of the single-mutant variants, but they are compensated to a smaller extent by the  $-T(\Delta\Delta S^\ddagger)$  term.

**Mutation effects on kinetic, urea  $m$  values for lipase irreversible denaturation are small, supporting modest changes in transition-state accessibility to solvent upon mutation**

Changes in protein energetic parameters are usually interpreted in terms of folding-unfolding processes and/or the associated changes in the protein surface accessible to the solvent.<sup>21-23</sup> The large  $\Delta\Delta H^\ddagger$  and  $\Delta\Delta S^\ddagger$  values that we have found would then appear to suggest significant mutation-

<sup>†</sup> It is apparent from the data included in Figures 2 and 3, however, that not all the mutant proteins studied show the same behaviour with respect to the mutation effect on activation enthalpy. Structural interpretations of these differences, however, seem difficult at this stage; in particular because the network of water-unsatisfied, broken internal contacts proposed to exist in the transition state seems to behave in a "cooperative-like" manner (as suggested by the large effects on activation enthalpy found for some of the variants).



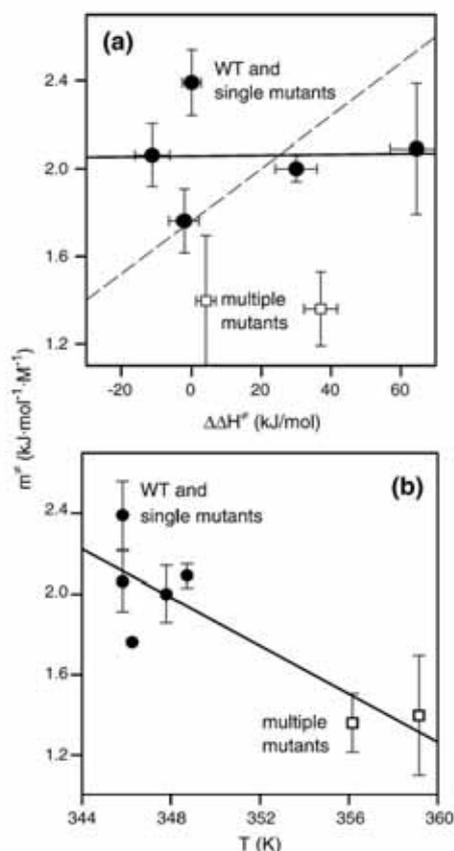
**Figure 3.** Plot of mutation effect on activation free energy at 353 K versus mutation effect on activation enthalpy for all the lipase variants studied here. Mutation effects on activation parameters were calculated as explained in Materials and Methods. For each variant, values corresponding to 0 M, 1 M and 2 M urea are included. Note that, since  $\Delta\Delta G^\ddagger = \Delta\Delta H^\ddagger - T\Delta\Delta S^\ddagger$ , this plot reveals the enthalpy-entropy compensation behaviour (see the text for details).

induced alterations in transition-state, which would change its exposure to solvent upon mutation. That is, according to this interpretation, the transition-state for the irreversible denaturation of lipase would show significant structural plasticity ("changing nature"), a feature that has been often reported for transition states of protein folding/unfolding processes.<sup>24-31</sup> If this interpretation in terms of changes in transition-state accessibility to solvent is correct, we could expect large mutation effects on kinetic  $m$  values for irreversible denaturation:

$$m^\ddagger = - \left( \frac{\partial \Delta G^\ddagger}{\partial [\text{urea}]} \right)_T = RT \cdot \left( \frac{\partial \ln k}{\partial [\text{urea}]} \right)_T \quad (4)$$

since the relation between  $m$  values (the denaturant concentration dependencies of free energy changes) and the corresponding changes in solvent accessibility is well-established.<sup>32</sup> Kinetic  $m^\ddagger$  values for lipase irreversible denaturation could be calculated from the urea dependence of the  $k$  values, as prescribed by equation (4), or, more conveniently, from the urea dependence of the transition temperature and activation energy values, as is explained in Materials and Methods. The resulting  $m^\ddagger$  values for WT lipase and the six variants are plotted versus  $\Delta\Delta H^\ddagger$  in Figure 4(a) and versus temperature in Figure 4(b).

The first point to notice about the calculated  $m^\ddagger$  values is that they are comparatively small. Thus, the  $m$  value for complete unfolding of lipase ( $m_{\text{unt}}$ ) can be estimated to be about  $17 \text{ kJ mol}^{-1} \text{ M}^{-1}$  from the relation of  $m$  with accessible surface area (ASA) change upon unfolding reported by Myers *et al.* (see Figure 2 in Myers *et al.*,<sup>32</sup> the lipase unfolding  $\Delta\text{ASA}$  is estimated to be about  $33,000 \text{ \AA}^2$  from standard ASA calculations on the native structure<sup>19</sup> and tripeptide models for the unfolded state). This gives values for the  $m^\ddagger/m_{\text{unt}}$  ratio smaller than 0.1, suggesting a native-like transition-state, at least in terms of exposure to solvent.



**Figure 4.** Kinetic urea  $m$  values ( $m^\ddagger$  values) for the thermal irreversible denaturation of WT lipase and the six variants studied here versus the mutation effect on activation enthalpy in aqueous solution (a) and versus temperature (average of the  $T_m$  values in aqueous solution for each variant). The broken line in (a) is the dependency expected if the mutation effects on activation enthalpy are attributed to changes in solvent accessibility (see the text for details). The continuous line in (a) is the linear, least-squares fit to all the single mutant data. The continuous line in (b) is the linear, least-squares fit to all data.

A second point to notice is that there is no clear correlation between  $m^{\ddagger}$  values and  $\Delta\Delta H^{\ddagger}$  values (Figure 4(a)). In fact, for the single-mutant variants, the  $m^{\ddagger}$  values cluster around a value of  $2 \text{ kJ mol}^{-1} \text{ M}^{-1}$  with no noticeable  $\Delta\Delta H^{\ddagger}$  dependence (Figure 4(a)). This is a shocking result if the mutation  $\Delta\Delta H^{\ddagger}$  values are interpreted in terms of changes in solvent accessibility<sup>‡</sup>, as the following illustrative calculations highlight. From the correlations between unfolding energetic parameters for globular proteins and ASA changes reported by Robertson and Murphy, an increase in enthalpy of  $39 \text{ J per } \text{Å}^2$  of exposed surface at  $347 \text{ K}$  can be easily deduced. Myers *et al.*<sup>32</sup> reported an increase in urea  $m$  values for globular protein unfolding of  $0.46 \text{ J mol}^{-1} \text{ M}^{-1}$  per  $\text{Å}^2$  of exposed surface. The two figures can be combined to give an expected value of  $0.012 \text{ mol}^{-1} \text{ M}^{-1}$  for the slope of a plot of  $m$  versus enthalpy changes. However, the experimental slope (essentially zero) of the plot of  $m^{\ddagger}$  versus  $\Delta\Delta H^{\ddagger}$  for the single-mutant variants is much smaller than that value (compare the continuous and broken lines in Figure 4(a)).

A third point to notice is that, although the  $m^{\ddagger}$  values for the multiple-mutant variants are significantly smaller than those for the WT and single-mutant ones, this appears to be a temperature effect associated with the higher  $T_m$  values for the former variants. Thus, a plot of  $m^{\ddagger}$  versus temperature including all the variants (Figure 4(b)) has a slope of  $-0.06(\pm 0.02) \text{ kJ mol}^{-1} \text{ M}^{-1} \text{ K}^{-1}$ , which agrees with the temperature dependence of the  $m^{\ddagger}$  values for the individual variants as expected from the linkage relationship:<sup>26</sup>

$$\left(\frac{\partial m^{\ddagger}}{\partial T}\right)_{[\text{urea}]} = \left(\frac{\partial \Delta S^{\ddagger}}{\partial [\text{urea}]}\right)_T = \frac{1}{T} \left(\frac{\partial \Delta H^{\ddagger}}{\partial [\text{urea}]}\right)_T - \frac{m^{\ddagger}}{T} \equiv \frac{1}{T} \left(\frac{\partial E_A}{\partial [\text{urea}]}\right)_T \quad (5)$$

where in the second equality we have used that  $\Delta G^{\ddagger} = \Delta H^{\ddagger} - T\Delta S^{\ddagger}$  and equation (4), and, in the third equality, we have used that  $\Delta H^{\ddagger} = E_A$  and that  $m^{\ddagger}/T$  is found to be negligible with respect to the  $(1/T) \cdot (\partial \Delta H^{\ddagger} / \partial [\text{urea}])$  term. Note that the linkage relationship (first equality in the equation) is actually a Maxwell thermodynamic relationship, i.e. a thermodynamic truism with no molecular or

mechanistic implications.<sup>‡</sup> In fact, equation (5) only indicates that the temperature dependence of  $m^{\ddagger}$  for a given variant can be calculated from the corresponding urea-concentration dependence of the activation energy values. The average of the calculated values for all the lipase forms is  $-0.07(\pm 0.03) \text{ kJ mol}^{-1} \text{ M}^{-1} \text{ K}^{-1}$ , in good agreement with the slope of the plot in Figure 4(b), which argues against effects of the studied mutations on  $m^{\ddagger}$  values. Of course, the decrease in  $m^{\ddagger}$  with temperature does suggest some degree of transition state plasticity. Interestingly, changes in kinetic  $m^{\ddagger}$  values with temperature have been reported for the unfolding processes of several proteins<sup>25,26</sup> and interpreted in terms of Hammond behaviour.<sup>24</sup>

#### A solvation/desolvation barrier likely contributes to the kinetic stability of lipase from *T. lanuginosa*

From all the above, we are led to conclude that the large observed mutation effects on activation enthalpy (and entropy) are associated to some structural feature of the transition state that is not linked to changes in overall solvent accessibility. Interestingly, several computational studies<sup>33-35</sup> have strongly suggested the possible existence of solvation/desolvation barriers in at least some protein folding/unfolding processes. In simple terms, such barriers would arise from an asynchrony between water escape and the formation of internal contacts in the folding process or, if viewed from a protein unfolding perspective, from the asynchrony between the breaking of internal contacts and water penetration. As a result, the transition state would be characterized by a network (or networks) of "partially broken" and "water-unsatisfied" internal contacts (see Figure 9 of Liu & Chan<sup>35</sup> for a pictorial illustration), which will necessarily imply a higher energy and will contribute therefore to the activation energy (enthalpy) of the process. In fact, it has been recently suggested<sup>35</sup> that solvation/desolvation is the likely origin of the intrinsic enthalpy barriers revealed by Arrhenius analysis of folding rates under constant native-stability conditions.<sup>35-37</sup>

We propose that a similar situation holds for the transition state of the irreversible thermal denaturation of lipase. That is, we suggest the existence in such transition state of a network (or networks) of partially broken, water-unsatisfied internal contacts. Such a network may be large, extended (since the

‡ One of the reviewers of this article has pointed out that there is nothing shocking with the lack of correlation between  $m^{\ddagger}$  and  $\Delta\Delta H^{\ddagger}$ , since the nature of the transition state is not known *a priori* and there is no reason to expect that the transition state is balanced energetically the same way as the native protein. Actually, we do agree with the reviewer: the lack of correlation is only shocking if the same energetic balance is assumed for the native and transition states. In fact, the analyses and discussion given in the text are aimed at suggesting a likely explanation for the different energetic balance in the transition state; i.e. the existence (associated to the solvation barrier) of a network of water-unsatisfied, broken internal contacts.

§ The linkage relationship is easily derived from the equality of the second cross derivatives of the activation free energy:  $(\partial^2 \Delta G^{\ddagger} / \partial T \partial [\text{urea}]) = (\partial^2 \Delta G^{\ddagger} / \partial [\text{urea}] \partial T)$ , an equality that essentially means that the order of derivation is inconsequential. Since the first derivatives are  $(\partial \Delta G^{\ddagger} / \partial T)_{[\text{urea}]} = -\Delta S^{\ddagger}$  and  $(\partial \Delta G^{\ddagger} / \partial [\text{urea}])_T = -m^{\ddagger}$  (equation (4)), the linkage equation follows immediately as a model-independent result.

transition state appears to be native-like) and able to sense and be altered (reduced, extended) by certain mutations, even if they are introduced at surface positions. Consequently, such mutations will cause significant changes in activation enthalpy but, also, in activation entropy, since solvation/desolvation barriers may have entropic components;<sup>35</sup> in addition, we may expect entropic contributions from the changes in side-chain conformational entropy<sup>38</sup> of the internal groups affected as well as from the possible non-constancy of the front factor in the Eyring equation<sup>39</sup> (note that deviations from the constant- $\tau$  assumption will appear as contributions to the activation entropy value; see equation (9) in Materials and Methods). Liu & Chan<sup>35</sup> have emphasized the role of enthalpy/entropy compensation in determining the free-energy height of solvation/desolvation barriers. Indeed, differences in the enthalpy/entropy compensation pattern are related to the enhanced kinetic stability of the multiple-mutant variants over that for the simple mutants (Figure 3). It must be noted, nevertheless, that mutations that mainly act by affecting unsatisfied internal interactions are not likely to cause changes in overall exposure to solvent, as measured by the  $m$  value, in agreement with our experimental results (Figure 4).

It is perhaps worth pointing out here that the general existence of water-solvation barriers that contribute to protein kinetic stability seems to be consistent with the well-known fact that dehydrated and lyophilized enzymes denature (irreversibly) at very high temperatures, indicating very high kinetic stability (see Nuñez-Olea & Sanchez-Ruiz<sup>40</sup> and references quoted therein). Finally, a general feature suggested by this work is the similarity between the transition states for protein unfolding and for protein irreversible denaturation. This is actually consistent with theoretical analyses of simple Lumry-Eyring models,<sup>1,20</sup> which show that, when irreversible alterations of non-native states occur very fast, the rate of irreversible denaturation is given by the rate of unfolding.

## Materials and Methods

### Protein expression and purification

Details on protein expression and purification are given by Minning *et al.*<sup>18</sup>

### Differential scanning calorimetry

DSC experiments were carried out using a VP-DSC calorimeter from MicroCal (Northampton, USA). Protein solutions for calorimetric experiments were prepared by exhaustive dialysis against the buffer (50 mM sodium acetate, pH 5). For experiments in the presence of urea the adequate amount of a stock urea solution in the same buffer was added. Protein concentrations were determined from the absorbance at 280 nm and urea concentrations from refraction index measurements. DSC

experiments were performed at different scan rates within the approximate 0.5–1.5 K/min range (in a few cases, additional experiments at 0.25 K/min were carried out); the actual scan rates of the experiments were calculated from the experimental temperature/time data, although they differ little from the nominal values. Additional details about the DSC experiments have been previously described.<sup>41</sup>

### Fittings of the two-state irreversible model

Fittings of the two-state irreversible model were carried out as follows. In the case of the lipase variants studied here, the process depicted in equation (1) would be followed by aggregation of F:  $N \rightarrow F \rightarrow \text{aggregated}$ , where the  $F \rightarrow \text{aggregated}$  step is likely responsible for the post-transition drop in calorimetric signal (note that F could already be an aggregated form of the protein which undergoes further aggregation). Our approach has been to include the effect of the aggregation step in the chemical baseline, so that we only need to model explicitly the  $N \rightarrow F$  process. Accordingly, we write the calorimetric signal as:

$$C_p^{\text{APP}} = C_p^{\text{PRE}} + (C_p^{\text{POST}} - C_p^{\text{PRE}})(1 - X_N) - \Delta H \cdot \left( \frac{dX_N}{dT} \right) \quad (6)$$

where  $X_N$  is the mole fraction of native state,  $\Delta H$  is the denaturation enthalpy (in this case the enthalpy change for the N to F transition) and  $C_p^{\text{PRE}}$  and  $C_p^{\text{POST}}$  are the pre- and post-transition baselines which, in our analysis, are taken to be linear functions of temperature. The slope of the post-transition baseline will be large and negative, reflecting the effect of aggregation. Note that the two first terms in the right-hand-side of equation (6) represent a smooth line that connects the pre- and post-transition heat capacity levels (i.e. the so-called chemical baseline; see broken lines in Figure 1(a) and (b)), while the last term describes the calorimetric transition (i.e., the "peak"). Note also that heat capacity on the left-hand-side of equation (6) is labelled "apparent" (superscript "APP") since, in this case (equation (1)), it reflects a kinetic process and, consequently, it is not a true thermodynamic property. Kinetic analysis of the model described in equation (1), as we have previously described,<sup>20</sup> leads to the following equations for  $X_N$  and  $dX_N/dT$ :

$$X_N = \exp \left[ -\exp \left( \frac{E_A \cdot \Delta T}{RT_m^2} \right) \right] \quad (7)$$

$$\frac{dX_N}{dT} = -\frac{E_A}{RT_m^2} \exp \left( \frac{E_A \cdot \Delta T}{RT_m^2} \right) \exp \left[ -\exp \left( \frac{E_A \cdot \Delta T}{RT_m^2} \right) \right] \quad (8)$$

where  $\Delta T = T - T_m$ . Each individual experimental DSC profile was fitted with equation (6) together with equations (7) and (8). Fittings were excellent in most cases, as is shown by the illustrative examples in Figure 1(a) and (b). Note that the excess enthalpy at any given temperature can be calculated as  $\langle H \rangle = (1 - X_N)(\Delta H)$  and that the value of  $\Delta H$  is obtained from the fitting. Therefore, once the fitting based on equation (6) has been carried out, calculation of the first-order rate constants for irreversible denaturation according to equation (3) is straightforward.

### Calculation of the mutation effects on transition-state energetics

The mutation effects on activation energetic parameters were calculated from the Arrhenius parameters ( $T^*$  and activation energy, see equation (2)) on the basis of transition-state theory using the following procedure. Within the framework of transition-state theory, the

Pre-transitional baselines were clearly linear in a very wide temperature range (from the lowest temperature of the scanning experiment, about 20 degrees, to the onset of the transition) and, therefore, posed no problems for chemical baseline determination. Although the low protein concentrations used here (about 0.1 mg/ml) preclude the accurate determination of absolute native heat capacities from the experimental calorimetric profiles, we found the experimental values to be reasonably close to those expected from the correlation with molecular mass suggested by Freire and co-workers<sup>44</sup> ( $28(\pm 2)$  kJ K<sup>-1</sup> mol<sup>-1</sup> from the average of the  $C_p$  values (at 315 K and in the absence of urea) for all the proteins studied here versus  $36(\pm 2)$  kJ K<sup>-1</sup> mol<sup>-1</sup> predicted for 315 K by the molecular mass correlation). Such qualitative agreement was also found to hold for the slopes of the native heat capacity function ( $0.2(\pm 0.1)$  kJ K<sup>-2</sup> mol<sup>-1</sup> from the average of the  $dC_p/dT$  values for all the proteins studied here versus  $0.14\pm 0.2$  kJ K<sup>-2</sup> mol<sup>-1</sup> predicted by the molecular mass correlation). On the other hand, post-transitional baselines appeared approximately linear (although with a large negative slope) only within a temperature range of a few degrees immediately above the transition (see Figure 1). This linear range was followed by a more pronounced and clearly non-linear drop in calorimetric signal. Attempts to include the non-linear drops in chemical baseline tracing were not satisfactory because: (a) no simple description of the post-transitional baseline (such as a second-order polynomial) was found to be adequate over a wide temperature range and, consequently, the fitting algorithm usually produced poor fittings in the transition region in order to achieve acceptable fittings of the non-linear drop in the post-transitional baseline; (b) using a second-order (or higher-order) polynomial to represent the post-transitional baseline was often found to produce large (and physically unrealistic) "undulations" in the chemical baseline within the transition temperature range. This is in contrast with the simpler procedure that we describe in the text and use throughout this work (including as post-transitional baseline only a short region above the transition and using a linear function to describe the calorimetric data in that region), which yields good fits and physically realistic chemical baselines that smoothly connect the pre- and post-transition heat capacity levels. Clearly, the procedure that we have used to define the chemical baseline is reasonable, although, to some extent, empirical. Therefore, some uncertainty is associated to the chemical baselines thus obtained. However, the main results of the analysis are robust with respect to chemical baseline tracing, as shown by the excellent agreement between the rate constants derived from the fittings of different calorimetric profiles corresponding to different scan rates (see Figure 2). Note that the comparison between the parameters derived from several profiles involves several, independent chemical-baseline determinations.

rate constant for irreversible protein denaturation is written as:

$$k = \tau \cdot \exp\left(-\frac{\Delta G^\ddagger}{RT}\right) = \tau \cdot \exp\left(\frac{\Delta S^\ddagger}{R}\right) \cdot \exp\left(-\frac{\Delta H^\ddagger}{RT}\right) \quad (9)$$

where  $\Delta G^\ddagger$ ,  $\Delta H^\ddagger$  and  $\Delta S^\ddagger$  stand for the activation free-energy, enthalpy and entropy. Strictly,  $\Delta H^\ddagger$  and  $\Delta S^\ddagger$  depend on temperature (according to the activation heat capacity value). However, such dependence cannot be observed within the quite narrow temperature range of the DSC transition (the temperature range for which experimental  $k$  values are available); we, therefore, take  $\Delta H^\ddagger$  and  $\Delta S^\ddagger$  as constants in this analysis.  $\tau$  in equation (9) is the front factor, i.e. the (hypothetical) value of  $k$  corresponding to  $\Delta G^\ddagger=0$ . There is considerable body of work regarding the value of  $\tau$  for folding/unfolding processes (see, for instance, Akmal & Muñoz<sup>45</sup> and references quoted therein). For the calculation purposes of interest here, however, we simply assume that  $\tau$  is temperature-independent (again, within the narrow temperature range experimentally accessible) and, further, that it has the same value for the seven lipase variants studied (see, however, the text for the implications of a non-constant front factor). Constancy of  $\tau$ ,  $\Delta H^\ddagger$  and  $\Delta S^\ddagger$  is consistent with the linear experimental Arrhenius plots (Figure 2(a)) and implies (compare equations (2) and (9)) that the activation enthalpy is equal to  $E_A$  and, hence, that mutation effects on activation enthalpy can be calculated as:

$$\Delta\Delta H^\ddagger = E_A(\text{variant}) - E_A(\text{WT}) \quad (10)$$

The assumption that the value of  $\tau$  is the same for the WT lipase and the six variants studied implies that mutation effects on activation entropy can be easily obtained from the  $T^*$  and  $E_A$  values using:

$$\Delta\Delta S^\ddagger = \frac{E_A}{T^*}(\text{variant}) - \frac{E_A}{T^*}(\text{WT}) \quad (11)$$

and the mutation effects on activation free-energy from the rate constants according to:

$$\Delta\Delta G^\ddagger = -RT \ln\left(\frac{k(\text{variant})}{k(\text{WT})}\right) \quad (12)$$

where the two  $k$  values must correspond to the temperature at which  $\Delta\Delta G^\ddagger$  is calculated.

### The calculation of kinetic urea $m$ values

The  $m^{\ddagger}$  values were calculated from the urea concentration dependence of the temperature of the maximum of the DSC transitions ( $T_m$ ) and the activation energy using a simple equation, which can be easily derived as follows. From equations (7) and (8) it is straightforward to deduce that, for  $T=T_m$ , the fraction of native protein and its temperature dependence are given by  $X_N=1/e$  and  $dX_N/dT = -E_A/(eRT_m^2)$  (with  $e=2.718...$ ). Also, the excess heat capacity at  $T_m$  would be given by  $C_p^{\ddagger} = -\Delta H \cdot (dX_N/dT) = E_A \cdot \Delta H / (eRT_m^2)$ . Now, since  $X_N$  is related to the DSC transition through  $X_N = (\Delta H - \langle H \rangle) / \Delta H$ , application of equation (3) to the temperature of the maximum leads to following expression for the rate constant at  $T_m$ :

$$k_m = \frac{v \cdot E_A}{RT_m^2} \quad (13)$$

The value of  $k_m$  will change with urea concentration due to the "direct" urea effect on  $k$  and also due to the "indirect" temperature effect (since  $T_m$  does change with urea concentration). Using the known rules of partial differentiation,<sup>43</sup> the total change of  $k_m$  with [urea] can be written as:

$$\frac{d \ln k_m}{d[\text{urea}]} = \left( \frac{\partial \ln k}{\partial T} \right)_{\text{obs}} \cdot \left( \frac{dT_m}{d[\text{urea}]} \right) + \left( \frac{\partial \ln k}{\partial [\text{urea}]} \right)_T \quad (14)$$

The partial derivative of  $\ln k$  with respect to [urea] is related to  $m^{\ddagger}$  through  $(\partial \ln k / \partial [\text{urea}])_T = m^{\ddagger} / (RT)$  (equation (4)), the partial derivative with respect to temperature is  $E_A / RT_m^2$  (Arrhenius equation) and  $k_m$  is given by equation (13). Equation (14) can then be solved for  $m^{\ddagger}$  to yield:

$$m^{\ddagger} = - \frac{E_A}{T_m} \left( \frac{dT_m}{d[\text{urea}]} \right) - RT_m \cdot \left( \frac{d \ln(E_A / RT_m^2)}{d[\text{urea}]} \right) \quad (15)$$

The derivatives involved in equation (15) are available from the experimental urea-concentration dependencies of  $T_m$  and  $E_A$  (for given scan-rates). Thus, the plots of  $T_m$  versus [urea] are highly linear, leading to accurate estimations of  $dT/d[\text{urea}]$ . Plots of  $\ln(E_A / RT_m^2)$  versus [urea] show some curvature, implying some uncertainty in the evaluation of the slope  $d \ln(E_A / RT_m^2) / d[\text{urea}]$ ; this is immaterial for  $m^{\ddagger}$  calculation, however, since the second-term in the right-hand-side of equation (15) is negligible when compared with the first-one. Once the values of the slopes are known,  $m^{\ddagger}$  values in the limit of low urea concentration are obtained from equation (15) using the  $E_A$  and  $T_m$  values in the absence of urea. For each protein system, the above calculation can be performed using the data corresponding to three scan-rates employed in this work, leading to three  $m^{\ddagger}$  values. The averages and the associated standard errors are shown in Figure 4. These associated errors are low, indicating good agreement between independent calculations of  $m^{\ddagger}$  and supporting the robustness of the calculation procedure based on equation (15). It is perhaps worth pointing out that the  $m^{\ddagger}$  values are essentially given by the urea concentration effect on  $T_m$  and that  $T_m$  values are accurately determined from the calorimetric profiles.

## Acknowledgements

We acknowledge financial support from the Spanish Ministry of Education and Science (Grant BIO2003-02229) and Feder Funds. We thank Dr Beatriz Ibarra-Molero and Dr Maria del Mar Garcia-Mira for critical readings of the manuscript.

## References

- Plaza del Pino, I. M., Ibarra-Molero, B. & Sanchez-Ruiz, J. M. (2000). Lower kinetic limit to protein thermal stability: a proposal regarding protein stability in vivo and its relation with misfolding diseases. *Proteins: Struct. Funct. Genet.* **40**, 58–70.
- Baker, D. & Agard, D. A. (1994). Kinetics versus thermodynamics in protein folding. *Biochemistry*, **33**, 7505–7509.
- Jaswal, S. S., Truhlar, S. M., Dill, K. A. & Agard, D. A. (2005). Comprehensive analysis of protein folding activation thermodynamics reveals a universal behaviour violated by kinetically stable proteases. *J. Mol. Biol.* **347**, 355–366.
- Schnyrov, V. L., Zhadan, G. G., Cobaleda, C., Muñoz-Barroso, I. & Villar, E. (1997). A differential scanning calorimetric study of Newcastle disease virus: identification of proteins involved in thermal transitions. *Arch. Biochem. Biophys.* **341**, 89–97.
- Sohl, J. L., Jaswal, S. S. & Agard, D. A. (1998). Unfolded conformations of alpha-lytic protease are more stable than its native state. *Nature*, **395**, 817–819.
- Jaswal, S. S., Sohl, J. L., Dans, J. H. & Agard, D. A. (2002). Energetic landscape of alpha-lytic protease optimizes longevity through kinetic stability. *Nature*, **415**, 343–346.
- Leikina, E., Merits, M. V., Kuznetsova, N. & Leikin, S. (2002). Type I collagen is thermally unstable at body temperature. *Proc. Natl Acad. Sci. USA*, **99**, 1314–1318.
- Persikov, A. V. & Brodsky, B. (2002). Unstable molecules form stable tissues. *Proc. Natl Acad. Sci. USA*, **99**, 1101–1103.
- Lynch, S. M., Boswell, S. A. & Colon, W. (2004). Kinetic stability of Cu/Zn superoxide dismutase is dependent on its metal ligands: implications for ALS. *Biochemistry*, **43**, 16525–16531.
- Manning, M. & Colon, W. (2004). Structural basis of protein kinetic stability: resistance to sodium dodecyl sulfate suggests a central role for rigidity and a bias toward beta-sheet structure. *Biochemistry*, **43**, 11248–11254.
- Forrer, P., Chang, C., Ott, D., Wlodawer, A. & Plückthun, A. (2004). Kinetic stability and crystal structure of the viral capsid protein SVP. *J. Mol. Biol.* **344**, 179–193.
- Jayaraman, S., Gantz, D. & Gursky, O. (2005). Structural basis for thermal stability of human low-density lipoprotein. *Biochemistry*, **44**, 3965–3971.
- Hammarström, P., Wiseman, R. L., Powers, E. T. & Kelly, J. W. (2003). Prevention of transthyretin amyloid disease by changing protein misfolding energetics. *Science*, **299**, 713–716.
- Petrassi, M. M., Johnson, S. M., Purkey, H. E., Chiang, K. P., Walkup, T., Jinag, X. *et al.* (2005). Potent and selective structure-based dibenzofuran inhibitors of transthyretin amyloidogenesis: kinetic stabilization of the native state. *J. Am. Chem. Soc.* **127**, 6662–6671.
- Wiseman, R. L., Johnson, S. M., Kelker, M. S., Foss, T., Wilson, I. A. & Kelly, J. W. (2005). Kinetic stabilization of an oligomeric protein by a single ligand binding event. *J. Am. Chem. Soc.* **127**, 5540–5541.
- Kirk, O., Borchert, T. V. & Fuglsang, C. C. (2002). Industrial enzyme applications. *Curr. Opin. Biotechnol.* **13**, 345–351.
- Sanchez-Ruiz, J. M. & Makhatazde, G. I. (2001). To charge or not to charge. *Trends Biotechnol.* **19**, 132–135.
- Minning, S., Glad, S.O., De Maria, L. & Borch, K. (2006). Protein stabilization by molecular-dynamics-focused directed evolution. in preparation
- Brozowski, A. M., Savage, H., Verma, C. S., Turkenburg, J. P., Lawson, D. M., Svendsen, A. & Patkar, S. (2000). Structural origins of the interfacial activation in *Thermomyces (humicola) lanuginosa* lipase. *Biochemistry*, **39**, 15071–15082.
- Sanchez-Ruiz, J. M. (1992). Theoretical analysis of Lumry-Eyring models in differential scanning calorimetry. *Biophys. J.* **61**, 921–935.
- Roberson, A. D. & Murphy, K. P. (1997). Protein

- structure and energetics of protein stability. *Chem. Rev.* **97**, 1251–1267.
22. Luque, I. & Freire, E. (1998). Structure-based prediction of binding affinities and molecular design of peptide ligands. *Methods Enzymol.* **295**, 100–127.
  23. Spolar, S. R. & Record, M. T. Jr (1994). Coupling of local folding to site-specific binding of proteins to DNA. *Science*, **263**, 777–784.
  24. Matouschek, A. & Fersht, A. R. (1993). Application of physical organic chemistry to engineered mutants of proteins: Hammond postulate behaviour in the transition state of protein folding. *Proc. Natl Acad. Sci. USA*, **90**, 7814–7818.
  25. Matouschek, A., Otzen, D. E., Itzhaki, L. S., Jackson, S. E. & Fersht, A. R. (1995). Movement of the position of the transition state in protein folding. *Biochemistry*, **34**, 13656–13662.
  26. Ibarra-Molero, B. & Sanchez-Ruiz, J. M. (1996). A model-independent, nonlinear extrapolation procedure for the characterization of protein folding energetics from solvent-denaturation data. *Biochemistry*, **35**, 14689–14702.
  27. Oliveberg, M., Tan, Y.-J., Silow, M. & Fersht, A. R. (1998). The changing nature of the protein folding transition state: implications for the shape of the free-energy profile for folding. *J. Mol. Biol.* **277**, 933–943.
  28. Oliveberg, M. (1998). Alternative explanations for “multistate” kinetics in protein folding: transient aggregation and changing transition-state ensembles. *Acc. Chem. Res.* **31**, 765–772.
  29. Lindberg, M., Tangrot, J. & Oliveberg, M. (2002). Complete change of the protein folding transition state upon circular permutation. *Nature Struct. Biol.* **9**, 818–822.
  30. Muñoz, V. (2002). Folding plasticity. *Nature Struct. Biol.* **9**, 792–794.
  31. Shen, T., Hofmann, C. P., Oliveberg, M. & Wolynes, P. G. (2005). Scanning malleable transition state ensembles: comparing theory and experiment for folding protein U1A. *Biochemistry*, **44**, 6433–6439.
  32. Myers, J. K., Pace, C. N. & Scholtz, J. M. (1995). Denaturant *m* values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4**, 2138–2148.
  33. Rank, J. A. & Baker, D. (1997). A desolvation barrier to hydrophobic cluster formation may contribute to the rate-limiting step in protein folding. *Protein Sci.* **6**, 347–354.
  34. Cheung, S. M., Garcia, A. E. & Onuchic, J. N. (2002). Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc. Natl Acad. Sci. USA*, **99**, 685–690.
  35. Liu, Z. & Chan, H. S. (2005). Desolvation is a likely origin of robust enthalpic barriers to protein folding. *J. Mol. Biol.* **349**, 872–889.
  36. Scalley, M. L. & Baker, D. (1997). Protein folding kinetics exhibit an Arrhenius temperature dependence when corrected for the temperature-dependence of protein stability. *Proc. Natl Acad. Sci. USA*, **94**, 10636–10640.
  37. Kuhlman, B., Luisi, D. L., Evans, P. A. & Raleigh, D. P. (1998). Global analysis of the effects of temperature and denaturant on the folding and unfolding kinetics of the N-terminal domain of the protein L9. *J. Mol. Biol.* **284**, 1661–16670.
  38. D’Aquino, J. A., Gomez, J., Hilser, V. J., Lee, K. H., Amzel, L. M. & Freire, E. (1996). The magnitude of the backbone conformational entropy change in protein folding. *Proteins: Struct. Funct. Genet.* **25**, 143–156.
  39. Kaya, H. & Chan, H. S. (2003). Solvation effects and driving forces for protein thermodynamic and kinetic cooperativity: how adequate is native-centric topological modeling? *J. Mol. Biol.* **326**, 911–931.
  40. Nuñez-Olea, J. & Sanchez-Ruiz, J. M. (1994). Self-assembly and protein stability. *Nature*, **370**, 105.
  41. Godoy-Ruiz, R., Perez-Jimenez, R., Ibarra-Molero, B. & Sanchez-Ruiz, J. M. (2004). Relation between protein stability, evolution and structure, as probed by carboxylic acid mutations. *J. Mol. Biol.* **336**, 313–318.
  42. Akmal, A. & Muñoz, V. (2004). The nature of the free energy barriers to two-state folding. *Proteins: Struct. Funct. Genet.* **57**, 142–152.
  43. Blinder, S. M. (1966). Mathematical methods in elementary thermodynamics. *J. Chem. Educ.* **43**, 85–92.
  44. Freire, E. (1995). Differential scanning calorimetry. In *Protein Stability and Folding. Theory and Practice* (Shirley, B. A., ed), pp. 191–218. Humana Press, Totowa, New Jersey.

Edited by F. Schmid

(Received 3 March 2006; received in revised form 1 May 2006; accepted 3 May 2006)  
Available online 19 May 2006

## SEGUNDO ARTÍCULO

"Energetic and structural consequences of desolvation/solvation barriers to protein folding/unfolding assessed from experimental unfolding rates"

Rodriguez-Larrea D, Ibarra-Molero B, Sanchez-Ruiz

JM.

*Biophysical Journal.* 2006 Sep 1;91(5):L48-50.

**INTRODUCCIÓN ARTÍCULO 2º**

La continuación natural del trabajo, presentado en esta Memoria, sobre las lipasas de *Thermomyces lanuginosa* (Rodríguez-Larrea, Minning et al. 2006) era tratar de incluir la barrera de solvatación en una imagen, principalmente energética, del estado de transición. Recordemos que éstas se determinaban experimentalmente en el trabajo anterior a partir de una discrepancia entre los valores  $m$ -cinéticos y las entalpías de activación (también denominadas a lo largo de este trabajo energías de activación).

Es necesario explicar, para clarificar los resultados, la naturaleza de esta discrepancia: Tanto los valores  $m$  como las entalpías correlacionan en general aceptablemente bien con los cambios de superficie expuesta al solvente (Robertson and Murphy 1997). Por tanto una alta energía de activación debería implicar un gran cambio en la superficie accesible al solvente. Sin embargo, los valores  $m$  (que dependen directamente de la superficie expuesta al solvente) indicaban que la superficie que se exponía de nuevo en el estado de transición era muy pequeña.

Por tanto, dos estimadores del cambio de área accesible al solvente arrojaban resultados dispares. Esta diferencia se le atribuyó a la existencia de barreras de solvatación, es decir, a la imposibilidad de que el solvente compense en el estado de transición algunas de las interacciones que han sido rotas en el estado de transición: Teníamos una evidencia energética. En el presente trabajo intentamos una descripción física, aunque sea a "baja resolución", y realizamos una estimación de la importancia de las barreras de solvatación en el universo de las proteínas.

Partimos de una entalpía de activación experimental (obtenidos a través de la bibliografía) que obtenemos a partir de experimentos de desnaturalización realizados a distintas temperaturas (un Arrhenius), de valores  $m$ -cinéticos experimentales y de la estructura de la proteína:

- La energía de activación experimental la dividiremos en dos componentes: una entalpía de hidratación y una entalpía de solvatación.

- Los valores  $m$  experimentales nos darán el área que se expone nuevamente en el estado de transición y se hidrata.

- La estructura de la proteína nos servirá para estimar el cambio de área accesible al solvente entre la estructura nativa y la proteína completamente desplegada.

Una vez tenemos esta información el primer paso es conocer la cantidad de área que se ha expuesto en el estado de transición. Para ello nos basamos en la relación entre el  $m$ -cinético y el  $m$  de equilibrio. Esto nos permite obtener un cambio de área accesible al solvente en el estado de transición y a partir de este una energía de activación (usando las correlaciones entre cambios en área accesible y cambios de entalpía).

Dado que disponemos de la energía de activación experimental y de una estimación de la misma a partir del cambio de área accesible expuesta en el estado de transición, consideraremos que la diferencia entre ambas está causada por la rotura de interacciones entre residuos de la proteína. Interacciones perdidas que aún no han sido compensadas por unas nuevas con el solvente. Por tanto, tenemos una entalpía atribuible a las barreras de solvatación.

A partir de este punto se tomaron las correlaciones publicadas por el Dr. E.Freire y colaboradores (Hilser, Gomez et al. 1996). En ellas se relacionan los cambios en entalpía con las superficies de interacción entre residuos. Esto permitiría, en cierto grado, realizar una estimación de la cantidad de área del estado de transición que había perdido todas sus interacciones y no había establecido aún interacciones con el solvente.

La figura final que se obtiene, una vez realizada la aproximación para un conjunto de 6 proteínas, es que la asincronía entre la rotura de las interacciones entre residuos y la formación de interacciones nuevas con el solvente genera barreras de solvatación que son especialmente importantes en el balance energético del estado de transición. Por otro lado y a nivel estructural el estado de transición presentaría, aproximadamente, el 50% del área que aún no ha sido expuesta al solvente carente de interacciones. En definitiva, proponemos una imagen estructural del estado de transición que explica algunas de las diferencias en los balances energéticos de los estados nativos y los estados de transición.

## Energetic and Structural Consequences of Desolvation/Solvation Barriers to Protein Folding/Unfolding Assessed from Experimental Unfolding Rates

David Rodríguez-Larrea, Beatriz Ibarra-Molero, and Jose M. Sanchez-Ruiz

Facultad de Ciencias, Departamento de Química Física, 18071-Granada, Spain

**ABSTRACT** Theoretical work has suggested the existence of solvation/desolvation barriers in protein folding/unfolding processes. We propose that the energetic and structural consequences of such barriers for the folding transition state can be assessed from experimental unfolding rates using well-established structure-energetics relationships. For a set of proteins of size within the 60–130 number-of-residues range, we find energetic effects associated to solvation/desolvation on the order of  $10^2$  kJ/mol. This supports that the folding transition states may be characterized by large networks of water-unsatisfied, broken internal contacts. In terms of buried surface, we estimate the typical network size to be on the order of several thousands of  $\text{\AA}^2$ , or  $\sim 50\%$  of the total change in accessible surface area upon unfolding. The analyses reported here thus suggest a clear structural picture for the different energetic balance of native and folding transition states.

Received for publication 27 April 2006 and in final form 19 June 2006.

Address reprint requests and inquiries to Jose M. Sanchez-Ruiz, Tel.: 34-958-243-189; E-mail: sanchezr@ugr.es.

Theoretical work has suggested the existence of enthalpic (energetic) desolvation barriers in at least some protein folding processes (1,2). In simple terms, such barriers are associated to the asynchrony between water escape and formation of internal interactions. Accordingly, the transition state would be characterized by a network of water-unsatisfied, broken internal contacts (see Fig. 1 for a pictorial illustration). In fact, desolvation has been suggested as the likely origin of the robust enthalpic barriers detected by Chevron-Eyring analysis of protein folding rates (3,4).

It must be noted that, when considered from a protein unfolding viewpoint, the barrier could be viewed as arising from the asynchrony between water penetration and the breaking of internal interactions and would be described as a "solvation" barrier (see Fig. 1). We show here that the energetic and structural consequences of solvation barriers for the folding transition state can be estimated from experimental unfolding rates using well-known structure-energetics correlations.

The enthalpy change associated to a protein conformational change includes contributions from breaking/formation of internal interactions ( $\Delta H_{int}$ ) and from the hydration/dehydration of the residues that become exposed-to-solvent/buried (5):

$$\Delta H = \Delta H_{int} + \Delta H_{hyd}. \quad (1)$$

For the conversion of the native to the unfolded state ( $N$  and  $U$  in Fig. 1), both contributions are linked (i.e., breaking of the internal interactions involving given residues implies the exposure of those residues to the solvent). In addition, both contributions have been shown to scale with changes in accessible surface area (5). Consequently, experimental en-

thalpy changes for protein unfolding have been shown to be well represented by the following empirical parametrization in terms of the unfolding changes in polar and apolar ASA:

$$\Delta H_{N \rightarrow U} = a \cdot \Delta ASA_{pol} + b \cdot \Delta ASA_{ap}, \quad (2)$$

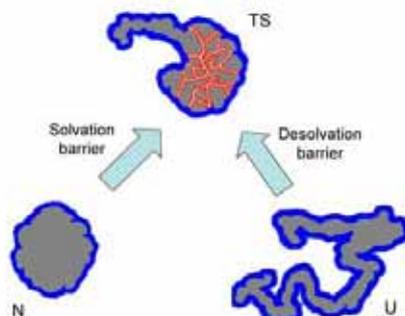
where  $a = -101.2 \text{ J}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$  and  $b = 169.5 \text{ J}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$  for  $T = 25^\circ\text{C}$  (6). Since native proteins bury an approximately constant proportion of polar and apolar ASA upon folding ( $0.417 \pm 3.4\%$  and  $0.583 \pm 3.4\%$ , respectively, as reported in Robertson and Murphy (7), Eq. 2 can be easily expressed in terms of the total unfolding ASA change:

$$\Delta H_{N \rightarrow U} = \alpha \cdot \Delta ASA_{N \rightarrow U}, \quad (3)$$

where  $\alpha = 11.7 \text{ J}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$  for  $T = 25^\circ\text{C}$ .

Consider now the unfolding activation energy.  $\Delta H_{N \rightarrow TS}$  has contributions associated to the surface exposed to the solvent upon formation of the transition state from the native protein (the *blue* surface in Fig. 1) and to the residues with broken internal interactions but not yet exposed to the solvent (the *red* surface in Fig. 1). The former contribution reflects compensated internal interactions and hydration terms (as in the global unfolding process) and, therefore, can be estimated from Eq. 3 with the activation change in ASA. The latter contribution ( $\Delta H^*$ ) is the enthalpic term associated to the solvation barrier we actually seek to calculate. Accordingly, the activation enthalpy can be written as

Biophysical Journal: Biophysical Letters



**FIGURE 1** Pictorial illustration of desolvation/solvation barriers to protein folding/unfolding. The surface shown in blue in the native (*N*), unfolded (*U*), and transition (*TS*) states is exposed to the solvent. The surface shown in red in the transition state represents broken internal contacts, which are not solvated by water.

$$\Delta H_{N \rightarrow TS} = \Delta H^* + \alpha \cdot \Delta ASA_{N \rightarrow TS}. \quad (4)$$

Values of  $\Delta H_{N \rightarrow TS}$  are experimentally available as activation energies for the unfolding process. Provided that some estimate is available for the activation change in ASA, Eq. 4 can be solved for  $\Delta H^*$ .  $\Delta ASA_{N \rightarrow TS}$  can be readily estimated from kinetic, denaturant  $m$  values, since denaturant  $m$  values are proportional to changes in surface exposed to the solvent (8).

The kind of calculation we have just outlined can be illustrated with the experimental data for hen egg white lysozyme unfolding reported in Ibarra-Molero and Sanchez-Ruiz (9). Fig. 2 shows Arrhenius plots for lysozyme unfolding rates at several high guanidine concentrations. These plots are essentially linear, supporting that the activation energy values can be taken as temperature-independent within the temperature range of the experimental data (18–45°C). They do depend somewhat on guanidine concentration, but the dependence appears linear (*inset* in Fig. 2) and allows the  $\Delta H_{N \rightarrow TS}$  value to be obtained as an extrapolation to zero denaturant concentration. We take the value thus obtained (208 kJ/mol) to be valid at 25°C. Under the assumption that denaturant  $m$  values and ASA values are linearly related (8),  $\Delta ASA_{N \rightarrow TS}$  is given by  $\Delta ASA_{N \rightarrow U} \cdot (m_{N \rightarrow TS} / m_{N \rightarrow U})$ .  $\Delta ASA_{N \rightarrow U}$  for lysozyme unfolding is calculated as  $14978 \text{ \AA}^2$  (10) and, from Ibarra-Molero and Sanchez-Ruiz (9), the kinetic and equilibrium  $m$  values are 2.9 and 10.5  $\text{kJ} \cdot \text{mol}^{-1} \cdot \text{M}^{-1}$ . Then,  $\Delta ASA_{N \rightarrow TS}$  is estimated as  $4090 \text{ \AA}^2$  (11) and the  $\alpha \cdot \Delta ASA_{N \rightarrow TS}$  term in Eq. 4 is 48 kJ/mol, significantly smaller than the experimental  $\Delta H_{N \rightarrow TS}$  value (208 kJ). The difference gives a large  $\Delta H^*$  value of 160 kJ/mol. This result is qualitatively robust; for instance, modifying the value of  $\alpha$  (Eqs. 3 and 4) by as much as plus/minus 50% only results in a change of  $\sim 30$  kJ/mol in the calculated  $\Delta H^*$  value. It must be noted, in addition, that we have used a tripeptides model to calculate the ASA of the

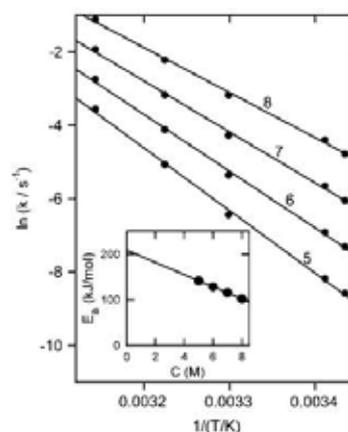
unfolded state (10); use of more compact models (12) would lead to even higher estimates of  $\Delta H^*$ .

We have carried out the above calculation with other protein systems for which the required experimental data are available in the literature (see Supplementary Material for details). The outcome of these calculations is summarized in Fig. 3 A as plots of  $\Delta H^*$  and  $\alpha \cdot \Delta ASA_{N \rightarrow TS}$  (the two contributions to the activation energy according to Eq. 4) versus  $\Delta H_{N \rightarrow TS}$ . In all cases, large values of  $\Delta H^*$  are found. In fact, the values of the slopes of the plots in Fig. 3 A indicate that, for the protein set studied,  $\sim 80\%$  of the activation energy value is to be attributed to the solvation barrier. A measure of the structural impact of such solvation barriers can be derived from the  $\Delta H^*$  values using the following equation:

$$\Delta H^* = \beta_{ap} \cdot U_{ap}^b \cdot ASA_{ap}^* + \beta_{pol} \cdot U_{pol}^b \cdot ASA_{pol}^* + \beta_{mix} \cdot U_{mix}^b \cdot ASA^*, \quad (5)$$

which is based upon the parametrization of the enthalpy changes for breaking of internal interactions reported by Freire and co-workers (5). Equation 5 takes into account the apolar-apolar, polar-polar, and polar-apolar (i.e., "mix") internal interactions and the  $U$  parameters are the corresponding energy-averaged separation distances. The  $U$  values are always close to 0.73 (see Table I in Hilser et al. (5)); using this, the values of  $\beta_{ap}$ ,  $\beta_{pol}$ , and  $\beta_{mix}$  given in Hilser et al. (5), and the known fractions of polar and apolar surface buried upon folding ( $0.417 \pm 3.4\%$  and  $0.583 \pm 3.4\%$ , respectively, as reported in Robertson and Murphy (7)), Eq. 5 can be simplified to

$$ASA^* \approx \Delta H^* / \beta \quad (6)$$



**FIGURE 2** Arrhenius plots for hen egg white lysozyme unfolding at several guanidine concentrations (numbers alongside the lines). The inset shows a plot of activation energy versus guanidine concentration; extrapolation to zero denaturant concentration yields the activation energy in water ( $\Delta H_{N \rightarrow TS} = 208$  kJ/mol).

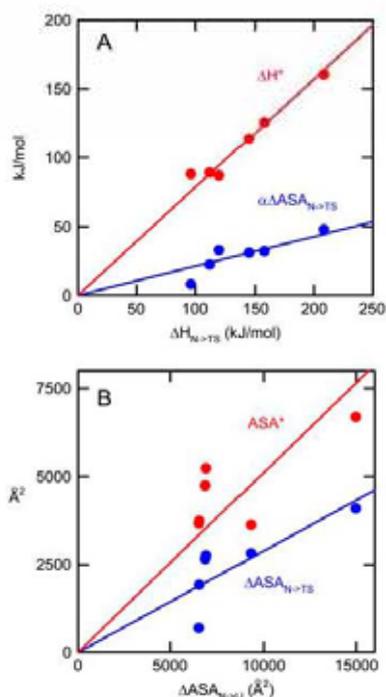
Biophysical Journal: Biophysical Letters

L49

Biophysical Journal: Biophysical Letters

with  $\beta = 24 \text{ J}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$ . As shown in Fig. 3 B, the  $\text{ASA}^*$  values thus estimated are surprisingly large: on the order of several thousands of  $\text{\AA}^2$ . This suggests that the broken, water-unsatisfied contacts may form large, extended networks in the transition states for protein folding/unfolding. In fact, the slopes of the plots of  $\Delta\text{ASA}_{\text{N}\rightarrow\text{TS}}$  and  $\text{ASA}^*$  versus  $\Delta\text{ASA}_{\text{N}\rightarrow\text{U}}$  of Fig. 3 B indicate that, for the protein set studied,  $\sim 30\%$  of the total unfolding ASA change is already exposed to the solvent in the transition state and  $\sim 50\%$  of that total ASA change is still buried but involved in broken internal contacts.

The analyses reported here focus on the energetic (enthalpic) consequences of solvation/desolvation barriers. Certainly, enthalpy is only half of the story, since the free energy associated to these barriers will have enthalpic as well as entropic components (4,13). Nevertheless, the analysis of activation enthalpies according to known structure-energetics relationships reveals the different energetic balance of the native and folding transition states and provides a clear qualitative picture of the structural consequences of solvation/desolvation for the transition states.



**FIGURE 3** Energetic (panel A) and structural (panel B) consequences of the solvation barrier for the unfolding transition state for the following proteins: hen egg white lysozyme, chymotrypsin inhibitor 2, apocytochrome B<sub>9</sub>, cold shock protein B (from *Bacillus subtilis*), coiled-coil peptide GCN4-p1, and IgG binding domain of protein L.

Biophysical Journal: Biophysical Letters

## SUPPLEMENTARY MATERIAL

An online supplement to this article can be found by visiting BJ Online at <http://www.biophysj.org>.

## ACKNOWLEDGMENTS

This work was supported by grant BJO2003-02229 from the Spanish Ministry of Science and Education, Feder Funds, and grant CVI-0771 from the "Junta de Andalucía".

## REFERENCES and FOOTNOTES

1. Rank, J. A., and D. Baker. 1997. A desolvation barrier to hydrophobic cluster formation may contribute to the rate-limiting step in protein folding. *Protein Sci.* 6:347-354.
2. Cheung, S. M., A. E. García, and J. N. Onuchic. 2002. Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc. Natl. Acad. Sci. USA.* 99:685-690.
3. Scalley, M. L., and D. Baker. 1997. Protein folding kinetics exhibits an Arrhenius temperature dependence when corrected for the temperature-dependence of protein stability. *Proc. Natl. Acad. Sci. USA.* 94:10636-10640.
4. Liu, Z., and H. S. Chan. 2005. Desolvation is a likely origin of robust enthalpic barriers to protein folding. *J. Mol. Biol.* 349:872-889.
5. Hilser, V. J., J. Gomez, and E. Freire. 1996. The enthalpy change in protein folding and binding: refinement of parameters for structure-based calculations. *Proteins.* 26:123-133.
6. These values at 25 °C are easily calculated from the reported values at the reference temperature of 60 °C (5), taking into account that the polar and apolar contributions to the unfolding enthalpy change with temperature according to the corresponding contributions to the unfolding heat capacity.
7. Robertson, A. D., and K. P. Murphy. 1997. Protein structure and the energetics of protein stability. *Chem. Rev.* 97:1251-1267.
8. Myers, J. K., C. N. Pace, and J. M. Scholtz. 1995. Denaturant *m* values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* 4:2138-2148.
9. Ibarra-Molero, B., and J. M. Sanchez-Ruiz. 1996. A model-independent, nonlinear extrapolation procedure for the characterization of protein folding energetics from solvent-denaturation data. *Biochemistry.* 35:14689-14702.
10. ASA values for the native states were calculated using a modification of the Shrake-Rupley algorithm, which randomly places 2000 points in the expanded van der Waals sphere representing each atom. A radius of 1.4 Å for the solvent probe and Chothia set for the protein atoms were used. ASA values for the unfolded states were calculated on the basis of a Gly-X-Gly tripeptides model.
11. Alternatively,  $\Delta\text{ASA}_{\text{N}\rightarrow\text{TS}}$  could be calculated solely from the kinetic *m* value by using the slope of the plot of guanidine-*m* versus ASA reported in Hilser et al. (5) ( $0.92 \text{ J}\cdot\text{mol}^{-1}\cdot\text{M}^{-1}\cdot\text{\AA}^{-2}$ ). The result obtained ( $3117 \text{ \AA}^2$ ), however, is close to that calculated using the method described in the text ( $4089 \text{ \AA}^2$ ) and the difference would not significantly modify the  $\Delta H^\ddagger$  value (from 160 kJ/mol to 172 kJ/mol) and the subsequent interpretations.
12. Ibarra-Molero, B., I. M. Plaza del Pino, B. Souhail, H. O. Hammou, and J. M. Sanchez-Ruiz. 2000. The sarcosine effect on protein stability: a case of nonadditivity? *Protein Sci.* 9:820-826.
13. Solvent entropy (4) and side-chain conformational entropy of the residues involved in broken internal interactions are among the likely contributors to the solvation activation entropy. However, estimation of  $\Delta S_{\text{N}\rightarrow\text{TS}}$  from experimental unfolding rates within the framework of transition-state theory is hampered by uncertainties associated with the value of the front factor in the Eyring equation. In any case, due to enthalpy-entropy compensation, a low free-energy barrier can be consistent with a high enthalpic solvation/desolvation barrier, as Liu and Chan have pointed out (4). In fact, the solvation/desolvation free-energy barrier is expected to vanish in cases of downhill folding.

L50

### 2.3-ARTICULO TERCERO

"Beyond Lumry-Eyring: an unexpected pattern of operational reversibility/irreversibility in protein denaturation"

Rodriguez-Larrea D, Ibarra-Molero B, de Maria L, Borchert TV, Sanchez-Ruiz JM.

*Proteins.* 2008 Jan 1;70(1):19-24.

**2.3.1-INTRODUCCIÓN ARTÍCULO 3º**

Este trabajo se enmarca en los estudios que se llevaron para caracterizar la energética del desplegamiento de la lipasa de *Thermomyces lanuginosa* (Rodríguez-Larrea, Minning et al. 2006), una de las proteínas industriales más importantes. En él profundizamos en el concepto de reversibilidad operacional en desplegamiento térmico. Presentamos dos ejemplos experimentales muy ilustrativos. Uno, la fosfoglicerato quinasa, que sigue el patrón esperado de reversibilidad/irreversibilidad según modelos simples de Lumry-Eyring, mientras que el otro ejemplo, la lipasa, claramente se desvia del comportamiento esperado. Estudios previos de calorimetría diferencial de barrido (Rodríguez-Larrea, Minning et al. 2006), pusieron de manifiesto que el proceso de desplegamiento térmico de lipasa se describía adecuadamente según un modelo de dos estados irreversible. Ello implica que la proteína no se encuentra en equilibrio en ningún momento durante la transición calorimétrica. Para comprobar este requisito (ausencia de reversibilidad) en calorimetría realizamos lo que frecuentemente denominamos "reheating", que no es más que realizar un segundo barrido sobre la misma muestra. Esto se debe hacer de un modo rutinario (Sanchez-Ruiz 1995).

Sorprendentemente uno de los "reheatings" realizados en nuestros experimentos con urea mostró una transición. Esto no es necesariamente excepcional, puesto que la urea, con su actividad anti-agregante, permite observar reversibilidad (siempre y cuando la agregación sea la única causa de la irreversibilidad y la urea la evite). Lo que hizo excepcional el resultado fueron los siguientes puntos:

-La transición calorimétrica obtenida en el reheating no era la misma que la obtenida en el primer barrido puesto que tenía una temperatura de desnaturalización algunos grados mayor.

-La transición en el reheating dependía de cual era la temperatura final del primer barrido. Esta dependencia tenía la siguiente naturaleza: Si el primer barrido llegaba justo al final de la transición, no había "reversibilidad". Si la temperatura final era mayor, se observaba la transición en el reheating.

Este último punto, en primer lugar, permite aseverar que la transición en el primer calentamiento de la muestra es irreversible. En segundo lugar supone que se da un comportamiento anómalo, puesto que las proteínas, cuando reversibles, tienden a reducir el grado de

reversibilidad a medida que incrementamos la temperatura a la cual la exponemos. Finalmente sugiere la existencia de un proceso a altas temperaturas (más allá de la transición), en el cual la urea juega un papel relevante, que determina si la proteína será capaz de replegarse al ser enfriada de nuevo. Con respecto al punto número uno, este sugiere que la especie que se obtiene tras el primer barrido es distinta de la original. Sin embargo retiene su capacidad catalítica y por tanto no debe diferenciarse sustancialmente de la proteína nativa.

Con esta información decidimos hacer una caracterización detallada y exhaustiva de la naturaleza del fenómeno. Tras numerosos experimentos con lipasa, se muestra claramente como el porcentaje de reversibilidad obtenido depende inversamente de la temperatura a la cual calentamos la muestra. Es decir, cuanto más calentamos la muestra más favorecemos la ocurrencia de fenómenos irreversibles que, en último término, provocan una menor reversibilidad observada en el segundo scan o reheating. Una vez analizada y agrupada toda la información obtenida, se procedió a sugerir un modelo fenomenológico de la máxima simplicidad y que fuera capaz de explicar razonablemente los resultados obtenidos con la Lipasa.

La importancia del presente trabajo no sólo radica en ser la primera caracterización de un fenómeno de esta naturaleza. En vista de la teoría del "protein landscape" (Onuchic, Luthey-Schulten et al. 1997) el presente trabajo sugiere la posibilidad de que existan diferentes estados nativos para una misma secuencia. La proteína mantendría su estado nativo "fisiológico" gracias a la presencia de barreras energéticas que impiden que otras estructuras compatibles con la secuencia sean accesibles. En este contexto, pensamos que el presente trabajo puede tener relevancia en la comprensión de las enfermedades relacionadas con amiloides y priones, para los que la conformación de proteína inicial posiblemente determina las características químicas y consecuencias biológicas de los procesos subsecuentes de agregación.



## SHORT COMMUNICATION

# Beyond Lumry–Eyring: An unexpected pattern of operational reversibility/irreversibility in protein denaturation

David Rodriguez-Larrea,<sup>1</sup> Beatriz Ibarra-Molero,<sup>1</sup> Leonardo de Maria,<sup>2</sup> Torben V. Borchert,<sup>2</sup> and Jose M. Sanchez-Ruiz<sup>1\*</sup>

<sup>1</sup>Facultad de Ciencias, Departamento de Química Física, 18071-Granada, Spain

<sup>2</sup>Research and Development, Novozymes A/S, Brødelysvej 26, 2880-Bagvaerd, Denmark

### ABSTRACT

We have found that, contrary to naïve intuition, the degree of operational reversibility in the thermal denaturation of lipase from *Thermomyces lanuginosa* (an important industrial enzyme) in urea solutions is maximum when the protein is heated several degrees above the end of the temperature-induced denaturation transition. Upon cooling to room temperature, the protein seems to reach a state with enzymatic activity similar to that of the initial native state, but with higher denaturation temperature and radically different behavior in terms of susceptibility to irreversible denaturation. These results show that patterns of operational reversibility/irreversibility in protein denaturation may be more complex than the often-taken-for-granted, two-situation classification (reversible vs. irreversible). Furthermore, they are consistent with the possibility of existence of different native or native-like states separated by high kinetic barriers under native conditions and they suggest experimental procedures to reach and study such “alternative” native states.

Proteins 2008; 70:19–24.  
© 2007 Wiley-Liss, Inc.

**Key words:** protein denaturation; reversibility; irreversibility; kinetic stability; aggregation; chemical denaturants; calorimetry

### INTRODUCTION

The reversibility of the denaturation process is one of the most fundamental issues in protein chemistry. Two situations are usually reported to occur.<sup>1</sup> (1) In some cases, the protein, previously denatured by a high concentration of a chemical denaturant or high temperature, is able to fold back to the native state upon its transfer to native conditions (room temperature, low concentration of denaturant, ...). The denaturation process is then said to be operationally reversible. Operational reversibility is often taken to mean that denaturation involves the unfolding of the native structure and that the unfolded protein molecules can solve the Levinthal paradox<sup>2</sup> in the time-scale of the renaturation experiment. Also, for an operational reversible process, it should be possible to set up a denaturation experiment in which conditions (temperature and denaturant concentration) are changed gradually in such a way that equilibrium between significantly populated states is always established. That is, operational reversibility may allow denaturation experiments to be carried out under conditions of thermodynamic reversibility and protein thermodynamic stability to be determined. (2) In other cases, the denatured protein is unable to fold back to the native state upon transfer to native conditions (in a given time scale) and denaturation is operationally irreversible. Operational irreversibility is generally attributed to alterations of unfolded or partially unfolded states (aggregation, autolysis, chemical alteration of residues, etc.) that may occur under denaturing conditions. Very often, operationally irreversible denaturation is found to be a nonequilibrium process which must be described in terms of rate equations and whose analysis

Grant sponsor: Spanish Ministry of Education and Science; Grant number: BIO2006-07332; Grant sponsor: Junta de Andalucía; Grant number: CVI-771; Grant sponsor: FEDER Funds.

\*Correspondence to: Jose M. Sanchez-Ruiz, Facultad de Ciencias, Departamento de Química Física, Fuentenueva s/n, 18071-Granada, Spain. E-mail: sancher@ugr.es

Received 13 February 2007; Accepted 17 April 2007

Published online 19 July 2007 in Wiley InterScience (www.interscience.wiley.com).

DOI: 10.1002/prot.21572

D. Rodríguez-Larrea et al.

leads to the determination of kinetic stability.<sup>1,3</sup> Understanding protein kinetic stability is of considerable biotechnological importance<sup>4</sup> and may also have relevance in the development of approaches to the inhibition of amyloidogenesis.<sup>5</sup>

Many fundamental aspects of protein denaturation, in particular those related with reversibility, can be discussed in terms of Lumry-Eyring models,<sup>1,3,6,7</sup> which take into account that overall irreversible denaturation consists of at least two-steps:



where F is a final, irreversibly denatured "state" (unable to fold back to the native state) and U stands for the unfolded state or, from a more general viewpoint, for the ensemble of unfolded and partially unfolded states (the non-native ensemble<sup>3</sup>). In terms of this Lumry-Eyring model, the situation 1 we have above-mentioned (operational reversibility) corresponds to the case in which the irreversible alteration step ( $U \rightarrow F$ ) is comparatively slow and does not take place significantly in the time-scale of the experiment. On the other hand, when the irreversible step is fast in the time scale of the experiment, the protein ends up in the final state and the process is overall irreversible (situation 2 described earlier). Furthermore, for a sufficiently fast irreversible step, U never gets significantly populated and the denaturation process can be phenomenologically described as a two-state kinetically-controlled conversion from the native state to the irreversibly denatured protein:



This case is usually known as the two-state irreversible model<sup>1</sup> and has been shown to describe adequately the thermal denaturation of many proteins (see Plaza del Pino *et al.*<sup>3</sup> and references quoted therein).

The influence of irreversible alteration steps [ $U \rightarrow F$  in the simple Lumry-Eyring model of Eq. (1)] on the overall denaturation process can often be modulated to some significant extent. For instance, in thermal denaturation experiments in which temperature is increased with time (scanning experiments, such as those carried out with differential scanning calorimetry); the degree of operational reversibility may depend on the temperature at which the experiment is stopped [provided that the limiting case represented by Eq. (2) has not been reached]. Furthermore, if the irreversible step is mainly due to aggregation it can be slowed down (or even prevented in some cases) by adding moderate amounts of chemical denaturants (urea and guanidine). Therefore, the expected behavior in thermal denaturation experiments is that operational reversibility will decrease as the final temperature of the experiment increases (since the rate of the  $U \rightarrow F$  step should increase with temperature) and will increase with the concentration of chemical denatur-

ants (since these substances disfavor protein aggregation). In this work, we illustrate this "normal" behavior with the thermal denaturation of yeast phosphoglycerate kinase (PGK) in guanidine solutions. However, in the course of an experimental study into the stability of lipase from *Thermomyces lanuginosa* (an important industrial enzyme) in urea solutions, we have found a pattern of reversibility/irreversibility which appears more complex (and far more interesting) than this "normal, expected" behavior. Actually, the reversibility/irreversibility pattern we have found for lipase does not fit the often-taken-for-granted, two-situation classification described in the opening paragraph in this "Introduction" (reversible vs. irreversible) and cannot be explained in terms of simple Lumry-Eyring models [such as that shown in Eq. (1)]. As we discuss in some detail in this work, the possibility of occurrence of such "unexpected" patterns of reversibility/irreversibility may have general implications for our understanding of protein kinetic stability and its consequences.

## METHODS

### Protein sample preparation

Yeast PGK was purchased from Sigma as an ammonium sulphate precipitate and used without further purification. PGK solutions for calorimetric experiments were prepared by exhaustive dialysis against 50 mM PIPES buffer pH 7.0. PGK concentrations were calculated spectrophotometrically from the absorbance at 280 nm. Further experimental details are as given in Galisteo *et al.*<sup>8</sup>

Lipase from *Thermomyces lanuginosa* was expressed and purified as previously described.<sup>4</sup> Lipase solutions for calorimetric experiments were prepared by exhaustive dialysis against 50 mM sodium acetate, pH 5. Further experimental details are as described in Rodríguez-Larrea *et al.*<sup>4</sup> Lipase activity was assayed by following the increase in absorbance at 405 nm of a solution of *p*-nitrophenylvalerate after addition of the enzyme. In all cases, linear changes of absorbance with time were observed.

For experiments in the presence of guanidine or urea, the adequate amounts of stock solutions (in the same buffer as the protein under study) were added. Urea and guanidine concentrations were determined from refraction index measurements.

### Differential scanning calorimetry

Differential scanning calorimetry (DSC) experiments were carried out using a VP-DSC microcalorimeter or a capillary VP-DSC microcalorimeter equipped with the auto-sampler attachment. Both calorimeters were from MicroCal (Northhampton, USA). Most DSC experiments were performed at 1.5 K/min but some addi-

## Reversibility in Protein Denaturation

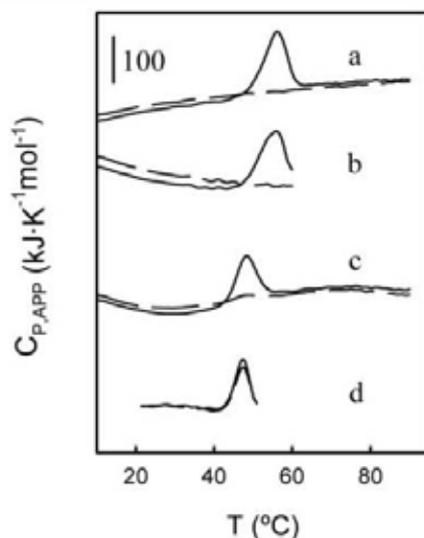
tional experiments were carried out at lower scan rates (1 and 0.5 K/min). The actual scan rates of the experiments were calculated from the temperature/time data, although they differ little from the nominal values. In all cases, operational reversibility was checked by performing a reheating run after stopping the first run at a given temperature ( $T_{\text{FINAL}}$ ) and cooling to room temperature. The fraction of reversibility was calculated as the ratio between the areas under the calorimetric transition in the reheating and first runs. Additional details about the DSC experiments have been previously described.<sup>4,9</sup>

Fittings of the two-state irreversible model [Eq. (2)] to the lipase DSC profiles were carried out as we have previously described in detail.<sup>4</sup>

## RESULTS AND DISCUSSION

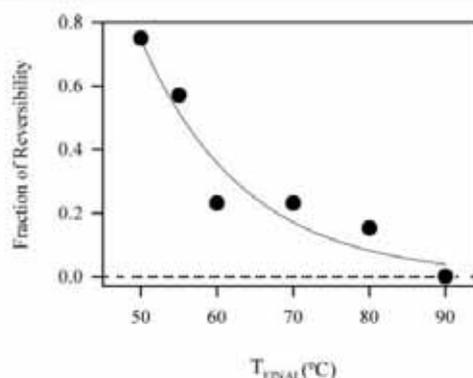
### Thermal denaturation of PGK in guanidine solutions: an expected pattern of operational reversibility/irreversibility

We have previously studied the thermal denaturation of PGK in aqueous solution by differential scanning calo-



**Figure 1**

Operational reversibility in the thermal denaturation of yeast PGK. First DSC runs (continuous lines) and reheating runs (dashed lines) in the absence and presence of guanidine (0.47M and 0.52M for the profiles c and d, respectively). Note that, in profiles b and d, the first scan was stopped immediately after the transition, while, in profiles a and c, the first scan was stopped at 50°C.



**Figure 2**

Operational reversibility in the thermal denaturation of yeast PGK. Effect of the temperature at which the first run is terminated ( $T_{\text{FINAL}}$ ) on the degree of operational reversibility for experiments carried out in the presence of 0.47M guanidine.

rimetry.<sup>8</sup> The process was found to be operationally irreversible and kinetically controlled, as shown by a strong scan-rate effect on the DSC transitions. However, the process was not found to follow the two-state irreversible model [Eq. (2)], but a more complex kinetics.<sup>8</sup> In any case, aggregation was identified as one of the main sources of irreversibility and, in subsequent work,<sup>10</sup> moderate amounts of guanidine were found to lead to significant operational reversibility.

In this work, we have systematically characterized the reversibility of PGK denaturation as a function of guanidine concentration and the final temperature of the first scan in DSC experiments (see Fig. 1 for some representative examples of the analyzed DSC profiles). In agreement with earlier findings,<sup>10</sup> we have found significant reversibility with guanidine concentrations above 0.4M approximately. Figure 2 shows the effect of the final temperature of the first scan on reversibility for DSC experiments carried out in the presence of 0.47M guanidine (in the absence of denaturant no reversibility is found, regardless of the value of  $T_{\text{FINAL}}$ ; data not shown). In the presence of 0.47M guanidine, reversibility is a maximum when the scan is terminated immediately after the completion of the DSC transition and decreases as  $T_{\text{FINAL}}$  is increased.

The results summarized in Figure 2 exemplify what we have referred to in the "Introduction" as the "normal, expected" behavior: since aggregation appears to be the main process responsible for irreversibility in this case, operational reversibility can be induced by moderate concentrations of a chemical denaturant; furthermore, for a

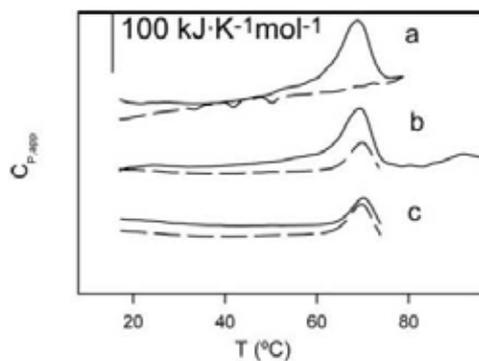
D. Rodríguez-Larrea et al.

given chemical denaturant concentration, reversibility decreases with increasing  $T_{\text{FINAL}}$ , since the rate of irreversible alteration processes is expected to increase with temperature.

**Thermal denaturation of lipase from *Thermomyces lanuginosa* in urea solutions: an unexpected pattern of operational reversibility/irreversibility**

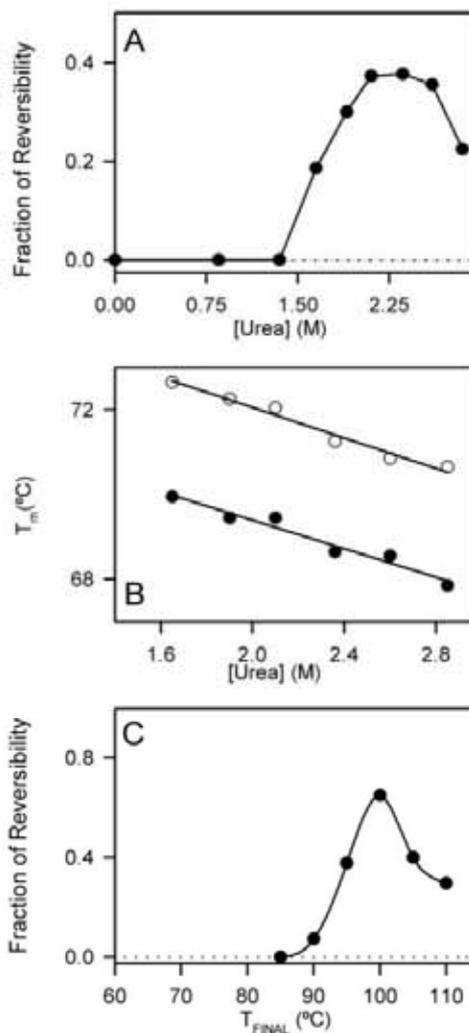
We have recently reported<sup>4</sup> that DSC thermograms for lipase thermal denaturation in the absence and presence of moderate urea concentrations (0–2M) show a distinct transition in the approximate 65–90°C range (the exact value of the temperature of the maximum of the transition depending on the variant, the urea concentration, and the scan rate of the DSC experiment). The denaturation transition was found to be operationally irreversible (absence of the transition in a reheating run when the first run had been stopped immediately after the transition) and kinetically-controlled. Aggregation appeared to be one of the main causes of irreversibility in this case, as suggested by the turbidity of the samples extracted from the calorimetric cells after completion of the experiment.

In this work, we have extended the study to urea concentrations higher than those previously employed.<sup>4</sup> Turbidity of the samples extracted from the calorimetric cells after completion of the experiment appears greatly diminished and we have found significant operational reversibility but, rather surprisingly, only when the first DSC run is terminated at temperatures several degrees above the end of the transition; if the first scan is termi-



**Figure 3**

Operational reversibility in the thermal denaturation of lipase from *Thermomyces lanuginosa*. (a) First DSC run (continuous line) and reheating run (dashed line) obtained after stopping the first run immediately after the transition. (b) Same as in a, but here the first scan was terminated at 100°C. (c) Reheating run from b (continuous line) and its reheating (third scan, dashed line). In all cases, the experiments were carried out at a scan rate of 1.5 K/min and in the presence of 2.5M urea.

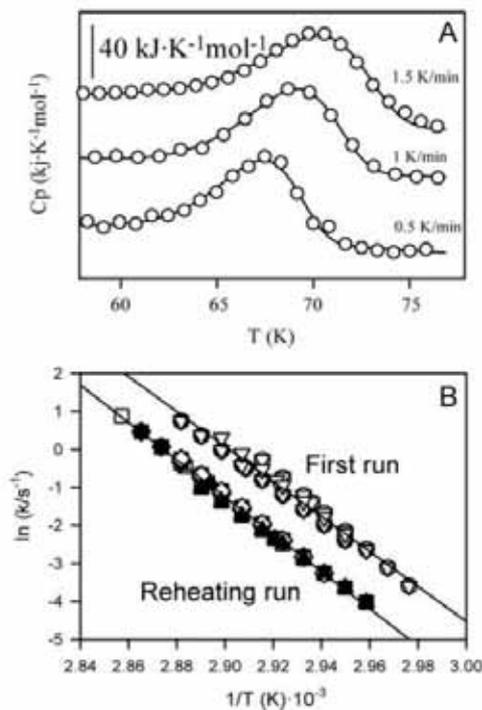


**Figure 4**

Operational reversibility in the thermal denaturation of lipase from *Thermomyces lanuginosa*. (A) Effect of urea on the degree of operational reversibility of the thermal denaturation of lipase. First runs were carried out at 1.5 K/min and terminated at  $T_{\text{FINAL}} = 95^\circ\text{C}$ . (B) Values for the temperature of the maximum of the transitions observed in the first (closed symbols) and reheating (open symbols) runs of the experiments described in A. (C) Effect of the temperature at which the first scan is terminated ( $T_{\text{FINAL}}$ ) on the operational reversibility of the thermal denaturation of lipase. Urea concentration was 2.6M and scan rate 1.5 K/min in all these experiments.

nated immediately after the transition, no reversibility is found (see Fig. 3 for some representative examples of the DSC profiles for lipase denaturation analyzed in this work). In fact, reversibility is maximum for a urea con-

## Reversibility in Protein Denaturation

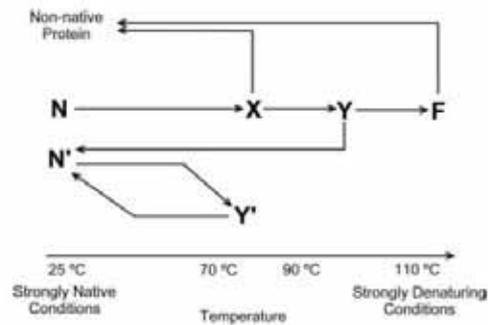


**Figure 5**

Analysis of the thermal denaturation of lipase from *Thermomyces lanuginosus* according to the two-state irreversible model. (A) Experimental DSC reheating transitions for the denaturation of lipase at the indicated scan rates. Open symbols are the experimental apparent heat capacity data and the continuous lines are the best fit of the two-state irreversible model. The profiles have been shifted in the y-axis for the sake of clarity. (B) Arrhenius plots for the thermal denaturation lipase. The values of the first-order rate constant were derived from the analysis of experimental DSC profiles for the first-run and reheating-run transitions, according to the two-state irreversible model, as described in Rodríguez-Larrea et al.<sup>6</sup> and illustrated in panel A. The different symbols correspond to the different scan rates of the original experiments. Note that the agreement between the rate constants derived from calorimetric experiments carried at different scan rates provides support for the applicability of the two-state irreversible model.

centration of about 2.4 M [Fig. 4(A)] and when the scan is terminated at about 100° [Fig. 4(C)]. The transition observed in the reheating run is, like the first-scan transition, kinetically controlled (as shown by the scan rate effect: see Fig. 5) and conforms to the two-state irreversible model (see Fig. 5). Again rather surprisingly, however, the reheating transition has a  $T_m$  temperature a few degrees higher than that for the original transition seen in the first run [see Fig. 4(B)] and shows operational reversibility [see third scan in Fig. 3(C)] regardless of the temperature at which second scans are terminated.

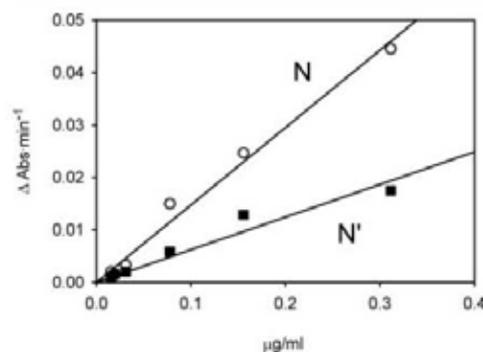
Clearly, the experimental results described above for lipase cannot be explained in terms of a simple Lumry–



**Figure 6**

Minimal scheme that summarizes the results of the operational reversibility experiments with lipase from *Thermomyces lanuginosus* (see text for details). Note that the double arrow between  $N'$  and  $Y'$  is only meant to indicate operational reversibility (not chemical equilibrium).

Eyring model, such as that embodied in Eq. (1). A minimal scheme that summarizes these results is proposed in Figure 6. Upon temperature increase, the native protein is converted to a “state”, X, which is unable to fold back to the native protein upon cooling; obviously, the  $N \rightarrow X$  conversion could involve lowly populated intermediate states ( $NU \rightarrow X$ , for instance). In any case, the  $N \rightarrow X$  conversion is taken to be responsible for the main transition seen in the first-scan thermograms. The combined effect of urea and higher temperature, transforms X into a new “state”, Y, which, upon cooling, folds to a native state,



**Figure 7**

Lipase activity assayed by following the increase in absorbance at 405 nm of a solution of p-nitrophenylacetate after addition of the enzyme. In all cases, linear changes of absorbance with time were observed. The figure shows a plot of activity (change in absorbance per unit time) versus lipase concentration in the assay for the N and  $N'$  forms. The  $N'/N$  activity ratio is 0.44 which qualitatively agrees with the degree of reversibility of the experiments used to prepare the  $N'$  form (about 0.6).

D. Rodríguez-Larrea et al.

$N'$ , which differs in some respects from  $N$  (although  $N$  and  $N'$  have similar enzymatic activities after correcting for the degree of operational reversibility; see Fig. 7). Thermal denaturation of  $N'$  leads to  $Y$ , which folds again to  $N'$  upon cooling.

## CONCLUDING REMARKS

We have given an experimental example in this work (PGK from yeast in guanidine solutions) of the kind of pattern of operational reversibility/irreversibility expected from simple Lumry-Eyring models [Eq. (1)]. We also provide one experimental example (lipase from *Thermomyces lanuginosa* in urea solutions) of a pattern of reversibility/irreversibility that departs from these seemingly intuitive expectations. We believe that such "unexpected" patterns may actually be common and that they have gone mostly unnoticed because researchers were unaware of the possibility.

Certainly, our experimental data on lipase denaturation in urea solutions cannot be interpreted in terms of simple Lumry-Eyring models, such as that given in Eq. (1). We have thus proposed a more complex scheme (Fig. 6). This proposal should be regarded as a minimal model that rationalizes the data reported in this work in a simple way, but which should be susceptible of modification/refinement in the light of additional experimental information. Note, in particular, that the processes involved in the scheme of Figure 6 are all of a kinetic nature (i.e., the stability of both  $N$  and  $N'$  is kinetic stability) and, consequently, the operational reversibility/irreversibility pattern observed may depend strongly on the time scale of the denaturation/renaturation experiments (the time-scale and time-history of the experiments shown in Figure 3 are the standard for protein DSC studies). Also, we do not have at this stage a structural model of "states"  $X$  and  $Y$  (for instance,  $X$  could be an aggregated protein which is disrupted at high temperature in the presence of urea).

The above uncertainties notwithstanding, our results have several implications of general interest. First, it is clear that operational reversibility/irreversibility in protein denaturation can be more complex than the two "common" situations often reported (reversible vs. irreversible: see opening paragraph in the "Introduction"). Also, the possibility of achieving operational reversibility by combining chemical denaturants and high temperature (prior to cooling to room to temperature) could perhaps find practical application in for instance the refolding to the functional states of proteins obtained as inclusion bodies.

Finally, several studies<sup>11–14</sup> have suggested that kinetic stability in proteins may be accompanied with the possibility of existence of different native or native-like states which are separated by high kinetic barriers under native conditions and which do not equilibrate in a bio-

logically relevant time-scale. The results reported here suggest experimental procedures to reach and study such "alternative" native states. In this regard, it is interesting to note that, although the difference between the structures of states  $N$  and  $N'$  (scheme of Fig. 6) is probably comparatively small ( $N$  and  $N'$  have similar activities and similar, although not identical,  $T_m$  values), it does bring about a radically different behavior in terms of operational reversibility and susceptibility to irreversible denaturation. It appears then, at least plausible that the analysis of such differences could provide information relevant for the understanding of amyloid and prion diseases, in which the starting protein conformation possibly determines the chemical features and biological consequences of subsequent aggregation processes.<sup>15</sup>

## REFERENCES

1. Sanchez-Ruiz JM. Theoretical analysis of Lumry-Eyring models in differential scanning calorimetry. *Biophys J* 1992;61:921–935.
2. Levinthal C. Are there pathways for protein folding? *J Chem Phys* 1968;85:44–45.
3. Plaza del Pino IM, Ibarra-Molero B, Sanchez-Ruiz JM. Lower kinetic limit to protein thermal stability: a proposal regarding protein stability in vivo and its relation with misfolding diseases. *Proteins* 2000;40:58–70.
4. Rodríguez-Larrea D, Minning S, Borchert TV, Sanchez-Ruiz JM. Role of solvation barriers in protein kinetic stability. *J Mol Biol* 2006;360:715–724.
5. Petrassi MM, Johnson SM, Purkey HE, Chiang KP, Walkup T, Jinag X, Powers ET, Kelly JW. Potent and selective structure-based dibenzofuran inhibitors of transthyretin amyloidogenesis: kinetic stabilization of the native state. *J Am Chem Soc* 2005;127:6662–6671.
6. Lumry R, Eyring H. Conformational changes in proteins. *J Phys Chem* 1954;58:110–120.
7. Klibanov AM, Ahern TJ. Thermal stability of proteins. In: Oxender DL, Fox CF, editors. *Protein engineering*. New York: Alan R Liss; 1987. pp 213–218.
8. Galisteo ML, Mateo PL, Sanchez-Ruiz JM. Kinetic study on the irreversible thermal denaturation of yeast phosphoglycerate kinase. *Biochemistry* 1991;30:2061–2066.
9. Godoy-Ruiz R, Perez-Jimenez R, Ibarra-Molero B, Sanchez-Ruiz JM. Relation between protein stability, evolution and structure, as probed by carboxylic acid mutations. *J Mol Biol* 2004;336:313–318.
10. Freire E, Murphy KP, Sanchez-Ruiz JM, Galisteo ML, Privalov PL. The molecular basis of cooperativity in protein folding. Thermodynamic dissection of interdomain interactions in phosphoglycerate kinase. *Biochemistry* 1992;31:250–256.
11. Sinclair JE, Ziegler MM, Baldwin TO. Kinetic partitioning during protein folding yields multiple native states. *Nat Struct Biol* 1994;1:320–326.
12. Fasshauer D, Antonin W, Subramaniam V, Jahn R. SNARE assembly and disassembly exhibit a pronounced hysteresis. *Nat Struct Biol* 2002;9:144–151.
13. Lundquist M, Sethson I, Johnson B-H. Transient interaction with nanoparticles "freezes" a protein in an ensemble of metastable near-native conformations. *Biochemistry* 2005;44:10093–10099.
14. Brujić J, Heermans RI, Walther KA, Fernandez JM. Single-molecule force spectroscopy reveals signatures of glassy dynamics in the energy of ubiquitin. *Nat Phys* 2006;2:282–286.
15. Tanaka M, Collins SR, Toyama BH, Weissman JS. The physical basis of how prion conformations determine strain phenotypes. *Nature* 2006;442:585–589.

## **3-RESUMEN Y CONCLUSIONES**

En la presente Tesis Doctoral hemos abordado dos tipos de estudios de relevancia en ingeniería de proteínas.

En primer lugar, nos hemos planteado analizar los factores que determinan la estabilidad cinética de proteínas. Para ello, hemos utilizando como sistema modelo la lipasa de *Thermomyces lanuginosa*, una proteína de enorme importancia industrial. Así, hemos caracterizado en detalle la desnaturalización térmica de la forma silvestre de la lipasa y de diversas variantes de estabilidad incrementada y obtenidas mediante una estrategia de evolución dirigida *in vitro* guiada por simulaciones de dinámica molecular. Nuestros estudios nos permiten concluir que la estabilidad de estas proteínas es de naturaleza fundamentalmente cinética y que el éxito de la estrategia antes citada se debe a que las simulaciones de dinámica molecular a alta temperatura identifican las regiones desplegadas (o perturbadas) de los estados de transición cinéticamente relevantes. Más aun, nuestros estudios experimentales revelan que dichos estados de transición presentan un balance energético diferente al correspondiente a los estados nativos. En particular, los estados de transición se caracterizan por la existencia de una red de interacciones internas rotas o debilitadas pero aun no satisfechas por interacciones con moléculas de agua. Esta asincronía entre la rotura de interacciones internas y la penetración del agua puede incrementar considerablemente la entalpía de los estados de transición, así como su energía libre, aunque en menor medida en este último caso debido a la existencia de compensación entalpía/entropía. Nos referimos a estas contribuciones energéticas como "barreras de solvatación". Las barreras de solvatación pueden ser una característica universal en procesos de desnaturalización y de plegamiento/desplegamiento en proteínas. Por ello, hemos desarrollado procedimientos generales para estimar su impacto energético y estructural a partir de información experimental acerca del acoplamiento cinético de los efectos de la temperatura y desnaturalizantes químicos. Finalmente, hemos proporcionado evidencia experimental de que las altas barreras energéticas asociadas a la estabilidad cinética pueden conducir a la existencia de estados nativos alternativos con propiedades diferentes. De hecho, creemos que nuestros resultados sugieren procedimientos generales para acceder a tales estados nativos "alternativos" y que el estudio sistemático de éstos podría tener implicaciones de interés para la comprensión de enfermedades relacionadas con amiloides y priones.

En segundo lugar, hemos analizado sistemáticamente la posibilidad de utilizar en ingeniería de proteínas la información evolutiva contenida en las bases de datos de secuencias, bases que, actualmente, crecen de forma exponencial. Para ello hemos planteado un modelo simple acerca de la estadística de la selección natural purificadora y, en base a este modelo y al análisis de alineamientos de secuencias, hemos construido una biblioteca combinatorial de variantes de una proteína modelo adecuada (tiorredoxina de *E. coli*). El estudio experimental de esta biblioteca demuestra la posibilidad de conseguir modulaciones simultáneas y a gran escala de diversas propiedades de las proteínas, propiedades relacionadas tanto con la estabilidad (cinética y termodinámica) como con la función biológica (actividad, mecanismo catalítico). Resulta casi innecesario resaltar el enorme interés biotecnológico de la optimización simultánea de proteínas en términos de varias propiedades ortogonales y, consecuentemente, el gran potencial de la estrategia basada en la información evolutiva que hemos desarrollado en este trabajo.

## 4-APENDICE 1

Artículo enviado para  
publicación

"Super-proteins from fitness-threshold  
selection statistics"

## **INTRODUCCIÓN APÉNDICE 1**

En principio, es posible asignar un valor de "fitness", entendido como la tasa reproductiva del organismo, a cada una de las secuencias de una proteína dada. El resultante "fitness landscape" puede considerarse como la convolución de dos factores: uno que va de genotipo a fenotipo con la que va de fenotipo a "fitness". A pesar de esta complejidad intrínseca, las interacciones con otras moléculas, la estabilidad, etc.. En cualquier caso, la relación específica entre fitness, espacio de secuencias, evolución y las propiedades moleculares de la proteína no está bien entendida, un hecho que dificulta los esfuerzos para aprovechar el enorme potencial de explotar las exponencialmente crecientes bases de datos en el diseño de proteínas y organismos. Mostramos aquí que una simple hipótesis evolutiva sobre la estadística de la selección natural que "purifica" por encima de un determinado límite selectivo es operacional y lleva a una optimización de las proteínas en múltiples características. Para ello obtenemos variantes de tiorredoxina de E.coli que muestran una optimización simultánea y de gran magnitud en estabilidad, cinéticas de plegamiento/desplegamiento y actividad oxidoreductasa. La espectroscopia de fuerza atómica de moléculas de tiorredoxina individuales también muestra una modulación de gran magnitud del mecanismo catalítico (eliminación de la ruta de alta fuerza y un incremento de la ruta de baja fuerza, que es dependiente de interacciones proteína-proteína) en una dirección que la hace similar a la tiorredoxina humana. En conjunto, estos hallazgos señalan a procedimientos generales (selección *in vitro* de varias características simultáneamente basandose en información bioinformática) para modificar a voluntad varias propiedades proteicas, un importante éxito para las aplicaciones biomédicas y tecnológicas. Dado que la modulación obtenida de este modo probablemente tenga un efecto en el "fitness" del organismo, también anticipamos aplicaciones en los campos que involucran diseño de organismos (tecnología microbiana y biología sintética).

# Super-proteins from fitness-threshold selection statistics

David Rodriguez-Larrea<sup>1</sup>, Raul Perez-Jimenez<sup>2</sup>, Julio M. Fernandez<sup>2</sup> &  
Jose M. Sanchez-Ruiz<sup>1</sup>

<sup>1</sup>Facultad de Ciencias, Departamento de Quimica-Fisica, Universidad de Granada, 18071-Granada, Spain. <sup>2</sup> Department of Biological Sciences, Columbia University, New York, New York 10027, USA.

Correspondence and requests for materials should be addressed to  
J.M.S.-R. ([sanchezr@ugr.es](mailto:sanchezr@ugr.es))

It is in principle possible to assign a fitness value (organism reproductive rate) to every sequence of any given protein. The resulting fitness landscape can be regarded as the convolution of two mappings: one from genotype to phenotype with another mapping from phenotype to fitness<sup>1</sup>. Despite this intrinsic complexity, we may expect organism fitness to depend on protein properties bearing upon catalysis, interactions with other molecules, stability, et cetera. However, the specific relation between fitness, sequence space, evolution and protein molecular properties is not understood, a fact which hampers efforts to tap the enormous potential for protein and organism engineering contained in the exponentially-growing sequence databases. Here we show that a simple evolutionary hypothesis on the statistics of purifying natural selection over a fitness threshold is operational and leads to protein multi-feature optimization. We thus obtain variants of *E. coli* thioredoxin showing simultaneous, large-scale optimizing modulations in stability, folding/unfolding kinetics and oxidoreductase activity. Single-molecule force-clamp spectroscopy also reveals a large-scale modulation of the catalytic mechanism (elimination of the high-force pathway and enhancement of the low-force, protein-interaction-dependent pathway) in the direction of making it similar to that in human thioredoxin. Overall, these findings point to general procedures (bioinformatics-focused in-vitro multi-feature selection) to modify at will several protein properties, an important goal for technological and biomedical applications. Since protein modulation achieved in this way will likely have an effect on organism fitness, we also anticipate applications in fields that involve organism engineering (microbial biotechnology, synthetic biology).

The evolutionary hypothesis explored in this work involves four assumptions, which are stated and briefly elaborated below.

- 1) Purifying selection occurs frequently during protein molecular evolution; i.e., mutations are not fixed when they modify protein properties in such a way that organism fitness is seriously compromised. In other words, a mutation in a given protein will not be accepted if it causes a relevant property of the protein to cross a critical fitness threshold (under the usual environment conditions for the organism or under frequently encountered stress situations). We believe this assumption (together with assumption 2 below) to be

generally consistent with recent experimental studies on the fitness landscape of randomly drifting proteins<sup>2</sup>.

2) A protein property may be affected by many different mutations (at many different positions) and, therefore, whether a given mutation is fixed or rejected may depend on the previous mutational background, even if the mutation has a comparatively constant effect on the property (i.e., even if the effect of the mutation is not affected by mutations at other positions). Accordingly, we may view acceptance/rejection of mutations as a stochastic process with probabilities related to the mutation effects on protein properties (an illustrative Monte Carlo simulation is included in the Supplementary Information). For instance, a mutation that strongly worsens a property will only be occasionally accepted (only when previous mutations have improved the property so that the mutation does not cause the property value to cross its fitness threshold) and, consequently, the opposite mutation (which improves the property) will be found much more often.

3) Natural purifying selection operates on the basis of many different protein properties. For instance, stability is certainly important in protein evolution, as vividly shown by the recently reported correlation between the denaturation temperatures of ancestral, resurrected proteins and the cooling trend for the ancient oceans<sup>3</sup>. However, properties related to function are also essential, as illustrated by the recent demonstration of evolutionary constraints associated to co-enzyme use<sup>4</sup>. The important general implication in the context of this work is the following: the statistics of fixation/rejection may identify mutations with significant effects on the protein molecular properties which are important for fitness, but it does not provide clues about which specific properties are being modulated by a given mutation (see the illustrative Monte Carlo simulation in Supplementary Material) . Actually, this explains why correlations between sequence statistics and mutation effects on a single property are weak<sup>5,6</sup> and why consensus approaches to protein stabilization often yield inconsistent results<sup>7</sup>.

4) The mechanism implied by points 1-3 above is taken to be operative, not only at the population genetics level, but also in longer evolutionary time-scales (at least, in some kind of average way). Accordingly, the relevant statistics of fixation/rejection can

be estimated from sequence alignments including proteins from different species.

The hypothesis embodied in assumptions 1-4, albeit simplistic and speculative, has an important merit: it is operational, in the sense that it leads to predictions that can be tested. In particular, it points to the possibility of modulating/optimizing proteins in terms of several properties on the basis of statistical information derived from sequence alignments. However, since it is not possible to establish from the sequence alignment analysis alone a correspondence between mutation and protein property, a test of the hypothesis requires the construction of a library of variants and its screening for the values of several protein properties. Library screening should also bypass uncertainties associated to the analysis of the alignments, to the possibility of non-additivity of mutation effects and, most importantly, to the fact that some mutations may be modulating properties that are important *in vivo* but for which *in vitro* assays have yet to be developed. Still, it is clear that the library must be very small in order to provide a convincing test of the hypothesis. We demonstrate in this work multi-feature optimization of a protein on the basis of the screening of a very small number of variants (about 20) from a small sequence-statistics-based combinatorial library (256 variants).

The library was constructed on the basis of an alignment of a hundred sequences<sup>8</sup> belonging to proteins from proteobacteria (Fig.1a). For each position in the *E. coli* thioredoxin sequence, a ratio of non-wt to wt frequencies was calculated. We recently showed<sup>8</sup> that the three highest-ranked mutations (A22P, I23V, P68A) lead to a triple-mutant variant with an activity (in the insulin reduction assay) similar to that of the wt protein from *E. coli*, but with a significantly enhanced stability. We use this triple mutant (which we shall refer to as V3) as background for a combinatorial library of thioredoxin genes which includes (besides A22P, I23V, P68A) all possible combinations of the 8 next-ranked mutations (D10A, D47A, Q50A, Q62A, Y70F, G74S, E85Q, A87V), making a total of 256 different genes. We randomly selected 23 genes for strain transformation, protein expression and purification. The 23 resulting variants were characterized (Figs. 1b and 1c) in terms of thermal stability (differential scanning calorimetry: DSC), kinetic stability (unfolding branches of the guanidine chevron plots) and activity (insulin reduction assay). Most variants show a

denaturation temperature value higher than that of the background V3 variant (and much higher than that of wt-thioredoxin). Kinetic stability is also greatly enhanced as indicated by the much slower unfolding rates of most of the variants. The two features are not independent (see correlation in Fig. 1b) suggesting natural selection for kinetic stability (i.e., a fitness-threshold associated to kinetic stability), which is reflected in thermodynamic stability and equilibrium denaturation temperature<sup>5</sup>. Remarkably, many of the studied variants display, in addition to a higher stability, a significantly enhanced catalytic activity. Moreover, stability and activity seem to be independent features (no significant correlation between them: see Fig. 2c). This suggests that a fitness threshold is associated to, at least some of the *in vivo* activities of thioredoxin<sup>9,10</sup> and that such activities are to some extent represented by the *in vitro* insulin-reduction assay.

All thioredoxin variants proceeding from the combinatorial library have a histidine-tag attached for ease of purification. Our results (Fig. 1) support that the presence of the His-tag has a comparatively small effect on stability and activity. Nevertheless, we deemed convenient to carry out a detailed characterization of a high-stability, high-activity variant obtained without a His-tag. For this purpose, we chose D10A/Q50A/G74S/E85Q/A87V (over the V3 background), a variant which, for reasons that become clear below, will be referred to as super-thioredoxin. Super-thioredoxin is identical to one of the "best" library variants, except that the His-tag and a Y70F mutation are not present (since an independent-mutation-effect analysis suggests that Y70F may decrease activity: see Supplementary Information). Super-thioredoxin shows enhanced denaturation temperature and activity when compared with the library variants (see Figs 1 and 2). In addition, when compared with wt-thioredoxin, it shows: 1) a denaturation temperature value about 20 degrees higher (Fig. 2a); 2) unfolding rates about 15000-fold slower (Fig. 2b); 3) Folding rates about ten-fold faster when folding is not determined by "roll-over processes" (Fig. 2b). 4) enzymatic activity (as measured by the insulin-reduction assay) about 80% higher (Fig. 2c).

We have further probed the enzymatic catalysis by super-thioredoxin using single-molecule, force-clamp spectroscopy<sup>11,12</sup>. In this approach, mechanical force is applied to a disulfide bond protein substrate and the reduction of these bonds by individual enzyme

molecules is monitored. Pioneering studies using this technology<sup>12</sup> indicated the existence of two separate pathways for disulphide bond reduction by wt-thioredoxin from *E. coli*. Pathway I (the major pathway at low force) involves a shortening of the substrate polypeptide by  $\sim 0.8 \text{ \AA}$  and likely corresponds to a mechanism for reduction that implies specific protein-protein (i.e., thioredoxin-substrate polypeptide) interaction and a reorientation of the substrate disulfide bond. Pathway II (the major pathway at high force) involves an elongation of the of the substrate disulfide bond by  $\sim 0.2 \text{ \AA}$ <sup>10</sup>. Fig. 3 shows the reduction rate versus force profile for super-thioredoxin determined in this work from single-molecule force-clamp experiments. Compared with wt-thioredoxin from *E. coli*, pathway I is about two-fold enhanced. The most surprising result, however, is the absence of pathway II (instead, a small contribution from an essentially force-independent pathway is revealed). Previous studies<sup>12</sup> have shown that pathway II is also greatly diminished in human thioredoxin and recent work (manuscript in preparation) supports that this is in general de case with eukaryotic thioredoxins, while a significant pathway II appears to be a general feature of prokaryotic thioredoxins. Elimination of pathway II in super-thioredoxin makes the catalytic mechanism of a prokaryotic thioredoxin similar to that of an eukaryotic thioredoxin and, therefore, highlights the enormous potential for protein modulation of the approach proposed in this work.

To summarize, we have achieved multi-feature protein optimization on the basis of a simple evolutionary hypothesis and the screening of a very small library suggested by a straightforward bioinformatics analysis of sequence alignments. This approach should be easily extended and refined (screening of larger libraries constructed on the basis of more complex bioinformatics analyses) and may have a strong impact on technological and biomedical applications of proteins which, very often, require their optimization with respect to several features: stability, solubility, activity, interactions with other molecules, serum half-life, immunogenicity, et cetera<sup>13</sup>. Certainly, the methodology is limited to features that are relevant for organism fitness and introduction of novel (non-natural) functions and activities, for example, must rely on computational design and in vitro directed evolution<sup>14-17</sup>. On the other hand, we anticipate that the approach proposed here may find applications in fields that involve organism engineering (such as microbial biotechnology and synthetic

biology), since the protein variants obtained on its basis may be expected to affect organism fitness. For instance, a recent proteomic analysis<sup>10</sup> has identified 80 proteins associated with thioredoxin and has involved thioredoxin in at least 26 cellular processes; likely, some of the most interesting consequences of the thioredoxin modulation achieved in this work will be revealed by in vivo fitness studies (work in progress). Finally, from the viewpoint of protein molecular evolution, our success in achieving multi-feature optimization supports the underlying assumptions of the approach used, in particular, the relevance and statistical nature of purifying natural selection associated to a wide range of protein molecular properties.

### **METHODS SUMMARY**

The library of thioredoxin variant sequences were assembled by overlapping synthetic oligonucleotides containing degenerate bases at designed mutation sites, based on the method described by Stemmer<sup>18</sup>. Thermal stability was determined by differential scanning calorimetry as previously described<sup>8,19</sup>. Folding-unfolding kinetics were characterized by following the time dependence of the fluorescence intensity after suitable guanidine-concentration jumps, as previously described<sup>8</sup>. Oxidoreductase activity was determined using the insulin-aggregation assay described by Holmgren<sup>20</sup>. Single-molecule atomic force microscopy experiments were carried out as recently described<sup>12</sup>. Very briefly, we used as protein substrate a polypeptide made of eight repeats of the I27 domain of human cardiac titin with engineered cysteines. A custom-built atomic force microscope controlled by an analogue proportional-integral-derivative feedback system<sup>21</sup> was employed to follow the reduction catalyzed by thioredoxin of individual disulfide bonds under a stretching force applied to the substrate.

1. Nowak, A. M. Evolutionary dynamics. Exploring the equations of life. Harvard University Press, Cambridge, MA, chapter 3 (2006).
2. Bershtein, S., Segal, M., Beterman, R., Tokuriki, N. & Tawfik, D. S. Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. Nature 444, 929-932 (2006).

3. Gaucher, E. A., Govindarajan, S. & Ganesh, O. K. Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature* 451, 704-707 (2008).
4. Miller, S. P., Lunzer M. & Dean, A. M. Direct demonstration of an adaptive constraint. *Science* 314, 458-461 (2006).
5. Godoy-Ruiz, R., Ariza, F., Rodriguez-Larrea, D., Perez-Jimenez, R., Ibarra-Molero, B. & Sanchez-Ruiz, J. M. Natural selection for kinetic stability is a likely origin of correlations between mutational effects on protein energetics and frequencies of amino acid occurrences in sequence alignments. *J. Mol. Biol.* 362, 966-978 (2006).
6. Sanchez, I. F., Tejero, J., Gomez-Moreno, C., Medina, M. & Serrano, L. Point mutations in protein globular domains: contributions from function, stability and misfolding. *J. Mol. Biol.* 363: 422-432 (2006).
7. Lehmann, M., Loch, C., Middendorf, A., Studer, D., Lassen, S. F., Pasamontes, L., van Loon, A. P. & Wyss, M. The consensus concept for thermostability engineering of proteins: further proof of concept. *Protein Eng.* 15, 403-411 (2002).
8. Pey, A. L., Rodriguez-Larrea, D., Bomke, S., Dammers, S., Godoy-Ruiz, R., Garcia-Mira, M. & Sanchez-Ruiz, J. M. Engineering proteins with tunable thermodynamic and kinetic stabilities. *Proteins* 71, 165-171 (2008).
9. Arnér, E. S. & Holmgren, A. Physiological functions of thioredoxin and thioredoxin reductase. *Eur. J. Biochem.* 267, 6102-6109 (2000).
10. Kumar, J. K., Tabor, S. & Richardson C. C. Proteomic analysis of thioredoxin-targeted proteins in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 101, 3759-3764 (2004).
11. Wiita, A. P., Ainaravapu, S. R. K., Huang, H. H. & Fernandez, J. M. Force-dependent chemical kinetics of disulfide bond reduction observed with single-molecule techniques. *Proc. Natl. Acad. Sci. USA* 103, 7222-7227 (2006).
12. Wiita, A. P., Perez-Jimenez, R., Walther, K. A., Gräter, F., Berne, B. J., Holmgren, A., Sanchez-Ruiz, J. M. & Fernandez, J. M. Probing the chemistry of thioredoxin catalysis with force. *Nature* 450, 124-127 (2007).
13. Lazar, G. A., Marshall, S. A., Plecs, J. J., Mayo, S. L. & Desjarlais, J. R. Designing proteins for therapeutic applications. *Curr. Opin. Struct. Biol.* 13, 513-518 (2003).
14. Bolon D. N. & Mayo, S. L. Enzyme-like proteins by computational design. *Proc. Natl. Acad. Sci. USA* 98, 14274-14279 (2001).

15. Alert, M., Rizk, S. S., Looger, L. L. & Hellinga, H. W. Computational design of receptors for an organophosphate surrogate of the nerve agent soman. *Proc. Natl. Acad. Sci. USA* 101, 7907-7912 (2004).
16. Seelig, B. & Szostak, J. W. Selection and evolution of enzymes from a partially randomized non-catalytic scaffold. *Nature* 448, 828-831 (2007).
17. Röthlisberger, D., Khersonsky, O., Wollacott, A. M., Jiang, L., DeChancie, J., Bekter, J., Gallaher, J. L., Althoff, E. A., Zhanghellini, A., Dym, O., Albeck, S., Houk, K. N., Tawfik, D. S. & Baker, D. Kemp elimination catalysts by computational enzyme design. *Nature*, published online March-19 (2008).
18. Stemmer, W. P. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* 370, 389-391 (1994).
19. Georgescu, R. E., Garcia-Mira, M. M., Tasayco, M. L. & Sanchez-Ruiz, J. M. Heat capacity analysis of oxidized Escherichia coli thioredoxin fragments (1-73, 74-108) and their noncovalent complex. Evidence for the burial of apolar surface in protein unfolded states. *Eur. J. Biochem.* 268, 1477-1485 (2001).
20. Holmgren, A. Thioredoxin catalyzes the reduction of insulin disulfides by dithiothreitol and dihydrolipoamide. *J. Biol. Chem.* 254, 9627-9632 (1979).
21. Schlierf, M., Li, H. & Fernandez, J. M. The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques. *Proc. Natl. Acad. Sci. USA* 101, 7299-7304 (2004).

**Acknowledgments.** This work was supported by an NIH grant to J.M.F. and Feder Funds, a grant from the Spanish Ministry of Science and Education and a grant from the "Junta de Andalucía" to J.M.S.-R.

**Author Contributions.** D.R.-L. designed, performed and analyzed experiments addressed at determining stability, activity and folding/unfolding kinetics. D.L.-R. also performed all protein purifications. R.P.-J. and J.M.F. designed, performed and analyzed the single-molecule atomic-force microscopy experiments. J.M.S.-R. designed research and wrote the paper. All authors discussed the manuscript.

**Author Information.** Correspondence and requests for materials should be addressed to J.M.S.-R. (sanchezr@ugr.es).

## **METHODS**

**Variant library generation.** The library of TrxA variant sequences were assembled by overlapping synthetic oligonucleotides containing degenerate bases at designed mutation sites, based on the method described by Stemmer<sup>18</sup>. The assembled sequences were amplified by PCR. After restriction digestion the PCR products were ligated into the NdeI and XhoI sites of the pET30a(+) vector. Ligation reaction was then transformed into *E. coli* DH10B cells. Plasmid sequencing in a significant number of colonies showed that more than the 50 % of the sequences actually contained Trx genes. The genes also coded for a His-tag and a thrombin recognition sequence at the N-terminal end. 23 different variants were randomly selected for protein purification and characterization. Their sequences are given in Table I (Supplementary Information).

**Purification of proteins from the variant-library.** Overexpressing cells BL-21 DE3 were transformed with the amplified library. Isolated colonies were grown in 10 ml LB media in the presence of antibiotic. Aliquots of all these cultures were stored in glycerol at -80 °C. Cultures were induced for protein expression with IPTG to an O.D.<sub>600</sub> equal to 0.6. After 6 h of incubation, cultures were checked for Trx overexpression by SDS-PAGE. Those colonies which overproduced Trx were first grown for plasmid purification and gene sequencing, and, subsequently, for protein production, according to the following procedure. Cells were grown in 250 ml of LB with kanamycin to an O.D.<sub>600</sub> of 0.6 were expression was induced with IPTG. After incubation during 6 hours, cultures were centrifuged at 10.000 x g for 10 minnutes and pellets were resuspended in binding buffer (containing 10 mM imidazole), and sonicated (4 cycles of 45 seconds). Lisates were centrifuged and supernatants were applied to an affinity His-GraviTrap (GE Healthcare) column. The column was washed with excess of binding buffer and the protein was eluted by applying 3 ml of elution buffer (containing 500mM imidazole). Buffer was then changed to Hepes 5 mM, pH 7 (the buffer in which all experiments were conducted) by using a Fast Desalting (GE Healthcare) column. Finally,

protein solutions were heated to 75 °C for 2 hours and any aggregated material was removed by centrifugation (heating at 75 °C does not affect our thioredoxin variants which show very high thermostability). Proteins were found to be pure by SDS-PAGE.

**Purification of WT-thioredoxin, the V3 variant and super thioredoxin.** His-tag versions of wt-thioredoxin and the V3 variant were obtained as explained above for the library variants. In addition, wt-thioredoxin, V3 and super-thioredoxin were also prepared without a His-tag as described below:

Plasmids pET 30a(+) encoding wild-type, V3 and super-thioredoxin were transformed into E. coli BL21 strain for protein overexpression. Cells were grown, starting from single colonies, at 37°C in LB with kanamycin. At O.D.<sub>600</sub> of 0.6, protein overexpression was induced with IPTG. After 6 h, cultures were centrifuged, cell pellets were resuspended in 1 mM EDTA, 30 mM TRIZMA buffer, pH 8.3, and lysed using a French press. The cell debris was centrifuged and the supernatant was collected and stirred with streptomycin sulfate (10% w/v) at 4°C overnight to precipitate nucleic acids. After centrifugation, the supernatant was loaded onto a 2 L Sephacryl S-100 high resolution (Amersham Pharmacia Biotech) gel filtration column equilibrated in 1 mM EDTA, 30 mM TRIZMA buffer, pH 8.3. Thioredoxin fractions were identified by SDS-PAGE, pooled, and applied to a 250 mL Fractogel EMD DEAE (M) (Merck) ion exchange column equilibrated in 1 mM EDTA, 30 mM TRIZMA buffer, pH 8.3. The protein was eluted by a salt gradient (between 0 and 0.5 M NaCl). The proteins were pure as determined by SDS-PAGE.

**Activity assays.** Activity of wt thioredoxin and different variants was determined by measuring the rate of insulin reduction spectrophotometrically at 650 nm as turbidity formation from the precipitation of the free insulin B chain<sup>20</sup>. 50 ml of 10 mg/ml insulin solution were added to the activity buffer (potassium phosphate 0.1 M, 2mM EDTA). Different amounts of each mutant were added to the cuvette and the reaction was started by addition of 10 ml DTT 0.1M, which causes thioredoxin reduction. The activity was determined as the maximum value of the slope in a plot of absorbance at 650 nm versus time (see representative profiles in Fig. 4: Supplementary material). Blank corrections were performed but were invariably small. For each thioredoxin variant, several activity determinations were carried out

at different protein concentrations. Specific activity values were derived from the linear dependence of activity with protein concentration (see Fig. 2c).

**Folding-Unfolding Kinetics.** Folding-unfolding kinetics were studied at 25 °C by following the time dependence of the protein fluorescence emission at 360 nm after suitable guanidine concentration jumps in a manner similar to that described<sup>8</sup>. Apparent folding-unfolding rate constants ( $k$ ) were calculated from the fittings of a first-order rate equation ( $I = I_{\infty} - A e^{-kt}$ ) to the experimental fluorescence intensity versus time profiles. These fits were excellent. For most of the library variants, only the unfolding branches of the Chevron plots were determined (representative examples are shown in Fig. 5: Supplementary material). Linear fits to these unfolding branches were used to obtain the unfolding rate constants at 4.2 M guanidine shown in Fig. 1b. For wt-thioredoxin and super-thioredoxin, "complete chevron plots (i.e., including folding and unfolding branches) were determined. Significant roll-over was observed in the folding branches, perhaps due proline isomerization effects. Interestingly, there is agreement between the roll-over regions of wt-thioredoxin and super-thioredoxin (see Fig. 2b). Complete chevron plots were fitted with a two-state kinetic model as previously described<sup>8</sup>:

$$\ln k = \ln(k_U + k_F) \tag{1}$$

$$\ln k_F = \ln k_{1/2} + \frac{m_F}{RT} (C - C_{1/2}) \tag{2}$$

$$\ln k_U = \ln k_{1/2} + \frac{m_U}{RT} (C - C_{1/2}) \tag{3}$$

where  $k_U$  and  $k_F$  are the unfolding and folding rate constants,  $m_U$  and  $m_F$  describe the guanidine-concentration ( $C$ ) effect on the activation free energies for unfolding and folding ( $m = -\partial\Delta G/\partial C$ ), and  $k_{1/2}$  is the value of both  $k_F$  and  $k_U$  at the midpoint guanidine concentration ( $C_{1/2}$ ). For the purpose of these two-state kinetic fits, the roll-over sections were excluded. However, in the case of wt-trx, a rather short folding branch is left after roll-over exclusion (see Fig. 2b) and we deemed necessary to fix the folding  $m$ -value for wt.trx in the value obtained for super-trx. The values derived from the fittings are given in Table II (Supplementary Information).

**Differential Scanning Calorimetry.** DSC experiments were carried out using a capillary VP-DSC calorimeter from MicroCal (Northampton, USA). DSC experiments were performed in 5 mM Hepes buffer, pH 7, at a scan rate of 2.5 K/min. Calorimetric cells were kept under an excess pressure to avoid degassing during scan and to allow scans to proceed about 100 °C without boiling. Additional details about the DSC experiments can be found in previous publications<sup>8,19</sup>. Reheating runs were routinely performed to check reversibility. Remarkably, super-thioredoxin denaturation shows a significant degree of reversibility (about 70%), despite of the fact that the first scan was terminated at a high temperature (120 °C). In fact, reversibility was found to be somewhat lower for the library variants, likely due to the presence of the His-tag. DSC profiles for the library variants were mainly used to obtain denaturation temperature values (Figs. 1b and 1c). In the case of wt-thioredoxin and super thioredoxin (prepared without his-tags) we carried out two-state fits (see Fig. 2a) which yielded denaturation temperature and enthalpy values of 89.3 °C and 445 kJ/mol (wt-thioredoxin) and 107.3 °C and 535 kJ/mol (super-thioredoxin). The difference between the denaturation enthalpy values indicates a denaturation heat capacity of about  $5 \text{ kJ}\cdot\text{K}^{-1}\cdot\text{mol}^{-1}$ , which is consistent with previous determinations<sup>19</sup>.

**Single-molecule atomic force microscopy.** We followed the procedure recently described<sup>12</sup>. Briefly, a custom-built atomic force microscope controlled by an analogue proportional-integral-derivative feedback system<sup>21</sup> was employed. The buffer used in all experiments was 10 mM Hepes, 150 mM NaCl, 1mM EDTA, pH 7.2 and contained 2 mM NADPH and 50 nM thioredoxin reductase. Single  $(\text{I27}_{\text{SS}})_8$  protein molecules were stretched by first pressing the cantilever on the coverslide at a constant force of 800 pN for 3s, then retracting to a constant force of 165 pN for 400 ms during the unfolding pulse. The indicated test-pulse force was applied for ~5 s. We summed and normalized the test-pulse portions of numerous recordings that contained only disulphide reduction events to obtain the experimental  $r$  value. Further details and a description of the model fitting to the experimental  $r$  versus force profiles can be found in Wiita et al.<sup>12</sup>.

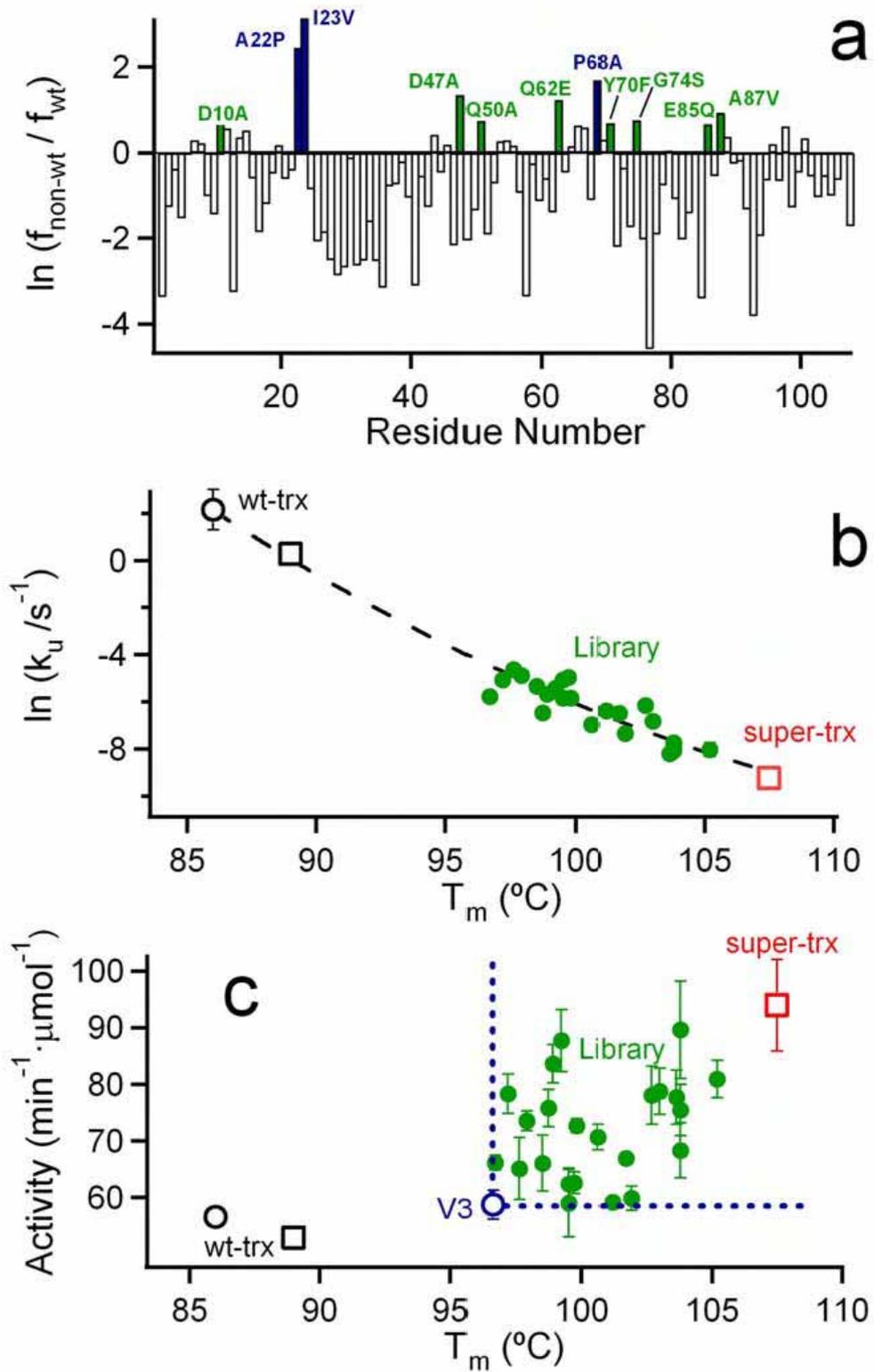
## Legends to the figures

Figure 1 | Sequence-statistics-based construction and experimental analysis of a combinatorial library of thioredoxin variants. a, We used an alignment including sequences from a hundred proteins belonging to proteobacteria, derived from a BLAST search of the UniProt/TrEMBL data base using the E. coli thioredoxin sequence as query<sup>8</sup>. For each position in the E. coli thioredoxin molecule, the ratio non-wt to wt frequencies was calculated (wt meaning here the residue present in E. coli thioredoxin and non-wt meaning the highest frequency residue which is not the wt one). The mutations introduced in the background variant (V3) and the eight positions selected for library construction are shown in blue and green, respectively. b, Plot of logarithm of unfolding rate at 4.2 M guanidine versus denaturation temperature including wt-thioredoxin, the variant V3, 23 variants selected from the combinatorial library and super-thioredoxin. The dashed line has no theoretical meaning and is shown to guide the eye. c, Plot of catalytic activity (insulin reduction assay) versus denaturation temperature for the same protein variants included in panel b. The dashed blue lines are meant to guide the eye to the fact that, when compared with V3, the library variants (green) have enhanced activity and denaturation temperature. In both panels (b and c), standard errors are not shown when they are smaller than the size of the marker, circles are used for protein variants having a His-tag and squares for variants without it.

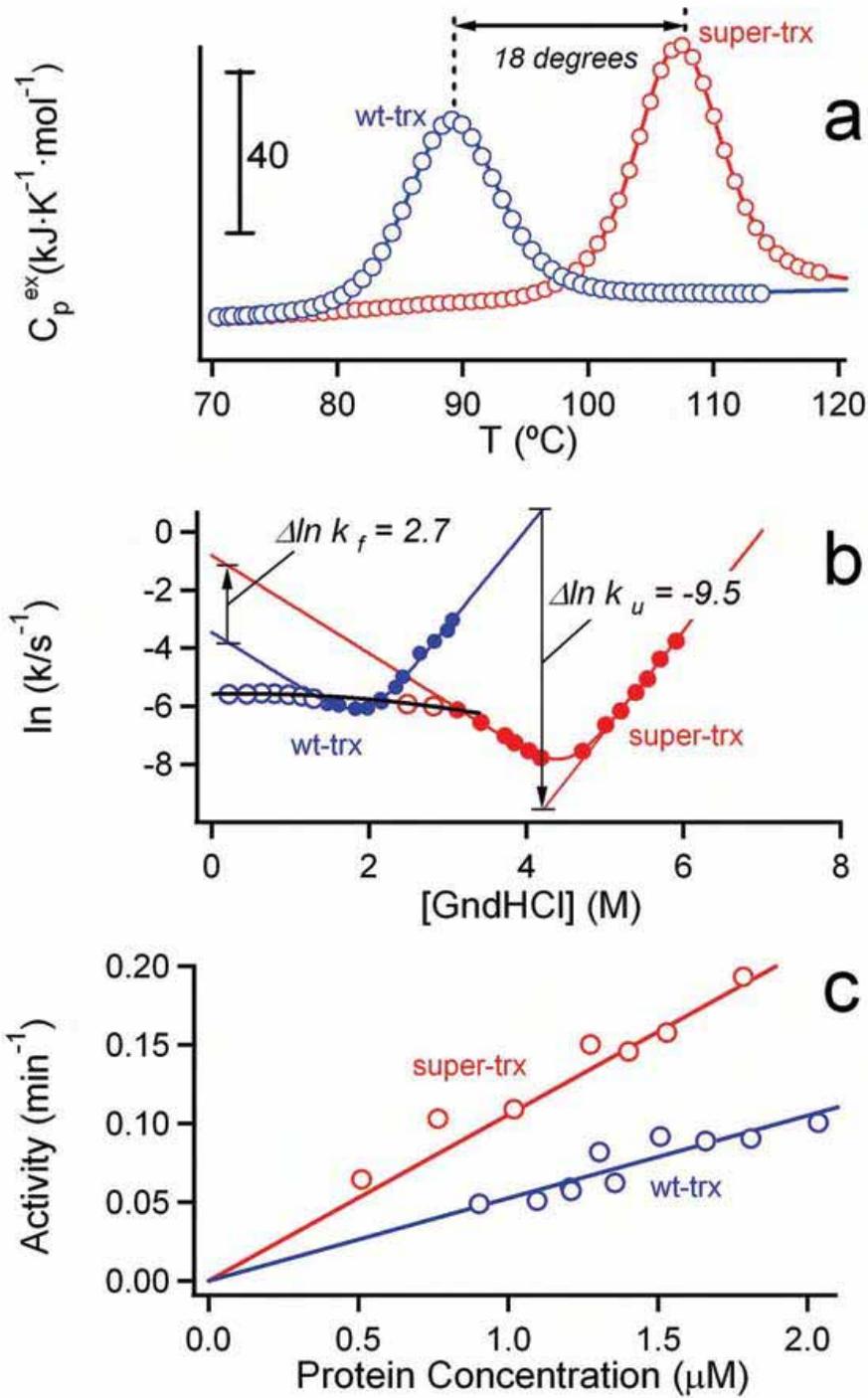
Figure 2 | Wild-type-thioredoxin versus super-thioredoxin: activity, stability and folding-unfolding kinetics. a, Differential scanning calorimetry thermograms for wt-thioredoxin and super-thioredoxin, showing a difference of 18 degrees in denaturation temperature. The continuous lines are the best fits of the two-state equilibrium model (see "Full Methods" for details). b, Plots of folding-unfolding rate constant versus guanidine concentration (chevron plots) showing a ~15000-fold difference in unfolding rates. The red and blue continuous lines are the best fits of a two-state kinetic model, excluding the roll-over regions (marked with a continuous black line). When not-determined by "roll-over processes", folding rate appears to be about 10-fold enhanced in super-thioredoxin (see "Full Methods" for details). c,. Plot of catalytic activity

(insuline reduction assay) versus protein concentration showing a ~80% enhancement in super-thioredoxin.

Figure 3 | Wild-type-thioredoxin versus super-thioredoxin: single-molecule atomic force microscopy. Rate constant for reduction of a disulfide protein substrate versus the force applied to the substrate. Data are derived from single-molecule force-clamp experiments. The continuous lines are the best fits of the two-pathway model<sup>11</sup>. The model is shown in the upper right corner of the figure, together with the equations giving the protein-concentration and force dependencies of the rates for the several steps. The parameter values derived from the fits are shown in the inserted table.

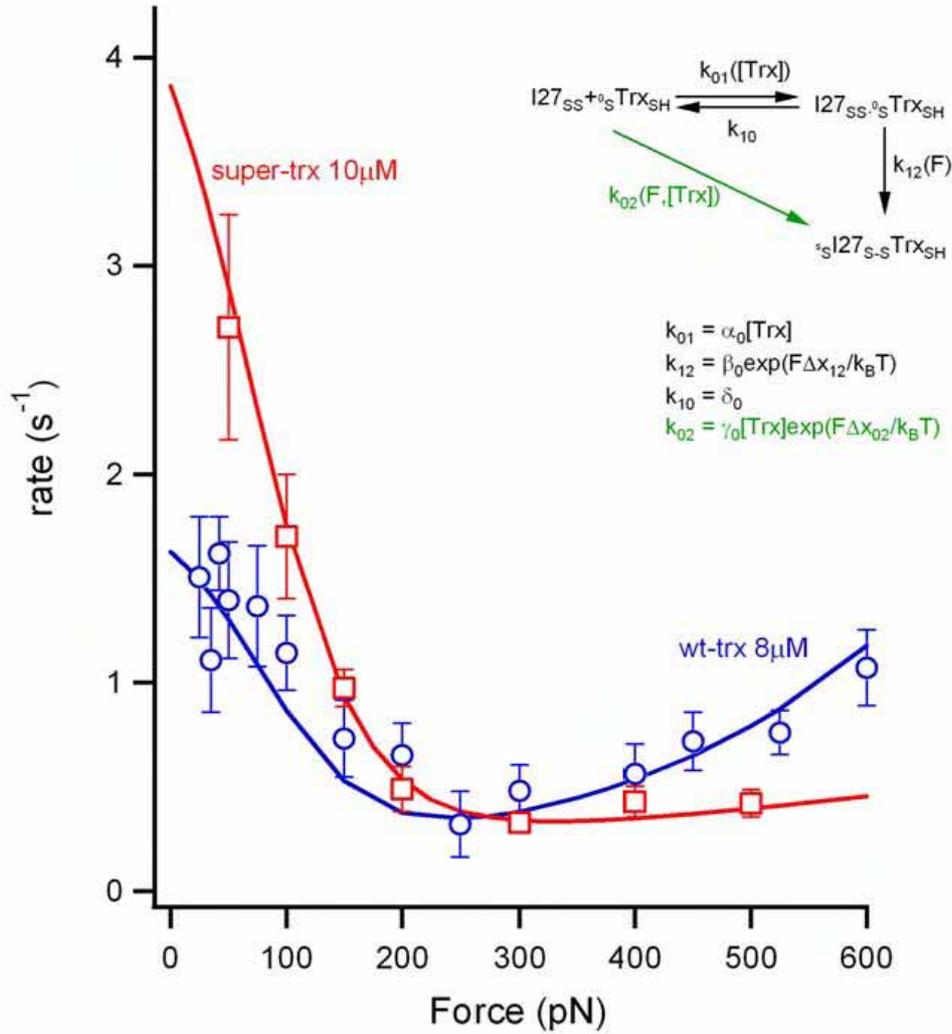


J.M.Sanchez-Ruiz, Figure 1  
 Figura AP 1.1 (Figure 1 EN EL TEXTO)



J.M.Sanchez-Ruiz, Figure 2

Figura AP1.2 (Figure 2 EN EL TEXTO)



Protein	$\alpha_0$ ( $\mu\text{M}^{-1} \cdot \text{s}^{-1}$ )	$\beta_0$ ( $\text{s}^{-1}$ )	$\gamma_0$ ( $\mu\text{M}^{-1} \cdot \text{s}^{-1}$ )	$\delta_0$ ( $\text{s}^{-1}$ )	$\Delta x_{12}$ ( $\text{\AA}$ )	$\Delta x_{02}$ ( $\text{\AA}$ )
WT-Trx	$0.22 \pm 0.03$	$21 \pm 6$	$0.019 \pm 0.008$	$3.2 \pm 0.8$	$-0.79 \pm 0.09$	$0.17 \pm 0.02$
Super-Trx	$0.43 \pm 0.09$	$28 \pm 2$	$0.038 \pm 0.012$	$4.1 \pm 0.5$	$-0.80 \pm 0.05$	$0.06 \pm 0.02$

J.M. Sanchez-Ruiz, Figure 3

Figura AP1.3 (Figure 3 EN EL TEXTO)

## 5-APENDICE 2

# DSC y Fluorescencia de proteínas

## **5.1-Calorimetría Diferencial de Barrido**

La calorimetría diferencial de barrido (DSC, Differential Scanning Calorimetry) es una técnica que permite la caracterización energética de cambios conformacionales inducidos por temperatura en polímeros biológicos y agregados supramoleculares. Es especialmente útil en el estudio de macromoléculas biológicas de las que no es posible, normalmente, disponer de grandes cantidades. El proceso de medida consiste en el registro del calor absorbido o emitido por una disolución a medida que es calentada a velocidad constante y se compara al solvente puro. De este modo se obtiene la capacidad calorífica aparente del soluto en función de la temperatura. Es interesante destacar que el DSC es la única técnica que permite acceder a la función de partición del sistema en estudio.

### **5.1.1- Características Técnicas**

Durante los años 70 se desarrolló la instrumentación necesaria para llevar a cabo este tipo de medidas. En términos generales, los microcalorímetros diferenciales de barrido comparten las siguientes características:

-Funcionan de modo adiabático.

-Las células de medida (muestra y referencia) son gemelas y se encuentran inmersas en la cámara adiabática de modo fijo, dándose una mayor reproducibilidad de los datos experimentales.

-La medida es diferencial. Cuando un proceso endotérmico tiene lugar en la célula de muestra, por ejemplo la desnaturalización de una proteína, la temperatura en dicha célula se retrasa con respecto a la de referencia. El sistema de retroalimentación del calorímetro suministra entonces una potencia extra proporcional a la capacidad calorífica de la solución de proteína (tomando el buffer como referencia), con objeto de mantener la temperatura de ambas células idéntica.

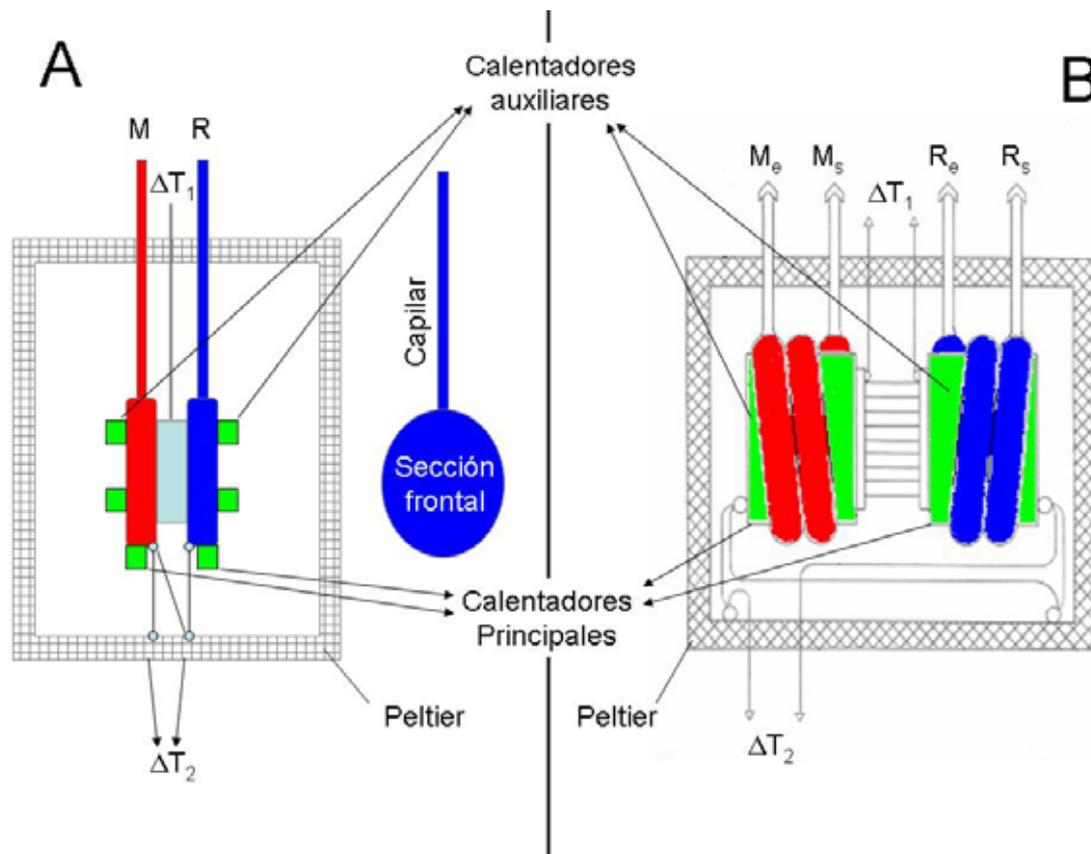
-La medida es dinámica. Se va produciendo a la vez que se realiza un incremento de la temperatura lineal con el tiempo.

- Carecen de agitación mecánica, eliminándose la contribución por efecto Joule, para lo cual se precisa que la relación superficie de calentamiento/volumen de muestra sea lo menor posible. De este modo el calentamiento es uniforme en toda la muestra.

La adiabaticidad se consigue aproximando a cero el intercambio de calor entre las células y una coraza externa cuya temperatura regula una peltier.

Rodeando las células y en contacto íntimo con éstas encontramos los calentadores principales que se encargan de suministrar la energía para el gradiente lineal de temperatura. Cada gradiente de temperatura tiene asociado un programa de pulsos de energía determinado (calibrado para disoluciones acuosas). En la coraza que rodea a las células encontramos serie de sensores de temperatura, que nos indicarán la temperatura absoluta de nuestro experimento. Entre ambas células encontramos un sensor de cristal ( $\text{Bi}_2\text{Te}_3$ ). Este componente se encarga de transformar diferencias de temperatura en señales eléctricas. Dicha señal es amplificada y transmitida a los calentadores auxiliares que igualan la temperatura de ambas células. Recapitulando, la diferencia de temperatura entre las dos células se transforma en una cantidad de calor, que será proporcional a nuestra señal.

El presente trabajo se ha realizado con dos microcalorímetros distintos que si bien comparten todo lo expuesto anteriormente manifiestan una serie de diferencias de importancia a nivel práctico. El primero de los microcalorímetros, un gran éxito de diseño por parte de Valerian Plotnikov (Plotnikov, Brandts et al. 1997) tiene células construidas en tántalo con forma de moneda (Figura DSC1-A). Se accede a ellas a través de un fino capilar de 1.5 mm de diámetro. Si bien el volumen efectivo es de 0.5 ml aproximadamente, a nivel práctico se precisa al menos de 1 ml de muestra para realizar un buen llenado de la célula. Este calorímetro opera entre 0 y 120 °C, con sobrepresión para evitar ebullición a velocidades de barrido que van de 0 a 90 °C/hora.



**Figura DSC1:** Para ambas figuras, se representa en color verde los calentadores principales y auxiliares. En rojo la célula de muestra y en azul la célula de referencia, entre las cuales se encuentra el sensor de cristal que suministra la diferencia de temperatura entre las células ( $\Delta T_1$ ). La señal  $\Delta T_2$  es la diferencia entre la coraza externa y las células. La temperatura absoluta la determinan sensores dispuestos en la coraza. A) Representación de las células del microcalorímetro VP-DSC de MicroCal. Las células tienen forma de moneda (ver sección frontal) y la muestra se introduce a través de un fino capilar (M para la celda de muestra y R para la celda de referencia). B) Representación de las células del microcalorímetro capilar VP-DSC de MicroCal. En este caso, dichas células tienen forma de espiral. Puede observarse que ahora cada una de las células tiene un conducto de salida ( $M_s$  y  $R_s$  para celdas de muestra y referencia respectivamente) y otro de entrada ( $M_e$  y  $R_e$  para celdas de muestra y referencia respectivamente). Además, el sistema posee una mejor distribución del calor a lo largo de la superficie de las células.

El microcalorímetro capilar está basado en el diseño anterior y permite la automatización. La diferencia inicial la encontramos en la forma de las células, tratándose en este caso de capilares en espiral de sólo 0.125 ml de volumen interno útil (Fig. DSC1-B). Esta estructura obliga a que cada una de las células tenga dos accesos, de manera que las células se pueden limpiar y cargar simplemente por desplazamiento. Además hace que la difusión de calor sea más homogénea, lo que evita la distorsión de la señal cuando las muestras agregan. Se han mejorado algunas características importantes del microcalorímetro, de modo que ahora el rango de temperaturas va de 0 a 130 °C, las velocidades de barrido de 0 a 250 °C/hora (recortándose considerablemente el tiempo experimental, lo cual posibilita realizar un gran número de barridos en poco tiempo) y permite su uso en modo manual o en modo automático. En nuestro caso, si bien se ha usado en ambos modos, se ha optado por la opción automática. Las muestras y las referencias se disponen en placas de 96 pocillos que se guardan dentro de un dispositivo termorregulado. Un brazo automático se encarga de tomar las muestras correspondientes e introducir las dentro de las células a través de un sistema de válvulas. Conectado a este sistema de válvulas se encuentra una entrada de nitrógeno que suministra la sobrepresión. Basta ahora programar los barridos a realizar y los procesos de limpieza. A partir de ahí no se precisa (normalmente) de intervención humana alguna. La capacidad máxima del aparato es de 298 ciclos de llenado.

En lo referente a la reproducibilidad de la señal hay que decir que en ambos calorímetros es excelente, dándose en ambos una desviación estándar de 1.5  $\mu$ cal/min en barridos a 90 grados/hora.

### **5.1.2- DSC de proteínas.**

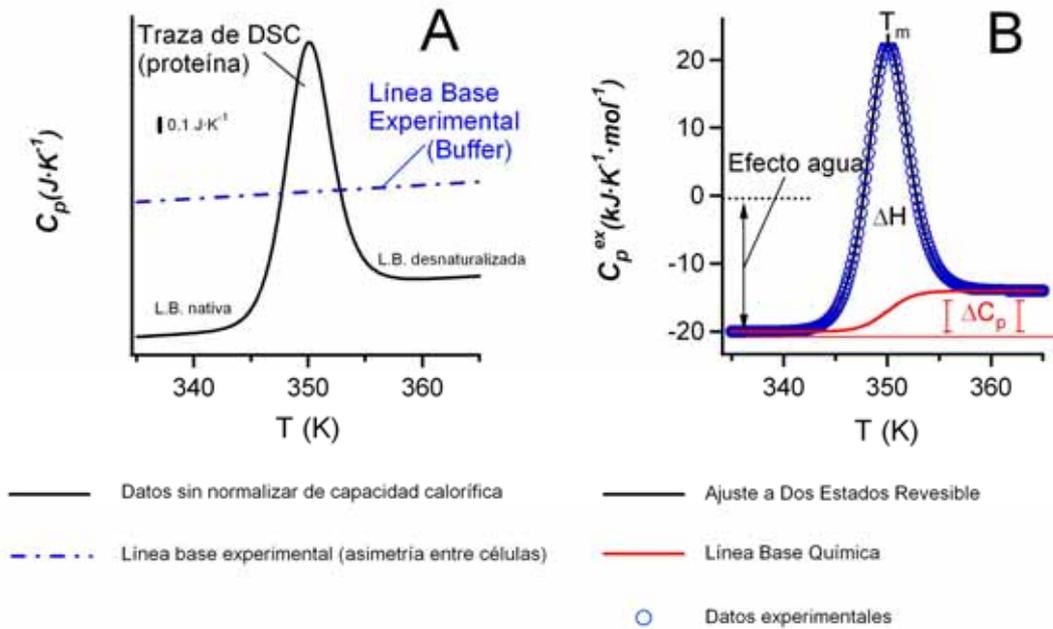
De entre las macromoléculas biológicas, las proteínas vienen siendo objeto de estudio frecuente de la microcalorimetría diferencial de barrido. Las características de nuestra muestra a la hora de realizar el experimento deben ser:

- Pureza.
  
- El solvente de referencia debe ser igual a aquél en el que se encuentra disuelta nuestra proteína. Por esto se hace necesaria la

diálisis de nuestras muestras; posteriormente usaremos en la célula de referencia el buffer del último cambio de la diálisis.

- Tanto la disolución de proteína como la de referencia deben estar filtradas y desgasificadas.

En primer lugar el aparato se debe estabilizar con el buffer en el que se encuentra la proteína hasta conseguir reproducibilidad en la señal. Este experimento es frecuentemente denominado línea base instrumental y nos permite eliminar la contribución a la señal debida a aspectos instrumentales (la asimetría entre las células de referencia y muestra). A continuación se cambia la disolución de la célula de muestra por nuestra disolución de proteína, que ha sido preparada a una concentración conocida (normalmente entre 0.1 y 2 mg/mL) y determinada con mucha precisión, dado que errores en la medida de la concentración alteran gravemente el posterior análisis de las trazas. El resultado es, normalmente, la obtención de un pico que refleja una transición entre diferentes estados inducida por el incremento de temperatura. Antes de detener nuestro experimento, es aconsejable permitir que el aparato enfríe las muestras y realice un barrido más sobre las mismas (reheating) (en la Figura DSC2-A se indican los elementos clásicos de un experimento de calorimetría). Esto suele dotarnos de una información vital a la hora del análisis e interpretación de los datos: Nos indica si el proceso de desnaturalización es reversible o irreversible en función de si la señal se recupera significativamente o no. Para analizar nuestros datos necesitaremos restar la línea base instrumental del barrido realizado sobre la proteína a fin de eliminar las asimetrías inherentes al aparato. Posteriormente se transforma la información suministrada por el aparato, potencia vs tiempo, a energía vs temperatura usando para ello la velocidad de barrido experimental. Y finalmente se normaliza por la concentración.



**Figura DSC2:** Perfiles simulados de DSC para  $\Delta H = 400 kJ/mol$ ,  $T_m = 350 K$  y  $\Delta C_p = 6 kJ \cdot K^{-1} \cdot mol^{-1}$ . Las capacidades caloríficas de los estados nativo y desnaturalizado se han tomado como independientes de la temperatura aunque lo usual es que exista cierta dependencia. A) Se muestran los elementos típicos de un experimento de DSC. En azul se muestra la Línea Base, en la práctica no se debe introducir la muestra hasta que ésta no es totalmente reproducible. La información que contiene es la asimetría entre células y ha de ser sustraída de los datos de capacidad calorífica de nuestra proteína (en negro). B) Una vez sustraída la línea base, se procede al ajuste. La línea roja representa la Línea Base Química, que da el cambio con la temperatura de la capacidad calorífica intrínseca de los estados implicados (obviamente sin considerar la contribución de la transición). El área que encierra el pico y la línea base química representa la entalpía del proceso. Por último se resalta la diferencia de los datos con respecto al cero (Efecto Agua). Esta diferencia se debe a que las moléculas de proteína desplazan un volumen de agua, fenómeno que no ocurre en la célula de referencia (el agua tiene mayor capacidad calorífica absoluta, de ahí el signo negativo). Una vez sustraído este efecto, tendríamos la capacidad calorífica absoluta de nuestra proteína.

En el mejor de los casos se obtiene un perfil semejante al mostrado en la (Figura DSC2-B). La figura simula una transición que

puede explicarse de acuerdo a uno de los modelos más comunes en desplegamiento de proteínas (como se verá más adelante, modelo de dos estados). Por ello, se pueden obtener fácilmente la diferencia de capacidad calorífica entre los estados implicados, la entalpía del proceso y de un modo sencillo la estabilidad de la proteína como función de la temperatura. Pero las proteínas pueden llegar a tener comportamientos muy divergentes como hemos visto en la INTRODUCCIÓN.

Cabe mencionar que un aparato como un microcalorímetro diferencial de barrido debe tener un mantenimiento. Sin embargo, los procesos de limpieza pueden llegar a alterar su funcionamiento. Por tanto es recomendable limpiar siempre bien las células con agua destilada (un litro de agua por célula). En caso de sospechar que persiste algún tipo de suciedad se van siguiendo protocolos de limpieza progresivamente más agresivos. En primer lugar, si el agua no limpió la impureza se prueba con un detergente comercial (Decon 70) y una incubación a 60 °C durante una hora. Si ello no funciona, se prueba con ácido fórmico al 10% e incubando a 60 °C. La última alternativa suele ser el uso de ácido nítrico al 20% e incubación a 60 °C durante una hora. En cualquier caso, siempre se debe evitar el uso de compuestos básicos, pues dañan la pared de las células del microcalorímetro.

### **5.1.3- Modelos para el análisis de datos de DSC de proteínas.**

El siguiente paso es el análisis de los datos conforme a un modelo teórico propuesto a priori. Como principio, diremos que será más adecuado aquel modelo que explica correctamente los datos imponiendo el menor número de suposiciones. En definitiva, tomaremos siempre que sea posible el modelo más simple y que explique coherentemente los datos.

Con este principio procederemos a decidir qué modelo es el más adecuado prestando atención a los siguientes elementos (Freire 1995; Sanchez-Ruiz 1995; Ibarra-Molero and Sanchez-Ruiz 2006):

- Criterio de reversibilidad: Se refiere a si la proteína puebla los distintos estados que dan lugar a la transición en equilibrio. En tal caso, normalmente, será aplicable la termodinámica de procesos reversibles.

- Análisis del efecto de concentración de proteína: Resulta indicativo de procesos de oligomerización de alguno de los estados implicados.

- Análisis del efecto de la velocidad de barrido: Indica la presencia de procesos cinéticos en nuestro experimento.

- Análisis de unión de ligandos.

Una vez dispongamos de esta información experimental, será posible elegir un modelo teórico para el análisis de dichos datos. En la figura DSC3 se representan algunos de los modelos más comúnmente utilizados. Cabe destacar que los modelos normalmente utilizados suponen estados discretos y definidos, es decir, un solo estado nativo y un solo estado desnaturalizado. Sin embargo las proteínas no pueblan un único estado nativo ni un único estado desnaturalizado, en realidad la proteína se encuentra en multitud de conformaciones parecidas ("ensembles"). No obstante, la aproximación de asumir estados discretos es válida siempre y cuando los "ensembles" tengan energías suficientemente diferentes (Kaya and Chan 2000).

Antes de entrar en el desarrollo de los modelos utilizados a lo largo del presente trabajo hay que mencionar la existencia de otro tipo de modelos (Kanehisa and Ikegami 1977; Munoz and Sanchez-Ruiz 2004). Dichos modelos son mucho más sofisticados dado que no presupone una exigencia inherente a los modelos tradicionales: la presencia de barreras energéticas robustas que separan los estados implicados en la transición. Estos modelos se han aplicado con éxito en proteínas reversibles con dos estados poblados diferenciados y proteínas "downhill" (Munoz and Sanchez-Ruiz 2004; Naganathan, Sanchez-Ruiz et al. 2005; Halskau, Perez-Jimenez et al. 2008).

MODELOS DE DESNATURALIZACIÓN  
REVERSIBLE

I



III

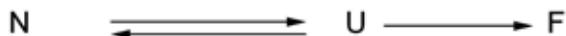


V



MODELOS DE DESNATURALIZACIÓN  
IRREVERSIBLE  
(Lumry-Eyring)

II



IV



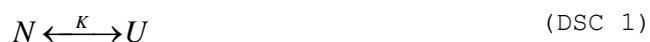
VI



**Figura DSC3:** En esta figura se representan algunos de los modelos de desnaturalización de proteínas más frecuentemente usados. Se considera N la proteína nativa, U la proteína reversiblemente desplegada y F el estado irreversiblemente desnaturalizado. I, III, V.-modelos de desnaturalización reversible. II, IV, VI.- Modelos de desnaturalización reversible con un paso de desnaturalización irreversible. Los modelos más sencillos (I y II) se expanden para asumir la presencia de fenómenos de oligomerización (III y IV) y fenómenos de unión a ligandos (L) (V y VI).

### 5.1.3.1- Modelo de dos estados reversible

El modelo de dos estados reversible (Figura DSC3-I) supone la presencia de sólo dos estados poblados significativamente, el estado nativo (N) y el estado desplegado (U). La proporción de dichos estados determina el valor de la constante de equilibrio  $K$  para una temperatura dada.



$$K = \frac{[U]}{[N]} \quad (\text{DSC 2})$$

Nuestros datos de calorimetría representan la capacidad calorífica de exceso  $C_p^{ex}$ , donde la capacidad calorífica de referencia en nuestro caso es la del estado nativo. Y como hemos visto en el apartado anterior, el área encerrada entre ésta y la línea base química suministra la entalpía del proceso  $\Delta H$ . Por tanto, el perfil queda definido por:

$$C_p^{ex} = \frac{\partial \langle \Delta H \rangle}{\partial T} \quad (\text{DSC 3})$$

$$\langle \Delta H \rangle = x_U \cdot \Delta H = \frac{K}{(1+K)} \cdot \Delta H \quad (\text{DSC 4})$$

Donde  $\langle \Delta H \rangle$  es la entalpía de exceso (la referencia es la línea base química). Haciendo uso de las ecuaciones de van't Hoff y Kirchoff para expresar las dependencias de  $K$  y  $\Delta H$  con la temperatura, respectivamente:  $\partial \ln K / \partial T = \Delta H / (R \cdot T^2)$  y  $\partial \Delta H / \partial T = \Delta C_p$ , obtenemos finalmente:

$$C_p^{ex} = \frac{(\Delta H)^2}{RT^2} \cdot \frac{K}{(1+K)^2} + \frac{K}{(1+K)} \cdot \Delta C_p \quad (\text{DSC 5})$$

El primer término de la ecuación DSC5 representa la transición y tiene forma de pico. El segundo término representa la suma de las capacidades caloríficas de los estados nativo y desplegado, suministra una curva de forma sigmoidal, la línea base química.

El valor de la constante  $K$  como función de la temperatura se deduce a partir de la diferencia de energía libre entre los estados nativo y desplegado  $\Delta G$ .

$$K = e^{(-\Delta G/R \cdot T)} \quad (\text{DSC 6})$$

Y los valores de  $\Delta G$  se deducen a partir de las ecuaciones habituales de la termodinámica. Tomando previamente una temperatura de referencia  $T_m$  a la cual  $x_U = x_N$  (comúnmente denominada temperatura de

desnaturalización), lo que implica  $\Delta G_m = 0$  y  $\Delta H_m = T \cdot \Delta S_m$  (siendo  $\Delta H_m$  la entalpía total del proceso atribuida a la  $T_m$ ). Asumiendo que el  $\Delta C_p$ , la diferencia en capacidad calorífica entre los estados nativo y desplegado, no varía significativamente,

$$\Delta G(T) = \Delta H_m \cdot \left[ 1 - \frac{T}{T_m} \right] + \Delta C_p \cdot \left[ T - T_m - T \cdot \ln \left( \frac{T}{T_m} \right) \right] \quad (\text{DSC7})$$

La ecuación DSC7 implica que el conocimiento de los valores de  $T_m$ ,  $\Delta H_m$  y  $\Delta C_p$  dan una descripción detallada de la estabilidad termodinámica del estado nativo con la temperatura. La dependencia de  $\Delta G$  con la temperatura ofrece un único máximo y dos puntos de corte a  $\Delta G = 0$  (Becktel and Schellman 1987; Schellman 1987) y es conocida como curva de estabilidad. Esto supone que la proteína desnaturaliza tanto a altas temperaturas como a bajas temperaturas, proceso denominado "desnaturalización fría" (Privalov, Griko Yu et al. 1986).

Por otro lado, resulta obvio que el hecho de que el modelo ajuste los datos no es demostración de su validez. Clásicamente se ha usado el cociente entre las entalpías calorimétricas y de van't Hoff como una medida de la validez del modelo de dos estados reversible (Freire 1995). Dado que ambas se pueden obtener de los datos experimentales "crudos" resulta un método independiente y válido con algunas salvedades (Zhou, Hall et al. 1999). La entalpía calorimétrica será igual al área encerrada entre la transición y la línea base química y es independiente del modelo usado en el análisis. La entalpía de van't Hoff rige como varía el área encerrada con la temperatura cuando el proceso es de dos estados, es decir, afecta a la forma del pico (su cálculo frecuentemente se hace a través de la anchura del pico):

$$\Delta H^{vH} = 4RT_m^2 \left[ \frac{\Delta C_{p,m}^{ex}}{\Delta H_m} \right] \quad (\text{DSC8})$$

Siendo  $\Delta C_{p,m}^{ex}$ ,  $\Delta H_m$  y  $T_m$  la capacidad calorífica de exceso en el punto medio de la transición (a la  $T_m$ ), la entalpía calorimétrica y

la antes referida  $T_m$ , respectivamente. Cuando el cociente  $r = \frac{\Delta H_m}{\Delta H^{vH}}$  es igual a 1, el proceso de desnaturalización es de dos estados reversible. Desviaciones con respecto a este valor suelen indicar la presencia de intermedios significativamente poblados, procesos de oligomerización, ...(Freire 1995).

### **5.1.3.2-Desnaturalización multiestado según el formalismo de la función de partición.**

Es interesante hacer mención en este trabajo a un modo más general de deducir modelos generales para el análisis de datos de DSC basado en el formalismo de la función de partición (Freire and Biltonen 1978). Este desarrollo permite introducir fácilmente en el modelo estados intermedios, procesos de unión a ligando y procesos de oligomerización. Toma al estado nativo como referencia y obtiene el peso estadístico de cada uno de los estados por su cociente con el nativo:

$$I_0 \leftrightarrow I_1 \leftrightarrow I_2 \leftrightarrow \dots \leftrightarrow I_{n-1} \leftrightarrow I_n \quad (\text{DSC9})$$

$$\omega_i = \frac{[I_i]}{[I_0]} = K_i = e^{(-\Delta_i^N G/RT)} \quad (\text{DSC10})$$

Ahora definimos la función de partición del sistema como la sumatoria de los pesos estadísticos,

$$Z = \sum_{i=0}^n \omega_i = \sum_{i=0}^n e^{(-\Delta_i^N G/RT)} = 1 + \sum_{i=1}^n e^{(-\Delta_i^N G/RT)} \quad (\text{DSC12})$$

Y la fracción de cada uno de los estados por el cociente entre el peso estadístico de ese estado y la función de partición,

$$x_i = \omega_i / Z = (1/Z) \cdot e^{(-\Delta_i^N G/RT)} \quad (\text{DSC13})$$

La entalpía de exceso, promedio de las contribuciones de todos los estados presentes, viene dada por:

$$\langle \Delta H \rangle = \sum_{i=0}^n x_i \cdot \Delta_N^i H = (1/Z) \cdot \sum_{i=0}^n \Delta_N^i H \cdot e^{(-\Delta_N^i G/RT)} \quad (\text{DSC14})$$

Y la capacidad calorífica de exceso:

$$C_p^{ex} = \partial \langle \Delta H \rangle / \partial T = (\langle \Delta H^2 \rangle - \langle \Delta H \rangle^2) / (RT^2) + \langle \Delta C_p \rangle \quad (\text{DSC15})$$

Donde  $\langle \Delta H^2 \rangle = \sum_{i=0}^n x_i \cdot (\Delta_N^i H)^2$ . El primer término de la ecuación DSC15 representa el desplazamiento del equilibrio de desnaturalización inducido por la temperatura y el segundo la capacidad calorífica media de la proteína  $\left( \langle \Delta C_p \rangle = \sum_{i=0}^n \Delta_N^i C_p \cdot x_i \right)$  a cada temperatura y corresponde a la línea base química.

La función de partición como función de la temperatura se obtiene por doble integración de la ecuación DSC14.

$$\ln[Z(T)] = \int_{T_0}^T (1/R \cdot T^2) \langle \Delta H \rangle \cdot dT = \int_{T_0}^T (1/R \cdot T^2) \cdot \left( \int_{T_0}^T C_p^{ex} \cdot dT \right) \cdot dT \quad (\text{DSC16})$$

Donde  $T_0$  es una temperatura de referencia a la cual sólo se encuentra poblado el estado nativo. Este tipo de análisis se basa en una propiedad única de la calorimetría diferencial de barrido, véase que en teoría es capaz de suministrar directamente la función de partición del sistema. Además, este análisis es fácilmente generalizable para incluir otros fenómenos como unión a ligandos,...(Thorolfsson, Ibarra-Molero et al. 2002).

### **5.1.3.3-Modelo de dos estados irreversible.**

El modelo de dos estados irreversible (un caso extremo de Fig. DSC3-II), considera la existencia de un proceso cinético de primer orden que opera sobre el estado desnaturalizado y cuya constante

cinética es mucho mayor que la constante cinética de replegamiento  $k \gg k_2$  (Sanchez-Ruiz 1992). Finalmente, asume que el proceso irreversible no conlleva un cambio de entalpía significativo en comparación al proceso de desplegamiento.



Donde  $k_1$  y  $k_2$  son las constantes cinéticas para los procesos de desplegamiento y replegamiento, respectivamente. Dado que como mencionamos antes  $k \gg k_2$ , deducimos dos puntos importantes: El estado desplegado U no se encuentra significativamente poblado y no se establecerá equilibrio químico entre N y U. Por lo que el proceso se puede simplificar a:



La variación de la concentración de proteína nativa, como decíamos, sigue una cinética de primer orden:

$$\frac{d[N]}{dt} = -k \cdot [N] \Rightarrow \frac{dx_N}{dt} = -k \cdot x_N \quad (\text{DSC19})$$

Que puede ser expresada también como variación de la concentración con la temperatura dividiendo en ambos términos por la velocidad de barrido,  $v$  ( $v = \frac{dT}{dt}$ ). Posteriormente, obtenemos la variación en la fracción de proteína nativa con la temperatura dividiendo ambos términos por la concentración total de proteína, con lo que finalmente obtenemos:

$$\frac{dx_N}{dT} = -\frac{1}{v} \cdot k \cdot x_N \quad (\text{DSC20})$$

Reordenando términos, llegamos a la siguiente integral:

$$\int \frac{1}{x_N} \cdot dx_N = \int -\frac{1}{v} \cdot k \cdot dT \quad (\text{DSC21})$$

La constante cinética es dependiente de la temperatura y dicha dependencia obedece a la ecuación de Arrhenius:

$$k = A \cdot e^{(-E_a/RT)} \quad (\text{DSC22})$$

Donde  $E_a$  es la energía de activación del proceso. Con ella y resolviendo la ecuación DSC21, deberíamos obtener la fracción de estado nativo como función de la temperatura para una velocidad de barrido dada. Sin embargo, se puede hacer una aproximación diferente que evita introducir la velocidad de barrido. Para ello definimos  $T_m$  como la temperatura a la cual se obtiene el máximo de  $C_p^{ex}$ . Esta elección no es trivial, puesto que dicha temperatura depende de la velocidad de barrido y presenta la peculiaridad de que  $dC_p^{ex}/dT = 0$ , lo cual implica que la derivada segunda de la concentración con la temperatura se haga también cero. Para realizar dicha aproximación, definimos la constante cinética como función de la concentración de estado nativo a partir de la ecuación DSC20:

$$k = -v/[N] \cdot \frac{d[N]}{dT} = -v \cdot \frac{d \ln[N]}{dT} \quad (\text{DSC23})$$

Igualando la ecuación DSC23 con la ecuación DSC22, derivando y simplificando:

$$\frac{d^2 \ln[N]}{dT^2} = \frac{d \ln[N]}{dT} \cdot \frac{E_a}{RT^2} \quad (\text{DSC24})$$

Podemos obtener también la derivada de  $\frac{d^2 \ln[N]}{dT^2}$  a través de la ecuación DSC23 directamente:

$$\frac{d^2 \ln[N]}{dT^2} = \frac{1}{[N]} \cdot \frac{d^2[N]}{dT^2} - \frac{1}{[N]^2} \left( \frac{d[N]}{dT} \right)^2 \quad (\text{DSC25})$$

Cuando  $T = T_m$ , se anula el primer término de la ecuación DSC25. Igualando las ecuaciones DSC24 y DSC25 y considerando siempre que ahora estamos a  $T = T_m$ , obtenemos:

$$\frac{E_a}{RT_m^2} \cdot \frac{1}{[N]} \left( \frac{d[N]}{dT} \right)_{T_m} = - \frac{1}{[N]^2} \cdot \left( \frac{d[N]}{dT} \right)_{T_m}^2 \quad (\text{DSC26})$$

Simplificando y haciendo el cambio  $\left( \frac{d \ln[N]}{dT} \right)_{T_m} = - \frac{A}{v} \cdot e^{(-E_a/RT_m)}$ ,

podemos obtener fácilmente:

$$\frac{A}{v} \cdot e^{(-E_a/RT_m)} = \frac{E_a}{RT_m^2} \quad (\text{DSC27})$$

Para solucionar la ecuación DSC21 y obtener la fracción de estado nativo como función de la temperatura, necesitamos evaluar la integral de Arrhenius. Esto se puede hacer bien numéricamente o bien realizando una expansión de Taylor. Dicha expansión se puede truncar en el segundo término dado que los subsiguientes términos apenas modifican los resultados (Sanchez-Ruiz, Lopez-Lacomba et al. 1988):

$$1/T = 1/T_m - \Delta T / T_m^2 \quad (\text{DSC28})$$

Con lo cual podemos dejar la ecuación DSC21 como:

$$\ln x_N = \int \frac{A}{v} \cdot e^{(E_a/RT_m)} \cdot e^{(-E_a \Delta T / RT_m^2)} \cdot dT \quad (\text{DSC29})$$

Sustituyendo con DSC27, nos queda la siguiente integral:

$$\ln x_N = \int \frac{E_a}{RT_m^2} \cdot e^{(-E_a \Delta T / RT_m^2)} \cdot dT \quad (\text{DSC30})$$

Y esta se resuelve para dar la fracción de estado nativo como función de la temperatura.

$$x_N = \exp \left[ -\exp \left( \frac{E_a \cdot \Delta T}{R \cdot T_m^2} \right) \right] \quad (\text{DSC31})$$

A partir de aquí resulta sencillo obtener la ecuación que describe nuestra traza calorimétrica. Los elementos que la componen serán:

$$C_p^{app} = C_p^{pre} + (C_p^{post} - C_p^{pre}) \cdot (1 - x_N) - \Delta H \cdot \left( \frac{dx_N}{dT} \right) \quad (\text{DSC32})$$

Siendo  $C_p^{pre}$  y  $C_p^{post}$  polinomios, normalmente de primer grado, que describen fenomenológicamente las dependencias de las capacidades caloríficas de los estados nativos y desplegado con la temperatura. Los dos primeros términos del segundo miembro de la ecuación DSC32 representan la línea base química y el último término representa la transición calorimétrica y queda definido:

$$C_p^{ex} = \frac{-\Delta H \cdot E_a}{R \cdot T_m^2} \cdot \exp \left( \frac{E_a \cdot \Delta T}{R \cdot T_m^2} \right) \cdot \exp \left[ -\exp \left( \frac{E_a \cdot \Delta T}{R \cdot T_m^2} \right) \right] \quad (\text{DSC33})$$

Cabe destacar que al igual que el modelo de dos estados reversible tenía un método independiente de mostrar su validez (cociente entre las entalpías calorimétricas y de van't Hoff), el modelo de dos estados irreversible también presenta la posibilidad de dicha comprobación (Conejero-Lara, Mateo et al. 1991). La energía de activación se puede obtener de diversas formas y todas ellas han de suministrar el mismo valor:

1.-A partir de experimentos realizados a distintas velocidades de barrido, podemos construir una gráfica de Arrhenius cuya pendiente nos indicará la energía de activación. Los valores de la constante para cada temperatura los obtendremos según:

$$k_{app} = \frac{v \cdot C_p^{ex}}{\Delta H - \langle \Delta H \rangle} \quad (\text{DSC34})$$

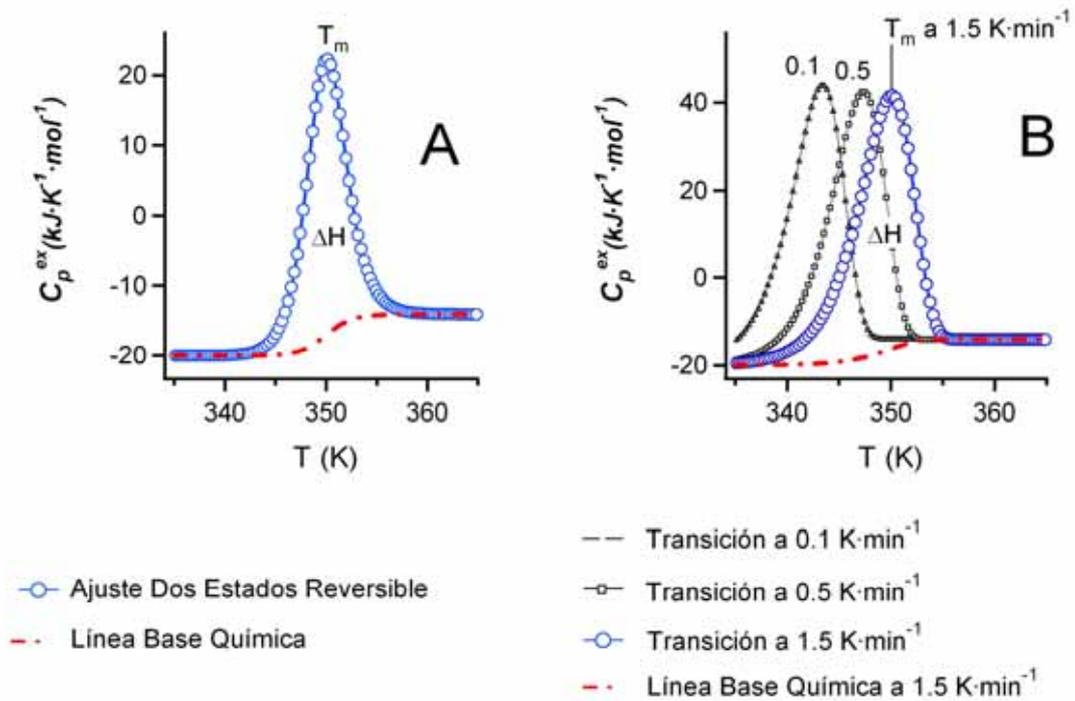
2.- Otro método es obtener la energía de activación a partir de varios experimentos a distinta velocidad de barrido usando la ecuación:

$$\ln\left(\frac{v}{T_m^2}\right) = \text{constante} - \frac{E_a}{RT_m} \quad (\text{DSC35})$$

Ecuación que se obtiene fácilmente a partir de la ecuación DSC27.

Aunque existen aún más métodos (Sanchez-Ruiz, Lopez-Lacomba et al. 1988; Conejero-Lara, Mateo et al. 1991; Sanchez-Ruiz 1992), el mensaje que ha de quedar claro es que no es suficiente con que el modelo ajuste a un solo juego de datos. Su validez se demuestra rigurosamente por coherencia entre los experimentos realizados a distintas velocidades de barrido, siendo aconsejable realizar un análisis global de los datos a distintas velocidades de barrido.

La figura DSC4 ilustra a continuación las diferencias observables entre un proceso de dos estados reversible y un proceso de dos estados irreversible.



**Figura DSC4:** Simulación de experimentos de DSC . A) transición reversible de dos estados (ecuación DSC5) para  $T_m = 350\text{ K}$ ,  $\Delta H = 400\text{ kJ/mol}$  y  $\Delta C_p = 6\text{ kJ}\cdot\text{K}^{-1}\cdot\text{mol}^{-1}$ . Las capacidades caloríficas de los estados nativo y desplegado se tomaron como independientes de la temperatura. La transición es simétrica con respecto a la  $T_m$ , punto donde las fracciones de proteína nativa y desplegada son iguales. Coincide, prácticamente, con el máximo de  $C_p^{ex}$ . B) Transiciones irreversibles de dos estados (ecuación DSC33) para  $E_a = 400\text{ kJ/mol}$  y  $\Delta H = 400\text{ kJ/mol}$  a distintas velocidades de barrido (0.1, 0.5 y 1.5  $\text{K}\cdot\text{min}^{-1}$ ). Las capacidades caloríficas de los estados nativo y desplegado se tomaron como independientes de la temperatura. La transiciones son asimétricas. El máximo de  $C_p^{ex}$  determina la  $T_m$ , que en este caso no supone que las fracciones nativa y desplegada sean iguales (de hecho a la  $T_m$ ,  $x_N = 1/e$ ). Los rasgos más distintivos de este tipo de procesos son la irreversibilidad, la asimetría y el desplazamiento de la transición con la velocidad de barrido.

## **5.2. Estudio del desplegamiento/ replegamiento de proteínas por emisión de Fluorescencia**

### **5.2.1.- Generalidades de la técnica y su uso en proteínas.**

Cuando irradiamos una molécula con una determinada longitud de onda algunos electrones de la misma absorben la energía de los fotones. Como consecuencia pasan a estados de mayor energía. Este electrón tiende a volver a su estado fundamental (de menor energía), para lo cual requiere "desprenderse" de la energía absorbida. Esta desexcitación puede ocurrir de distintos modos: de manera no radiativa, bien a través de interacciones con otras moléculas del entorno bien por solapamiento del estado excitado con modos vibracionales de alta energía del estado fundamental, o de modo radiativo. Cuando la relajación se produce por la emisión de un fotón, hablaremos de fluorescencia si la transición no implica cambio en la multiplicidad del electrón (cambio de spin) y de fosforescencia si esta se produce (lo cual resulta en un importante desfase temporal entre la excitación y la relajación).

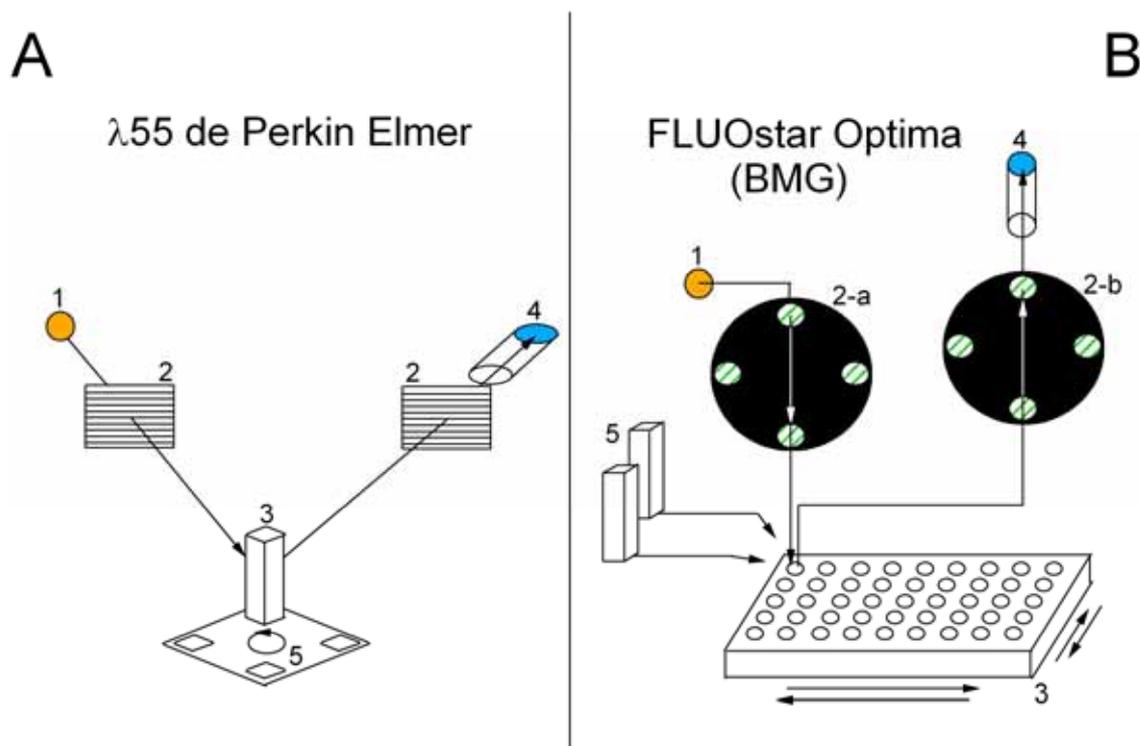
La emisión de fluorescencia presenta una serie de características que es importante resaltar:

- Hay un desplazamiento hacia el rojo (longitudes de onda mayores, es decir, menos energéticas) del espectro de emisión en relación al de absorción.
- Dada la competencia entre las distintas formas de desexcitación, fundamentalmente no radiativas, el número de excitaciones no se traduce en el mismo número de emisiones. La relación entre el número de fotones emitidos por fluorescencia y el número total de fotones absorbidos se conoce como rendimiento cuántico (cuyo valor está comprendido, obviamente, entre cero y uno).
- La intensidad de fluorescencia que emite una muestra será proporcional a la concentración de fluoróforo, al coeficiente de extinción molar, a la intensidad de luz incidente y al rendimiento cuántico.

- Existe el fenómeno de la desactivación (quenching) de fluorescencia. Se debe normalmente a la interacción de alguna de las moléculas presentes en la disolución (como oxígeno molecular) con los fluoróforos. Produce una relajación del fluoróforo no radiativa y por lo tanto causa una disminución en la intensidad de fluorescencia. La temperatura es un factor que conviene controlar especialmente, ya que afecta al rendimiento cuántico y a los fenómenos de quenching.

Con respecto a la proporcionalidad con la concentración de fluoróforo, hay que resaltar que ésta sólo se produce con muestras bastantes diluidas debido a efectos de filtro interno, filtro externo, perdiéndose la linealidad a medida que la concentración va en aumento. Por ello, es siempre necesario hacer una recta patrón de la proteína con la que vamos a trabajar, a fin de valorar el rango de concentraciones en el que podemos realizar los experimentos (García-Segura J.M. 2002).

En el presente trabajo se han utilizado dos espectrofluorímetros. El primero de ellos es el  $\lambda 55$  de Perkin Elmer, (ver Figura FLU1-A). Su diseño se basa en la medida de muestras dispuestas en cubetas de cuarzo. Se ha acoplado un dispositivo multicubeta que permite la medida de un máximo de cuatro muestras. La razón de esta limitación es que el ángulo de medida de la emisión es de  $90^\circ$  con respecto a la excitación. Este instrumento posee dos monocromadores (uno para la excitación y otro para la emisión) y gracias a unos "slits" tiene capacidad para regular el ancho de banda de las longitudes de onda seleccionadas.



**Figura FLU1:** Representación esquemática de los fluorímetros utilizados. A) λ55 de Perkin Elmer. Apréciase el ángulo de 90° entre los haces de excitación y emisión. 1, fuente de luz (arco de Xenon). 2, monocromadores de rejilla. 3, cubeta de muestra. 4, tubo fotomultiplicador. 5, dispositivo multicubeta. B) FLUOstar OPTIMA de BMG. En este aparato los haces de excitación son dirigidos por guías ópticas. 1, Fuente de luz (arco de Xenon). 2, Dispositivos con filtros ópticos para la excitación (2-a) y emisión (2-b). 3, placa de muestras. Las flechas de los laterales indican la dirección de desplazamiento para medir distintos pocillos. 4, tubo fotomultiplicador. 5, dispositivos de inyección.

El otro espectrofluorímetro utilizado es el FLUOstar OPTIMA de BMG, Figura FLU1-B. Presenta la característica de realizar las medidas en un ángulo de 180° con respecto a la excitación. Esto es posible gracias a que unas guías ópticas sitúan los haces de luz sobre la muestra, que puede por ello disponerse en placas de distinto formato (usualmente de 96 pocillos y construidas en polipropileno). Adicionalmente, está dotado de 2 sistemas de inyección de muestra y un sistema de termorregulación, todo lo cual permite realizar una gran cantidad de experimentos con un mínimo de participación humana. Sin embargo en detrimento del aparato podemos mencionar su versatilidad y calidad óptica. No es posible, por ejemplo, seleccionar longitudes de

onda más que a través de filtros ópticos dispuestos estáticamente (con una anchura de banda de selección inamovible).

Las características de la técnica y del fenómeno de fluorescencia hacen que esta metodología presente dos características propicias para el estudio de proteínas: alta sensibilidad (se requiere poca muestra) y alta especificidad. Esta última característica hace que sea posible no sólo estudiar residuos específicos de la proteína, sino además hacerlo como función del ambiente que los rodea.

Los fluoróforos naturales en proteínas son: Trp, Tyr y Phe (si bien esta última presenta una sensibilidad demasiado baja, unas mil veces menor a la del Trp), que tienen los máximos de excitación a 280, 274 y 257 nm respectivamente.

Las medidas de fluorescencia realizadas en el presente trabajo se dirigieron a estudiar el desplegamiento y replegamiento de proteínas. Dado que en los trabajos presentados nuestra proteína de trabajo era tiorredoxina, en primer lugar se realizaron medidas del espectro de emisión del Trp en muestras de proteína nativa y proteína desplegada. Finalmente se eligió como longitud de onda de emisión aquella que presentaba una mayor diferencia entre los espectros nativo y desplegado.

Como se ha mencionado anteriormente, el efecto de la temperatura sobre la muestra posee contribuciones adicionales a las diferencias entre los estados nativo y desplegado. Por tanto a la hora de desnaturalizar la proteína se optó por el uso de agentes químicos tales como la urea o el cloruro de guanidinio. Su relación con la energética de las proteínas se discutirá en el siguiente apartado.

### **5.2.2.- Energética del equilibrio en plegamiento/desplegamiento de proteínas seguido por fluorescencia.**

Los osmolitos son pequeñas moléculas orgánicas involucradas en la regulación de la presión osmótica y que afectan a la estabilidad de las proteínas. Estos pueden bien estabilizar el estado nativo o bien el estado desplegado según interaccionen preferentemente con uno u otro (Street, Bolen et al. 2006). Como resultado de dicha interacción desplazan el equilibrio hacia la forma con la que interactúan

preferentemente. El uso de estos agentes está apoyado en dos suposiciones importantes:

- La interacción con la proteína es inespecífica (no hay sitios de unión). La idea generalmente aceptada es que interaccionan con los enlaces peptídicos expuestos al solvente.
- El efecto en la energética es directamente proporcional a la concentración del osmolito, con lo que podremos definir una constante  $m_{1/2}$  tal que:

$$m_{1/2} = -\frac{\partial \Delta G}{\partial [\text{osmolito}]} \quad (\text{Ec. FLU1})$$

Esto supone que el efecto del osmolito (desnaturalizante a partir de ahora) sobre la diferencia de energía libre entre los estados implicados es lineal con la concentración (lo que ha dado lugar a los comúnmente conocidos como métodos de extrapolación lineal, LEM). Dicha suposición presenta en ocasiones desviaciones a concentraciones diluidas de desnaturalizante. Especialmente cuando el agente utilizado es cloruro de guanidinio, que por ser una sal establece interacciones electrostáticas con cargas presentes en la proteína, dando lugar a un apantallamiento de interacciones electrostáticas entre cargas de la proteína que contribuirían a la estabilidad (Ibarra-Molero and Sanchez-Ruiz 1996; Pey, Rodriguez-Larrea et al. 2008). Finalmente, y hecha esta salvedad, cuando trabajamos en proteínas que manifiestan dos estados significativamente poblados reversiblemente, podemos definir la constante de equilibrio como:

$$K = e^{\left(\frac{(C-C_{1/2})m_{1/2}}{R \cdot T}\right)} \quad (\text{Ec. FLU2})$$

Siendo  $C$  la concentración de desnaturalizante (nuestra variable independiente) y  $C_{1/2}$  la concentración de desnaturalizante para la cual  $x_N = x_D (\Delta G = 0)$ .

Para obtener la ecuación que nos permita analizar los datos de un experimento de equilibrio, simplemente tendremos que igualar la

magnitud de la propiedad medida a la suma de las contribuciones de cada uno de los estados implicados.

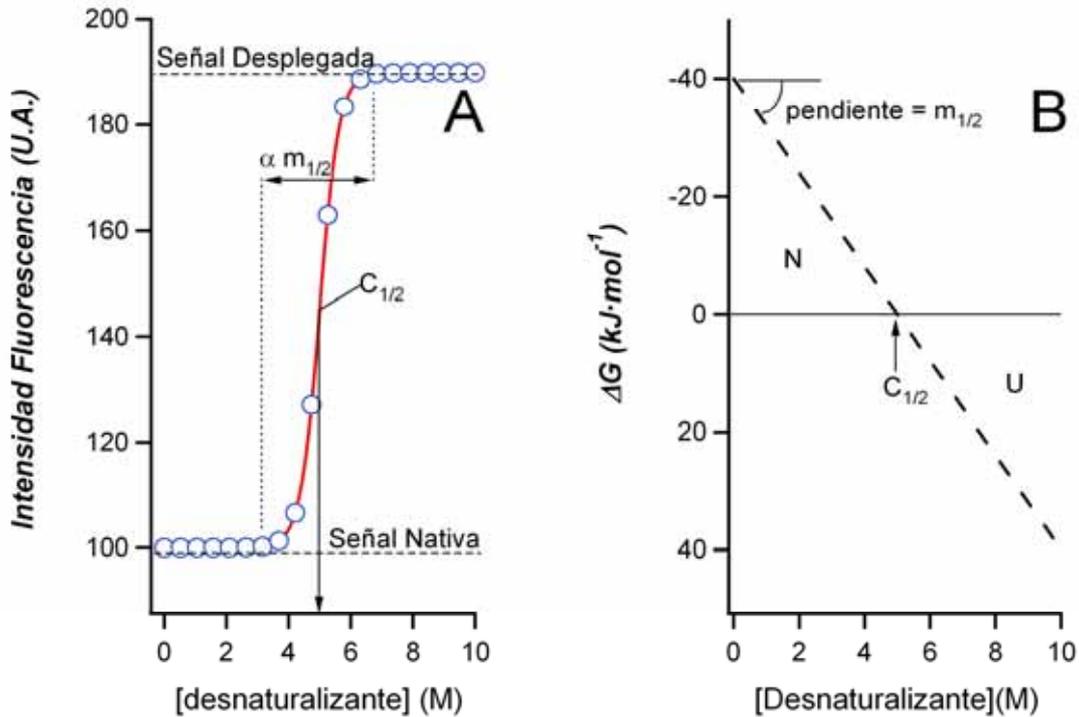
$$I_T = I_N \cdot x_N + I_D \cdot x_D \quad (\text{Ec. FLU3})$$

$$I_T = I_N + \Delta_N^D I \cdot x_D \quad (\text{Ec. FLU4})$$

$$I_T = I_N + \Delta_N^D I \cdot \left( \frac{\exp\left(\frac{(C - C_{1/2}) \cdot m_{1/2}}{R \cdot T}\right)}{1 + \exp\left(\frac{(C - C_{1/2}) \cdot m_{1/2}}{R \cdot T}\right)} \right) \quad (\text{Ec. FLU5})$$

Donde  $I_T$  es la intensidad de fluorescencia total (la que medimos),  $I_N$  es la intensidad de fluorescencia de nuestra muestra nativa y  $\Delta_N^D I$  es la diferencia entre las intensidades de la muestra totalmente desnaturalizada y la muestra totalmente nativa (es decir, la amplitud del cambio). No se expresa la dependencia de estas intensidades con la concentración de desnaturalizante dado que normalmente no existe.

A nivel práctico, para la realización del experimento tendremos que preparar muestras de proteína frente a distintas concentraciones del desnaturalizante. Prestando especial cuidado en que todas las muestras contengan exactamente la misma concentración de proteína y en que hemos dejado incubar la proteína el tiempo suficiente para que se llegue al equilibrio entre los estados implicados. Si cumplimos con ello y nuestra proteína presenta una desnaturalización reversible de dos estados obtendremos normalmente el perfil mostrado en la figura FLU2-A. Tras el ajuste, podremos extrapolar para obtener la estabilidad de la proteína en ausencia desnaturalizante (Figura FLU2-B) con el uso de los parámetros ajustables del modelo,  $C_{1/2}$  y  $m_{1/2}$



**Figura FLU2:** Caracterización del equilibrio de desnaturalización de una proteína hipotética por fluorescencia a 25 °C. A) Simulación de experimento y análisis de resultados para  $m_{1/2} = 8 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{M}^{-1}$ ,  $C_{1/2} = 5 \text{ M}$ ,  $I_N = 100 \text{ u.a.}$  e  $I_U = 190 \text{ u.a.}$  B) Dados los parámetros anteriores y conociendo la dependencia de la energía libre de desnaturalización con la concentración de desnaturalización, esto es los métodos de extrapolación lineal (ecuación FLU1), podemos extrapolar de  $C_{1/2} = 5 \text{ M}$  a  $C = 0 \text{ M}$  y obtener un  $\Delta G = 40 \text{ kJ}\cdot\text{mol}^{-1}$  en ausencia de desnaturalizante a 25 °C.

### 5.2.3.- Cinéticas de desplegamiento/ replegamiento seguidas por fluorescencia.

Como hemos mencionado en el apartado anterior, cuando una proteína es incubada con un agente desnaturalizante cambiamos la diferencia de energía entre los estados implicados (para un dos estados reversible, los estados nativo y desplegado). Como respuesta a este cambio, la proteína evoluciona y comienza a poblar nuevos estados distintos del nativo. Los experimentos cinéticos se encargan de seguir este cambio como función del tiempo.

El modelo cinético más sencillo que podemos suponer es un dos estados reversible y esto implica que la proteína sólo puebla los estados nativo y desnaturalizado. Pero, a diferencia de los experimentos de equilibrio, aquí no sólo importan los estados poblados. Son de vital importancia posibles intermedios cinéticos. Nuestro modelo, además de implicar sólo a dos estados poblados, supondría que el paso de uno a otro se realiza a través de una sola fase cinética. Finalmente, y para conseguir el modelo más sencillo, trataremos la desnaturalización como si fuera una reacción química elemental de orden uno.

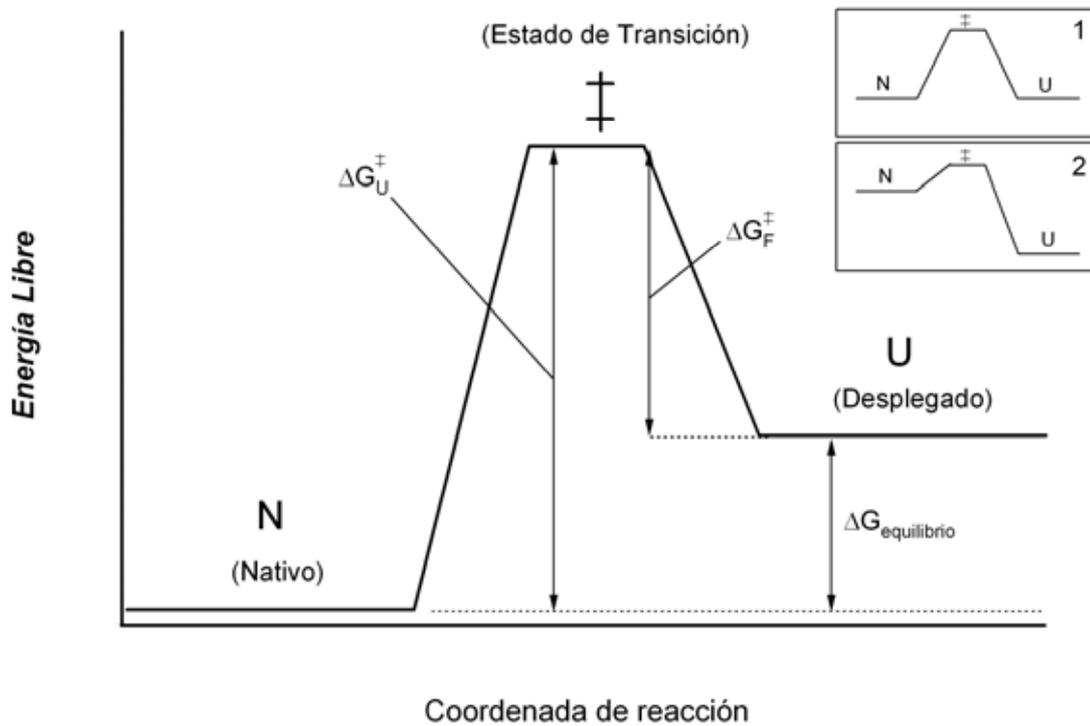


$$K = \frac{[U]}{[N]} = \frac{k_U}{k_F} \quad (\text{Ec. FLU7})$$

Donde  $k_U$  y  $k_F$  son las constantes de velocidad de los procesos de desplegamiento y replegamiento respectivamente. Pero cuando existe reversibilidad nunca se observan procesos cinéticos "puros", sino la suma de ambos. Por ello, definimos la constante de velocidad aparente  $k_{app}$ :

$$\ln k_{app} = \ln(k_U + k_F) \quad (\text{Ec. FLU8})$$

La descripción energética de estos procesos resulta hoy en día un tema controvertido. Una visión generalmente aceptada es que cada constante cinética se corresponde con un proceso limitado por la presencia de una barrera de energía libre. Por ello, se supone que la molécula pasa por un estado de máxima energía (estado de transición) a medida que progresa en la coordenada de reacción (Figura FLU3).



**FIGURA FLU3:** Coordenada de reacción en un modelo de desplegamiento de dos estados para condiciones donde se favorece el estado nativo. Se representa la energía de cada uno de los estados implicados frente a un hipotético parámetro de orden que los caracteriza. Separando los estados nativo y desplegado hay una barrera de energía libre que es el paso limitante para la reacción de desplegamiento/replegamiento. Según las condiciones experimentales la diferencia de energía libre entre los estados nativos y desplegados varía. La altura de la barrera que experimentan varía igualmente con las condiciones experimentales. Panel Incrustado 1: Coordenada de reacción en la zona de la transición. Panel Incrustado 2: Coordenada de reacción en condiciones desnaturalizantes.

La teoría del estado de transición, elaborada entre otros por Henry Eyring, asume que una molécula ocupa los estados fundamentales y el estado de máxima energía con una probabilidad que es función de la energía de cada uno de estos estados. En otras palabras, se establece el equilibrio entre los estados implicados y el estado de transición (ello no quiere decir que este último se pueble). Asumiendo la veracidad de este postulado, resulta posible describir la energética de un proceso cinético "puro" (que sólo implique un estado nativo y un estado de transición) a través de las leyes de la termodinámica.

$$k = A \cdot e^{\left(\frac{-\Delta G^\ddagger}{RT}\right)} \quad (\text{Ec. FLU9})$$

Donde A (el factor preexponencial) es una constante que para la rotura de un enlace covalente débil de una molécula simple en estado gaseoso es igual a  $\frac{k_b \cdot T}{h}$ , siendo  $k_b$  la constante de Boltzmann,  $T$  la temperatura,  $h$  la constante de Planck y  $\Delta G^\ddagger$  es la diferencia en energía libre de Gibas entre el estado nativo/desplegado y el estado de transición. El valor de esta constante es una medida de la velocidad de difusión de la molécula en el estado de transición. Obviamente su magnitud en el caso de proteínas en disolución ni tiene el mismo valor, ni es constante, ni es conocido (Knott and Chan 2006). Dado que los procesos cinéticos más rápidos observados en proteínas rondan el nanosegundo (Eaton, Munoz et al. 1997; Munoz, Thompson et al. 1997) (Yang and Gruebele 2003), si tomáramos un valor para el coeficiente preexponencial en la ecuación de Eyring nos deberíamos aproximar a esta escala temporal.

Hecha esta salvedad, queda claro que a causa del desconocimiento del coeficiente preexponencial resulta en todo imposible determinar el valor absoluto, en términos de energía libre, de la barrera al plegamiento/depiegamiento. Pero, y muy importante, sí que es factible hablar de diferencias de energía libre entre dos procesos cinéticos del mismo sistema (suponiendo simplemente que el valor del coeficiente preexponencial no varía).

Asumiendo los principios de la teoría del estado de transición y recordando que la dependencia de la energía libre de activación con la concentración de desnaturizante es lineal (Ec. FLU1), podremos definir dos nuevas constantes para la dependencia de las energías de activación de plegamiento/depiegamiento con la concentración de desnaturizante:

$$m_U^\ddagger = -\partial \Delta G_U^\ddagger / \partial C \quad (\text{Ec. FLU10})$$

$$m_F^\ddagger = -\partial \Delta G_F^\ddagger / \partial C \quad (\text{Ec. FLU11})$$

Por tanto, con el uso de estas ecuaciones y de las ecuaciones Ec. FLU8 y Ec. FLU9, podemos definir la constante de velocidad aparente en función de la concentración de desnaturalizante como (Figura FLU4-B):

$$\ln k_{app} = \ln \left( k_{1/2} \cdot \left( \exp \left( \frac{m_U^*}{R \cdot T} (C - C_{1/2}) \right) + \exp \left( \frac{m_F^*}{R \cdot T} (C - C_{1/2}) \right) \right) \right) \quad (\text{Ec. FLU12})$$

Teniendo en cuenta que  $k_{1/2}$  es el valor de  $k_F$  y  $k_U$  cuando  $C = C_{1/2}$  (por lo que  $k_F = k_U$ ).

Finalmente, sólo nos queda establecer cómo obtener la  $k_{app}$  a partir de nuestros experimentos de desplegamiento/replegamiento. Dado que el proceso sería de primer orden y con una sola fase cinética, la variación temporal en la señal registrada es proporcional a las poblaciones y a la señal emitida por cada una de las especies (Figura FLU4-A).

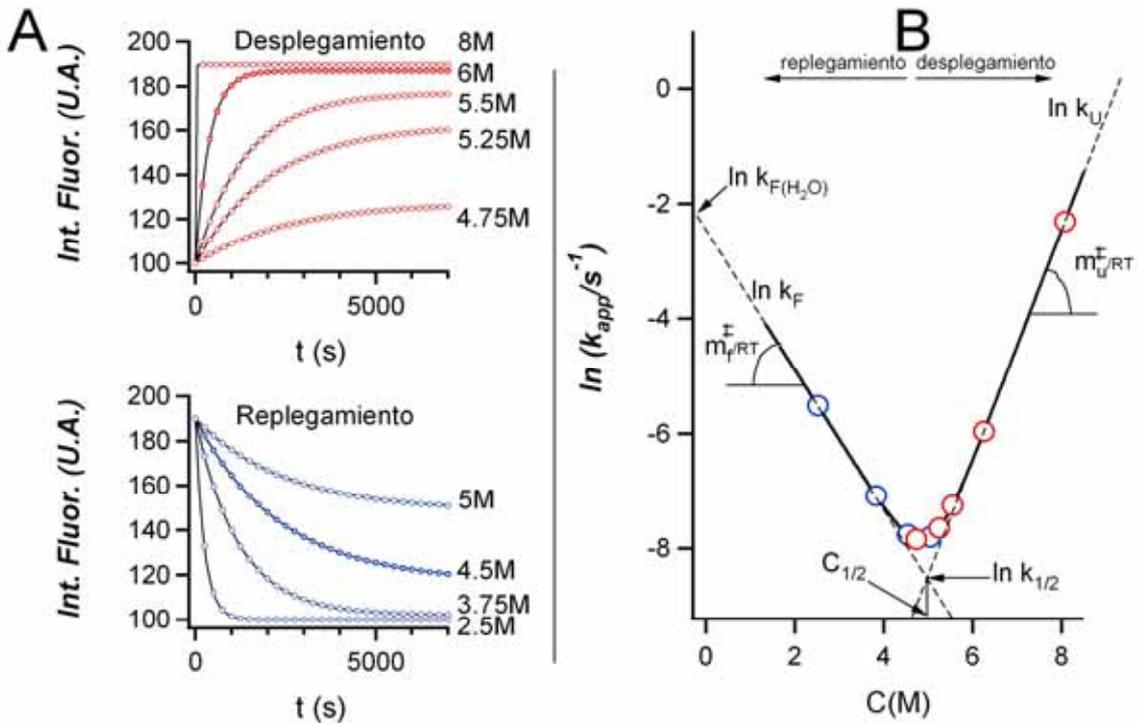
$$I = I_{eq} + \Delta I_0^{eq} \cdot e^{\left(-t/\tau\right)} \quad (\text{Ec. FLU13})$$

Siendo  $I_{eq}$  la intensidad de fluorescencia cuando se ha llegado al equilibrio,  $I_0^{eq}$  la diferencia entre la intensidad a tiempo cero y en el equilibrio y  $\tau$  el tiempo de relajación (valor de tiempo al cual la concentración de reactivo a disminuido  $1/e$  del valor inicial). Finalmente, la constante  $k_{app}$  se define como:

$$k_{app} = \frac{1}{\tau} \quad (\text{Ec. FLU14})$$

A nivel experimental hay que señalar que los factores a controlar experimentalmente son temperatura y concentración de

desnaturalizante, mientras, la concentración de proteína no afecta a los valores de la constante de velocidad.



**Figura FLU4:** Simulación de cinéticas para una proteína 2 estados reversible sin intermediarios cinéticos. Nótese que bajo estas condiciones es posible extraer parámetros de equilibrio.

( $C_{1/2} = 5M$ ,  $m_U^\ddagger = 5kJ \cdot mol^{-1} \cdot M^{-1}$ ,  $m_F^\ddagger = -3kJ \cdot mol^{-1} \cdot M^{-1}$ ,  $m_{1/2} = 8kJ \cdot mol^{-1} \cdot M^{-1}$  y

$\ln k_{1/2} = -8.52$ ). A) Parte superior, experimentos de desplegamiento

inducidos por desnaturalizante y seguidos por fluorescencia. Para ello se transfiere una alícuota de proteína nativa a disoluciones con concentraciones variables de desnaturalizante. La concentración de desnaturalizante se determina a posteriori por el índice de refracción de la muestra (mostradas al final de la cinética). A) Parte inferior, experimentos de replegamiento seguidos por fluorescencia. En este caso la muestra de partida se encuentra desnaturalizada en una disolución a alta concentración de desnaturalizante. El replegamiento se induce por dilución. Igualmente, se determina la concentración de desnaturalizante final por la medida del índice de refracción. La determinación de la constante cinética se hace por ajuste a FLU13 y uso de FLU14. B) Representación del logaritmo natural de las constantes cinéticas frente a las concentraciones de desnaturalizante (representación usualmente conocida como Chevron). El código de colores indica de cual de las figuras del panel A se extrajeron los

datos. El análisis de este tipo de representación ha sido ampliamente descrito en diferentes trabajos (Beasty, Hurle et al. 1986), (Hurle and Matthews 1987; Hurle, Michelotti et al. 1987) (Perry, Onuffer et al. 1989) (Chen, Baase et al. 1989; Chen and Schellman 1989) (Ibarra-Molero and Sanchez-Ruiz 1996). En recta discontinua se representan los valores de  $\ln k_U$  y  $\ln k_F$  obtenidos de a partir de la ecuación FLU12, cuyos valores sólo se aproximan a los del  $\ln k_{app}$  en las regiones alejadas de la transición (a alta y baja concentración de desnaturalizante para  $\ln k_U$  y  $\ln k_F$  respectivamente).

#### 5.2.4.- Ensayos para la búsqueda de intermedios de equilibrio y cinéticos.

Una vez caracterizada nuestra proteína a través de los experimentos de equilibrio y cinéticos descritos anteriormente (apartados 5.2.2 y 5.2.3), no es suficiente su conformidad con los modelos propuestos para asumir que hemos determinado correctamente el mecanismo de desplegamiento/replegamiento que sigue la proteína. Nos encontramos en la misma situación descrita anteriormente (apartados DSC 5.1.3.1 Y DSC 5.1.3.3), donde necesitábamos una prueba independiente de la validez del modelo de dos estados.

Hay que tener en cuenta que en los experimentos descritos seguimos variaciones en la fluorescencia de triptófanos en función a cambios en el ambiente que los rodea. Por tanto cabe la posibilidad de que existan intermediarios en el equilibrio e intermediarios cinéticos cuyos triptófanos no experimenten cambios en su ambiente con respecto al estado nativo. En este caso, no serían detectados y nuestro modelo, aunque ajustase bien los datos, no sería válido.

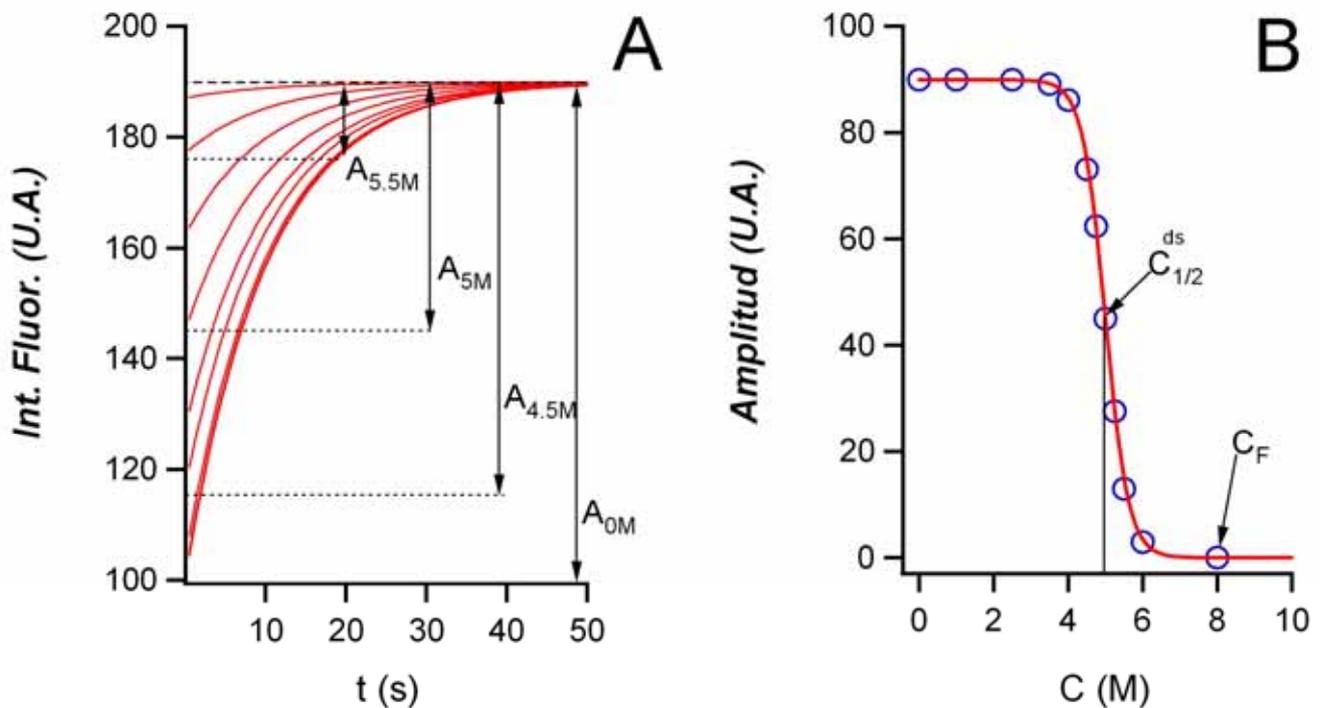
Clásicamente, la coherencia entre experimentos hechos con distintas técnicas espectroscópicas (que miden propiedades diferentes de nuestro sistema) se ha tomado como el mejor indicador de la validez del modelo de dos estados. Alternativamente podemos desarrollar una metodología experimental ligeramente distinta que, bajo ciertas suposiciones, nos permite confirmar que el modelo que hemos usado para ajustar nuestros datos de equilibrio y cinéticos es un buen modelo para describir el comportamiento de la proteína.

#### **5.2.4.1.- El experimento de Doble Salto para la detección de intermedios de equilibrio.**

Este experimento está dirigido a detectar posibles intermedios en el equilibrio de desplegamiento en proteínas (Mucke and Schmid 1994) (Schmid 1992). Se basa en que dicho intermedio presentaría una barrera energética al desplegamiento mucho menor que el estado nativo. Asumiendo esto como cierto, sería posible detectarlo con la siguiente metodología experimental:

Se preparan alícuotas de proteína en presencia de de diferentes concentraciones de desnaturizante. Cada alícuota debe tener la misma cantidad de proteína y una cantidad de desnaturizante conocida. Después de esperar un tiempo prudencial (a fin de que se equilibren las poblaciones de nativo, desplegado y cualquier posible intermedio) tendríamos la curva de equilibrio descrita en la figura FLU2. Iniciamos el experimento midiendo la cinética de desnaturización cuando cada una de estas alícuotas es llevada a una concentración alta de desnaturizante (Figura FLU5-A).

Es muy importante que la concentración final de desnaturizante sea igual en todos los experimentos. Resulta también fundamental que las concentraciones de proteína sean iguales para cada uno de los ensayos y que comencemos a medir las muestras exactamente en el mismo momento. Si esto se cumple, todas las cinéticas tendrán el mismo tiempo de relajación y su amplitud será proporcional a la fracción de estado nativo (Figura FLU5-B). La suposición hecha era que cualquier intermedio de equilibrio presentaba una barrera al desplegamiento mucho menor que el estado nativo. Por ello, cuando se expone a una concentración alta de desnaturizante despliega dentro del tiempo muerto y no contribuye ni a la amplitud ni a la constante de velocidad aparente.



**FIGURA FLU5:** Simulación de un experimento de doble salto. A) Cinéticas de desnaturalización de alícuotas de proteína incubadas en distintas concentraciones de desnaturalizante al ser transferidas a condiciones fuertemente desnaturalizantes. Todas las cinéticas tienen el mismo tiempo de relajación. El posible intermedio desnaturalizaría dentro del tiempo muerto del experimento y no contribuye, por tanto, a la señal. B) Representación de las amplitudes obtenidas del ajuste de las cinéticas de A) a la ecuación FLU13, frente a la concentración de desnaturalizante a la que se realizaron dichas cinéticas. El análisis de la curva se realiza con el segundo término de la ecuación FLU5, la variable independiente pasa a ser la amplitud y sustituimos  $\Delta_N^I I$  por un parámetro,  $A$ , que representa el cambio de amplitud en intensidad de fluorescencia entre el estado nativo y el estado desplegado final que obtenemos. Si los valores de  $C_{1/2}^{ds}$  y  $m_{1/2}^{ds}$  son iguales a los derivados de la Curva de Equilibrio y del Chevron, ello es indicativo de la validez del modelo de dos estados reversible por desnaturalización con agentes químicos.

Si la curva obtenida es la misma que la obtenida en experimentos de equilibrio o es consistente con el Chevron ( $C_{1/2}^{ds} = C_{1/2}^{chev}$  y  $m_{1/2}^{ds} = m_U^{\ddagger, chev} + m_F^{\ddagger, chev}$ ), consideraremos válido el modelo de dos estados reversible por desnaturalización con agentes químicos.

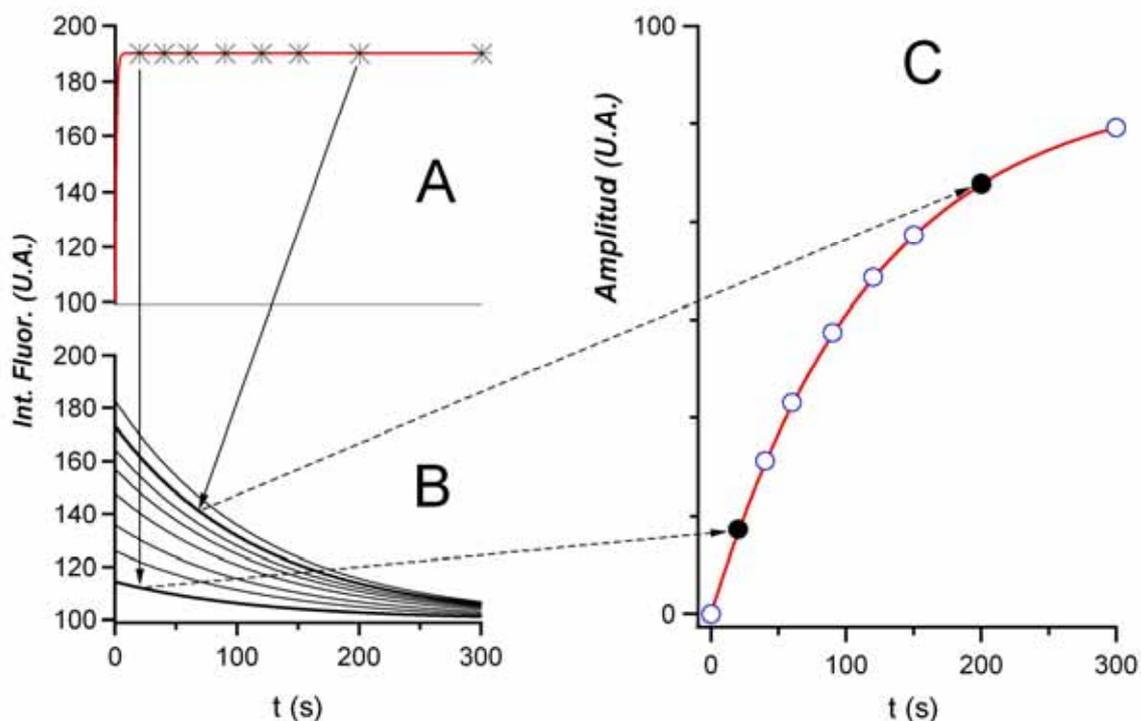
#### **5.2.4.2.-El experimento de Doble Salto Cinético**

Este experimento está dirigido a detectar posibles intermedios cinéticos (Kiefhaber and Schmid 1992). Se basa en el mismo principio anterior, donde los intermedios no experimentan la misma barrera energética que los estados nativo/desplegado al desplegamiento/replegamiento. Ello permite seguir independientemente el desplegamiento/replegamiento del estado nativo/desplegado ya que estos tienen una barrera mucho mayor que los separa de los demás. Sin embargo en este caso el ensayo se puede dirigir tanto a analizar las fases cinéticas en el plegamiento como en el desplegamiento, puesto que ambas fases no tienen por qué ser equivalentes.

La metodología consiste en analizar las amplitudes de la cinética de regreso de un experimento de plegamiento o desplegamiento como función del tiempo. Para ello prestaremos especial atención en mantener constantes las concentraciones de proteína y las concentraciones de desnaturalizante iniciales y finales (en este caso la concentración de desnaturalizante no va a ser una variable).

Como ejemplo, detallaremos la metodología en un experimento de doble salto cinético en la dirección de desplegamiento (examinar la fase de desplegamiento). Tomamos una muestra de proteína y la llevamos a una alta concentración del desnaturalizante (Figura FLU6-A). Este será nuestro tiempo cero. Vamos tomando alícuotas cada cierto tiempo y las llevamos a una disolución en ausencia del desnaturalizante (Figura FLU6-B). Finalmente, tomaremos la amplitud de esta última cinética y la representamos frente al tiempo. En caso de existir una fase cinética "oculta", la amplitud de esta última cinética irá variando de alícuota a alícuota (es decir, con el tiempo) y al representar la amplitud frente al tiempo, obtendremos una curva exponencial de cuyo análisis sacaremos el tiempo de relajación del fenómeno (Figura FLU6-C).

La ausencia de fases cinéticas "ocultas" en este análisis no es suficiente para garantizar la validez del modelo de dos fases cinéticas (esto es, un solo proceso cinético de desplegamiento y un solo cinético de replegamiento)



**Figura FLU6:** Simulación de un doble salto cinético a partir de la simulación de la Figura FLU4. Las flechas indican el proceso secuencial que se ha de seguir cuando trabajemos en el laboratorio. A) Simulación de la cinética que vamos a examinar ( $\tau = 1s$ ,  $8.52 M$ ). A la hora de realizar el experimento, se transfiere una disolución concentrada de proteína a una disolución concentrada de desnaturizante (no hace falta que registremos la cinética). En este momento se inicia el contador de tiempo. B) Cinéticas después de diluir una alícuota de la disolución anterior ( $\tau = 120s$ ,  $1.46M$ ). Como se puede observar, las amplitudes varían según el momento en el que realicemos el experimento. C) Representación de las amplitudes obtenidas en B frente a los tiempos a los que se tomaron de A las alícuotas. Se obtiene una exponencial con un tiempo de relajación de 300 segundos. El caso simulado es un clásico ejemplo de isomerización de Prolinas en el estado desnaturizado. Las Prolinas en el estado nativo suelen encontrarse casi en exclusividad como isómero -cis. Cuando desnaturizamos la proteína, gracias a la disminución de las restricciones estéricas, las Prolinas isomerizan lentamente de las

formas -cis a -trans (cinética descrita por el panel C). Al realizar la transferencia a baja concentración de desnaturizante, las proteínas, para poder plegarse, tienen que volver a isomerizar todas sus Prolinas a -cis (cinética descrita en el panel B).

## **APENDICE 3**

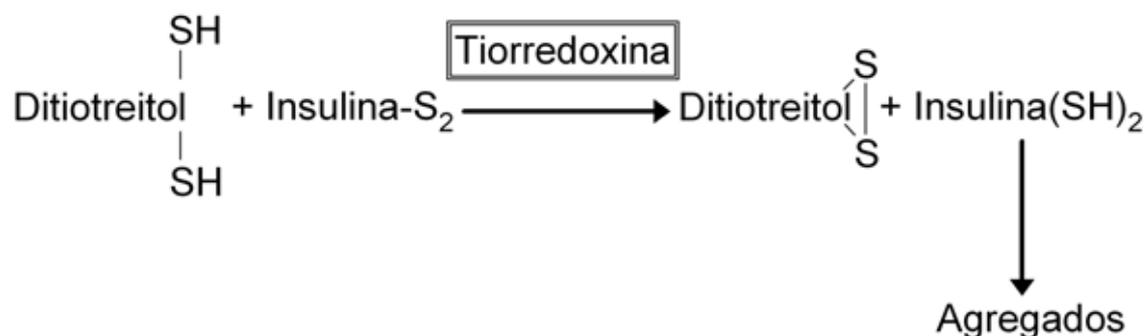
### **ENSAYOS DE ACTIVIDAD DE TIORREDOXINA Y LIPASA**

## 6.1-Ensayo de Actividad de Trx-Insulina

### 6.1.1- Generalidades

Tiorredoxina es una proteína que pertenece a la familia de las oxidoreductasas. Tras ser reducida por la tiorredoxina reductasa (dependiente de NADPH) es capaz de reducir grupos cisteína oxidados en otras proteínas (Powis and Montfort 2001). Si bien, como en el presente ensayo, tiorredoxina también puede ser reducida por agentes químicos.

El presente ensayo fue desarrollado por Arne Holmgren (Holmgren 1979) y se sigue con ligeras modificaciones. En él se mide espectrofotométricamente la turbidez de una disolución de insulina como función del tiempo a la vez que tiorredoxina reduce los puentes disulfuro de acuerdo con el siguiente esquema de reacción:

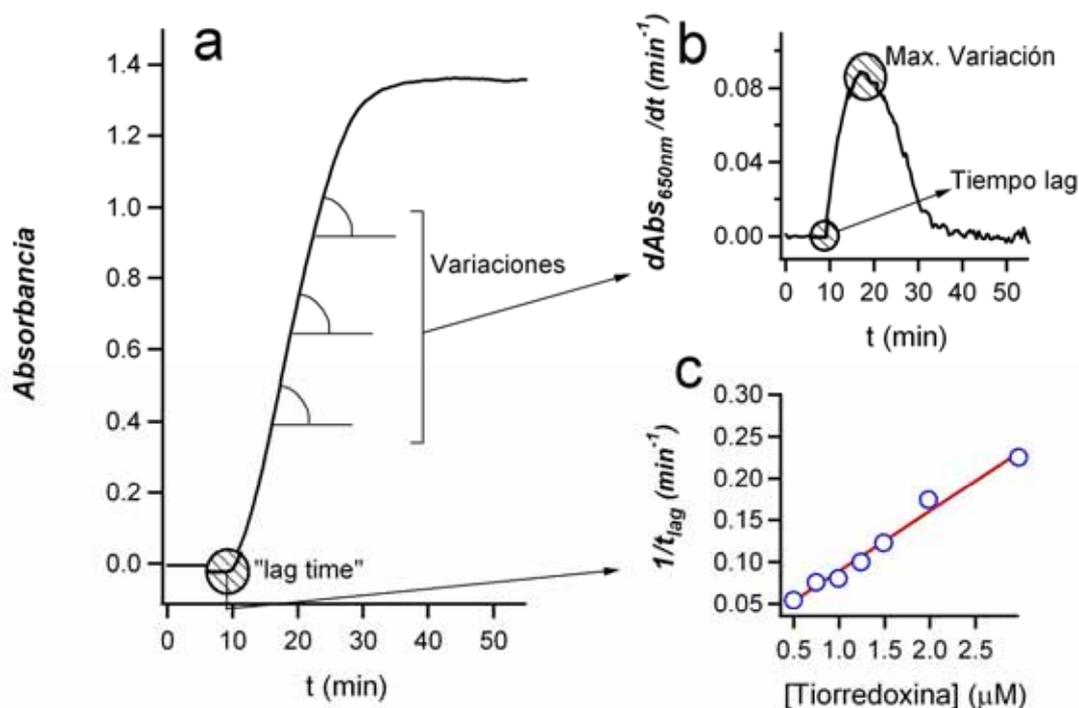


**Figura INS1:** Esquema de la reducción de insulina por DTT catalizada por tiorredoxina. La posterior agregación provoca un aumento de la absorbancia a 650nm.

La insulina es una proteína compuesta de dos cadenas polipeptídicas (A y B) unidas entre sí por dos enlaces disulfuro. Tras la reducción por parte de la tiorredoxina los enlaces disulfuro se vuelven a formar de un modo aleatorio entre distintas cadenas, de modo que se van produciendo agregados de insulina. Cuando estos agregados llegan a tener un tamaño considerable comienzan a dispersar la luz, lo cual resulta en un aumento de la absorbancia a 650nm.

La actividad de la tiorredoxina está relacionada con la velocidad a la cual agrega la insulina. El parámetro más útil para caracterizar la actividad es la velocidad máxima a la cual agrega la insulina para una concentración de tiorredoxina dada FIG INS2. Si bien

también es posible utilizar como parámetro el tiempo transcurrido hasta que se hace aparente la agregación ("lag time"). FIG INS2



**Figura INS2:** a) Ensayo de actividad de tiorredoxina wt a 1.48 μM, se indican cuáles son los parámetros que usaremos para estimar la actividad (bien el "lag time" o bien la variación de absorbancia). b) Se muestra la derivada de cada una de las líneas rectas que ajustan 3 valores de absorbancia consecutivos. A partir de esta gráfica podemos tomar la variación máxima o el "lag time" como índice de la velocidad de reacción. c) Se muestra la recta patrón para tiorredoxina wt cuando nuestro parámetro de actividad es el "lag time".

### 6.1.2- Metodología experimental

Así como el DTT debe ser preparado fresco cada día, el buffer de reacción (fosfato potásico 0.1M, EDTA 2mM, pH 6.5) y el stock de insulina (10 mg/mL) pueden almacenarse a -20 °C.

La preparación del stock de insulina es uno de los puntos en los que hay que prestar especial atención. En primer lugar hay que asegurarse de que la insulina que hemos comprado esté en el rango de las 30 unidades USP/mg. Posteriormente, pesamos en una balanza de precisión 50mg de insulina. Los disolvemos en un vaso de precipitado

de 5mL con 4mL de buffer Tris-HCl 50mM pH 8.3 y en agitación suave hasta que se disuelva completamente.

A continuación introducimos el electrodo fino del pHmetro y monitorizamos continuamente el pH. Con el fin de eliminar cualquier posible microagregado realizaremos un salto de pH del siguiente modo: Se añade gota a gota una disolución de HCl 1M hasta llegar a un pH entre 2 y 3 (aproximadamente 150 $\mu$ L). La disolución se volverá de un aspecto lechoso al pasar por pH 4-5 y al llegar a pH 3 se tornará nuevamente translúcida (en ocasiones hace falta aumentar ligeramente la agitación a fin de conseguir el completo clarificado de la disolución). Una vez lleguemos a este punto hay que devolver la disolución a pH 8.3 rápidamente, para lo cual añadiremos gota a gota una disolución 1M de NaOH (aproximadamente 150 $\mu$ L).

Para llegar a una concentración de 10mg/mL enrasamos con agua en un matraz aforado de 5mL. Con el fin de evitar pérdidas debido a restos de la disolución adheridos al electrodo y al vaso de precipitados, lavamos ambos con el agua que usaremos para enrasar. Los 5 mL finales se dividen en alícuotas convenientemente y se congelan a -20 °C (hay que tener en cuenta que por precaución una vez descongelada una alícuota ya no se volverá a congelar).

Una vez tengamos todos los reactivos preparados, se conservan en hielo. Se termostatan las cubetas en el espectrofotómetro y se selecciona la longitud de onda de medida (650 nm). En cada una de las cubetas mezclamos los reactivos en las cantidades y orden siguientes:

-784 $\mu$ L de buffer fosfato potásico 0.1M, EDTA 2mM, pH 6.5.

-50 $\mu$ L de insulina 10mg/mL

-166 $\mu$ L de proteína en Hepes 5mM, pH 7

-10 $\mu$ L de DTT 0.1M

A continuación mezclamos con una pipeta pasteur e iniciamos la medida. Si nuestra intención es medir el "lag time" (ver figura INS2-A), deberemos iniciar un cronómetro en el momento de añadir el DTT, en caso contrario sólo es necesario no demorarse más allá de 2 minutos antes de iniciar la medida (ver figura INS2-A).

Los principales factores a tener en cuenta a la hora de realizar el experimento serán:

-Medida la absorbancia en un espectrofotómetro como función del tiempo a una longitud de onda de 650nm durante unos 40 minutos.

-Control preciso de la temperatura a la cual se desarrolla el experimento (37 °C). Para ello, se utilizará un baño termostaticado que recircule agua alrededor de las celdas donde se lleva a cabo la reacción.

-La concentración de ditioneitol (DTT) debe ser 1mM final. Las medidas de actividad varían fuertemente con la concentración de este reactivo, dado que es el encargado de reducir en un primer momento a la tiorredoxina. Este reactivo se tiene que preparar fresco dado que se degrada fácilmente.

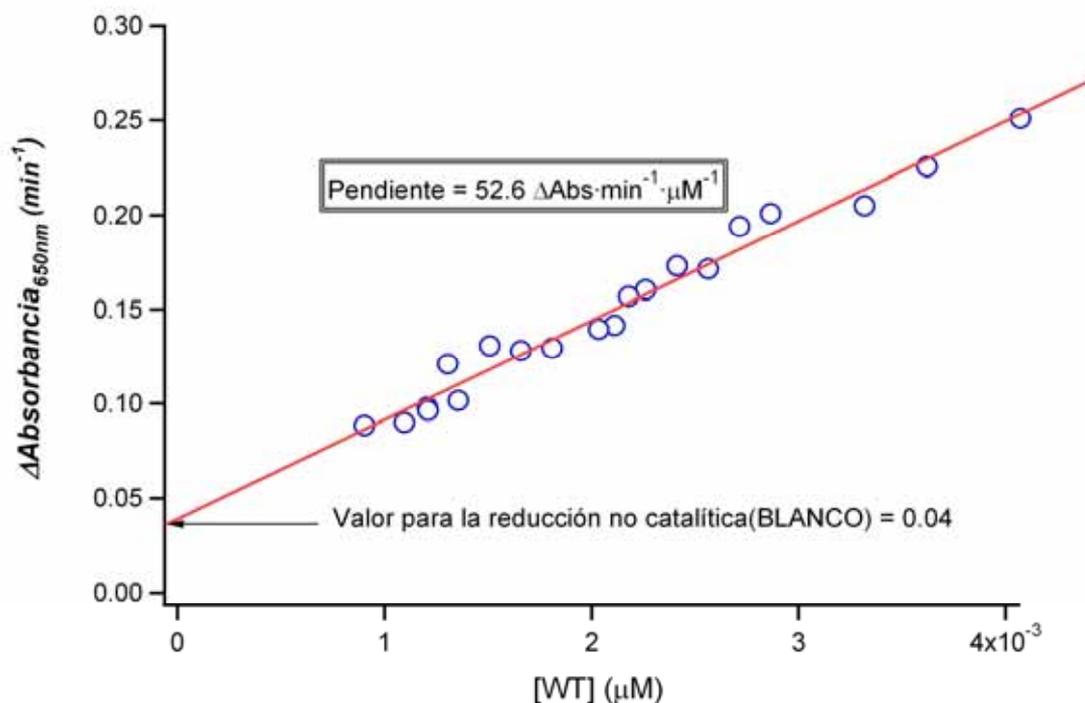
-La concentración de nuestro stock de insulina debe ser de 10mg/ml. De esta concentración depende la amplitud de la cinética, y variaciones en esta amplitud afectan a la velocidad.

-El buffer de reacción es fosfato potásico 0.1M, EDTA 2mM y pH 6.5. Hay que tener en cuenta que los valores de actividad dependen fuertemente del pH.

-La variación de la velocidad máxima de agregación con la concentración de tiorredoxina es sólo lineal dentro de un rango. Cuando la velocidad es superior a 0.2-0.25 U.A./min se pierde la linealidad con la concentración. Por ello es mejor trabajar con muestras de proteína hasta 3µM (0.035mg/mL) final (ver FIG INS3).

El componente no catalítico de la reacción, es decir, la reducción de los puentes disulfuro de la insulina por el DTT, se obtiene de un experimento en el que no se añade tiorredoxina. Si bien a nivel práctico resulta más apropiado obtenerlo por extrapolación de la recta patrón, dado que el tiempo para la reducción no catalítica es demasiado largo y acaban apareciendo artefactos que distorsionan los resultados (sedimentación de los agregados) (FIG INS3).

Finalmente destacar que siempre que realicemos ensayos de actividad sobre formas mutantes se ha de realizar un número de experimentos con la forma silvestre (wt) y comprobar que los datos obtenidos caen en la recta patrón (FIG INS3).



**Figura INS3:** Recta patrón para el ensayo de actividad de tiorredoxina por agregación de la insulina. El conjunto de datos proviene de experimentos realizados en distintos días y con stocks de reactivos y proteína diferentes.

## 6.2-Ensayo de Actividad de Trx-TrxReductasa

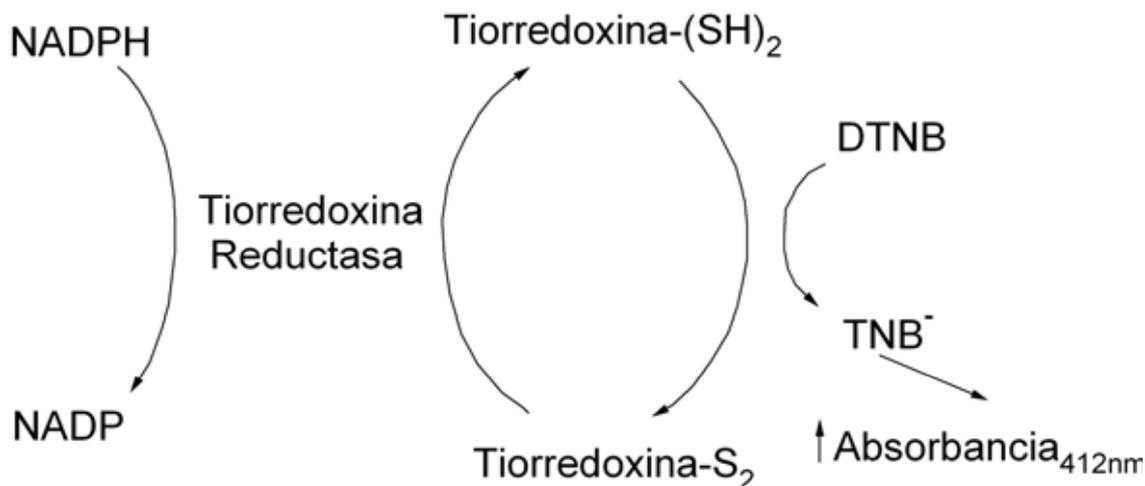
### 6.2.1-Generalidades

Tiorredoxina es una proteína que pertenece a la familia de las oxidoreductasas. Tras ser reducida por la tiorredoxina reductasa (dependiente de NADPH) es capaz de reducir grupos cisteína oxidados en otras proteínas (Powis and Montfort 2001).

El presente ensayo de actividad se basa en la capacidad del 5, 5'-ditiobis-(ácido 2-nitrobenzoico) (DTNB, también conocido como reactivo de Ellman) para reducirse en la presencia de grupos tiol (S-OH), liberándose ácido 2-nitro-5mercaptobenzoico, que se ioniza en

disolución a pH básico y resulta en un producto de un fuerte color amarillo (Ellman 1959).

Fue desarrollado por Laurent, Moore y Reichard (Laurent, Moore et al. 1964) y posteriormente modificado por Slaby y Holmgren (Slaby and Holmgren 1975). Es "grosso modo" una cuantificación de los grupos tiol de tiorredoxina, que tras ser oxidados por el DTNB vuelven a reducirse por un sistema que implica a la tiorredoxina reductasa:



**Figura RED1:** Esquema de la reducción de DTNB por tiorredoxina en un sistema dependiente de NADPH y tiorredoxina reductasa.

Hay que tener en cuenta que en el presente ensayo de actividad no sólo estamos viendo la capacidad catalítica de tiorredoxina, sino también la efectividad en la reducción de ésta por la tiorredoxina reductasa. Por tanto, la velocidad a la cual se forma el producto de la reacción dependerá linealmente tanto de la concentración de tiorredoxina como de la concentración de tiorredoxina reductasa.

### 6.2.2-Metodología experimental

Este ensayo de actividad precisa de tiorredoxina-reductasa comercial, cuya concentración puede variar de un lote a otro, por lo que como se indicó anteriormente necesitamos hacer siempre un control con tiorredoxina wt, de seroalbúmina bovina a 5 mg/mL, EDTA 0.2M, buffer Tris-HCl 1M pH8, NADPH a 10mM, DTNB 50mM en etanol y por supuesto tiorredoxina, con la que normalmente trabajamos en hepes 5mM

pH7. Una vez tengamos todos los reactivos preparados, se conservan en hielo. Se termostatan las cubetas en el espectrofotómetro y se selecciona la longitud de onda de medida (412nm). En cada una de las cubetas mezclamos los reactivos en las cantidades y orden siguientes:

-865 $\mu$ L de  $\text{adH}_2\text{O}$

-50 $\mu$ L de Tris-HCl 1M, pH 8

-10 $\mu$ L de EDTA 0.2M, pH 7.5

-10 $\mu$ L de BSA 5mg/mL en  $\text{adH}_2\text{O}$

-25 $\mu$ L de NADPH 10mM en  $\text{adH}_2\text{O}$

-10 $\mu$ L de tiorredoxina en Hepes 5mM, pH 7

A continuación se esperan unos minutos para que se termostaticen las muestras y añadimos:

-20 $\mu$ L de tiorredoxina reductasa (preparada añadiendo 20 $\mu$ L del stock comercial en 980 $\mu$ L de buffer Tris 30mM, EDTA 1mM, pH 8.3)

-10 $\mu$ L de DTNB 50mM en etanol.

Justo a continuación se mezcla con ayuda de una pipeta pasteur y se inicia la medida. Es importante que los volúmenes pipeteados sean exactos, por lo que se usarán preferentemente jeringuillas de precisión en lugar de micropipetas.

El efecto de la reducción no catalítica del DTNB se obtiene a partir de un experimento en el que no se añade tiorredoxina. El resultado debe ser una línea recta de pendiente cero.

El resultado en cada uno de los ensayos es una línea recta de cuya pendiente obtenemos el valor de la actividad para la concentración de proteína que estuviéramos usando (FIG RED2-A). A partir de experimentos a distintas concentraciones, obtenemos una recta cuya pendiente es proporcional a la actividad específica (FIG RED2-B).

Los principales factores a tener en cuenta a la hora de realizar el experimento serán:

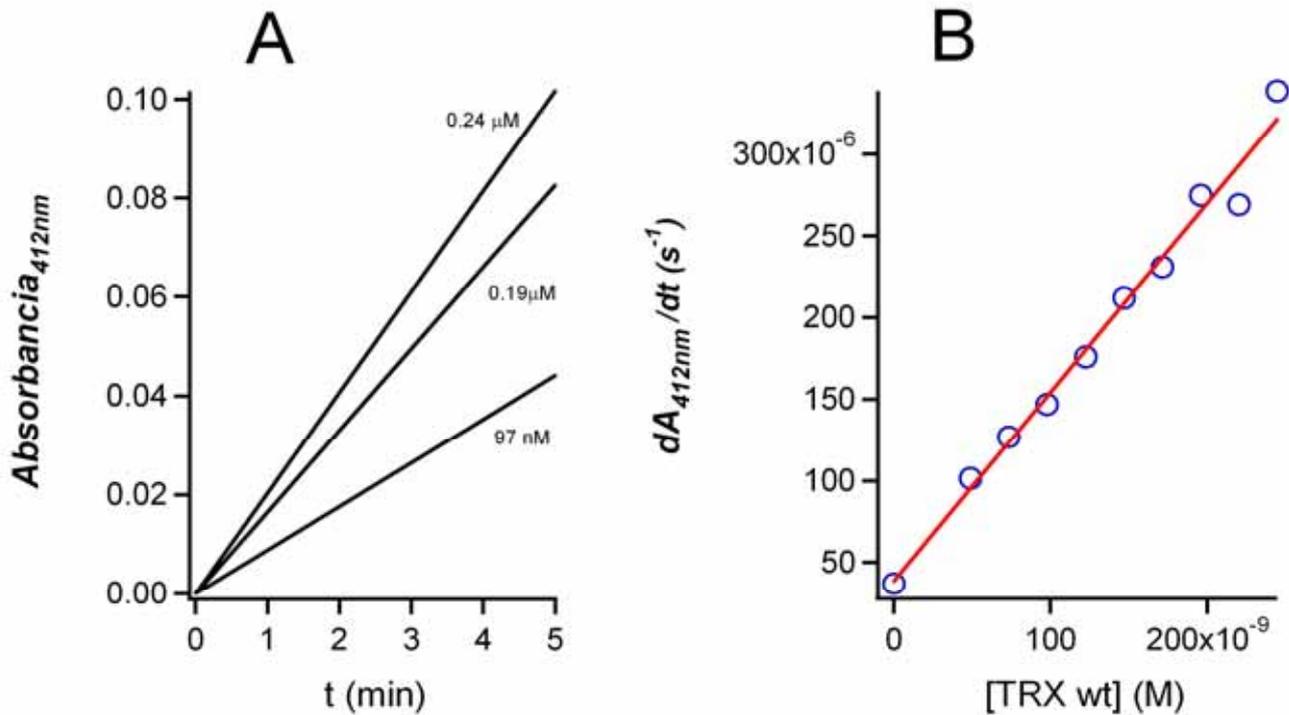
-Medida la absorbancia en un espectrofotómetro como función del tiempo a una longitud de onda de 412nm durante 5 minutos.

-Control preciso de la temperatura a la cual se desarrolla el experimento (37 °C). Para ello, se utilizará un baño termostático que recircule agua alrededor de las celdas donde se lleva a cabo la reacción.

-Dado que la actividad dependerá de la tiorredoxina reductasa, será necesario hacer siempre una recta patrón de wt con cada alícuota nueva de tiorredoxina reductasa y referenciar nuestros resultados a ella.

-El NADPH es el único reactivo que debe prepararse fresco cada día dada su inestabilidad.

-El rango de concentraciones en el cual el ensayo es lineal es bastante amplio, pero procuraremos no exceder 0.5µM final de Tiorredoxina (0.006 mg/mL finales).

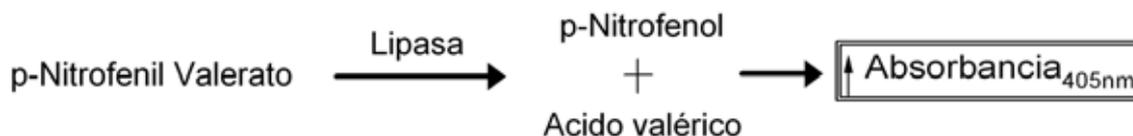


**FIG RED2:** A-Experimentos de determinación de actividad de tiorredoxina en la reducción acoplada a tiorredoxina reductasa de DTNB. A- Tres cinéticas de reducción del DTNB a distintas concentraciones (0.24μM, 0.19μM y 97nM) B-Recta patrón de tiorredoxina wt obtenida, entre otras, de las pendientes de A.

### 6.3-Ensayo de Actividad de Lipasa

#### 6.3.1-Generalidades

Las lipasas son unas de las proteínas con más usos biotecnológicos. Son capaces de hidrolizar enlaces éster, frecuentemente encontrados en sustancias lipídicas. Es posible encontrar en la bibliografía numerosos ensayos de actividad (Gupta, Rathi et al. 2003). La metodología seguida en este trabajo fue proporcionada gentilmente por Novozymes. Esencialmente sigue el siguiente esquema de reacción:



**Figura LIPl:** Esquema de la reacción de hidrólisis del enlace éster del p-Nitrofenil Valerato por parte de la lipasa.

Esencialmente se trabaja con un sustrato que lleva un enlace éster. Tras la hidrólisis del mismo resulta p-nitrofenol, que absorbe a 405nm. La velocidad de la reacción para cada concentración de proteína viene dada por la variación en la absorbancia por unidad de tiempo.

### 6.3.2-Metodología experimental

El buffer de dilución se compone de: 0.4% Triton X-100, 10mM CaCl<sub>2</sub>, 50mM Tris-HCl, pH 7. El sustrato se prepara diluyendo 117µL de p-nitrofenil valerato en 10 mL de isopropanol (dando una disolución con un ligero color amarillento, si este color gana en intensidad con el tiempo deberemos preparar nuevamente el sustrato). La lipasa, en los trabajos realizados, se encontraba en buffer acetato sódico 50mM pH5 a una concentración aproximada de 0.2 mg/mL.

La disolución de trabajo se obtiene al diluir el sustrato 1:100 con buffer de dilución (100µL en 9.9mL).

Una vez tengamos todos los reactivos preparados, se conservan en hielo. Se termostatan las cubetas en el espectrofotómetro y se selecciona la longitud de onda de medida (405nm). Preparamos la siguiente disolución:

Añadimos 990µL de disolución de trabajo a cada cubeta y la reacción comienza en el momento de añadir 10µL de proteína (en nuestro caso a pH 5, AcNa 50mM) a cada una de las cubetas con una jeringuilla Hamilton. Tras ello mezclamos bien con una pipeta pasteur y comenzamos la medida.

Como se muestra en la figura LIP2, analizaremos la dependencia de la velocidad de la reacción con la concentración de lipasa para obtener un valor proporcional a la actividad específica.

Los principales factores a tener en cuenta a la hora de realizar el experimento serán:

-Medida la absorbancia en un espectrofotómetro como función del tiempo a una longitud de onda de 405nm durante unos 15 minutos. Tomaremos como resultado la parte lineal del experimento.

-Control preciso de la temperatura a la cual se desarrolla el experimento (25 °C). Para ello, se utilizará un baño termostaticado que recircule agua alrededor de las celdas donde se lleva a cabo la reacción.

-Tanto el buffer de dilución como el sustrato pueden almacenarse a -20°C durante un mes. Pero la mezcla de ambos es inestable y se debe preparar fresca cada día.

-El ensayo mantiene la linealidad con la concentración de proteína al menos hasta los 0.4µg/mL, pero a concentraciones más altas la reacción deja de ser de orden 0, por lo que procuraremos no exceder esta concentración final de proteína.

-El pH tiene que estar correctamente controlado, dado que a pH ácido el producto de la reacción carece de absorbancia y varía fuertemente entre pH neutro y pH básico.

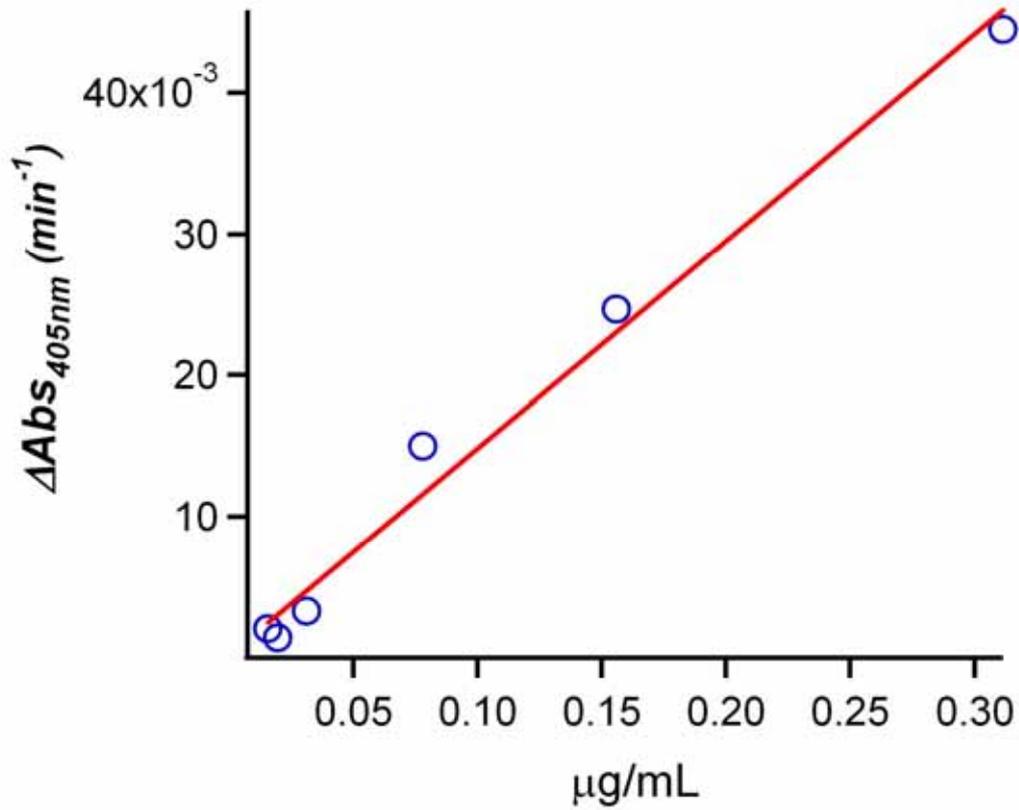


Figura LIP2: Se muestra la recta patrón de la lipasa de *Thermomyces Lanuginosa* obtenida a partir de la velocidad de degradación del p-nitrofenil valerato a distintas concentraciones de proteína. En el trabajo 3 se realizaron medidas comparativas entre wt y la forma alternativa de la lipasa (misma concentración de sustrato).

## **APENDICE 4**

# **PURIFICACIONES DE PROTEÍNA**

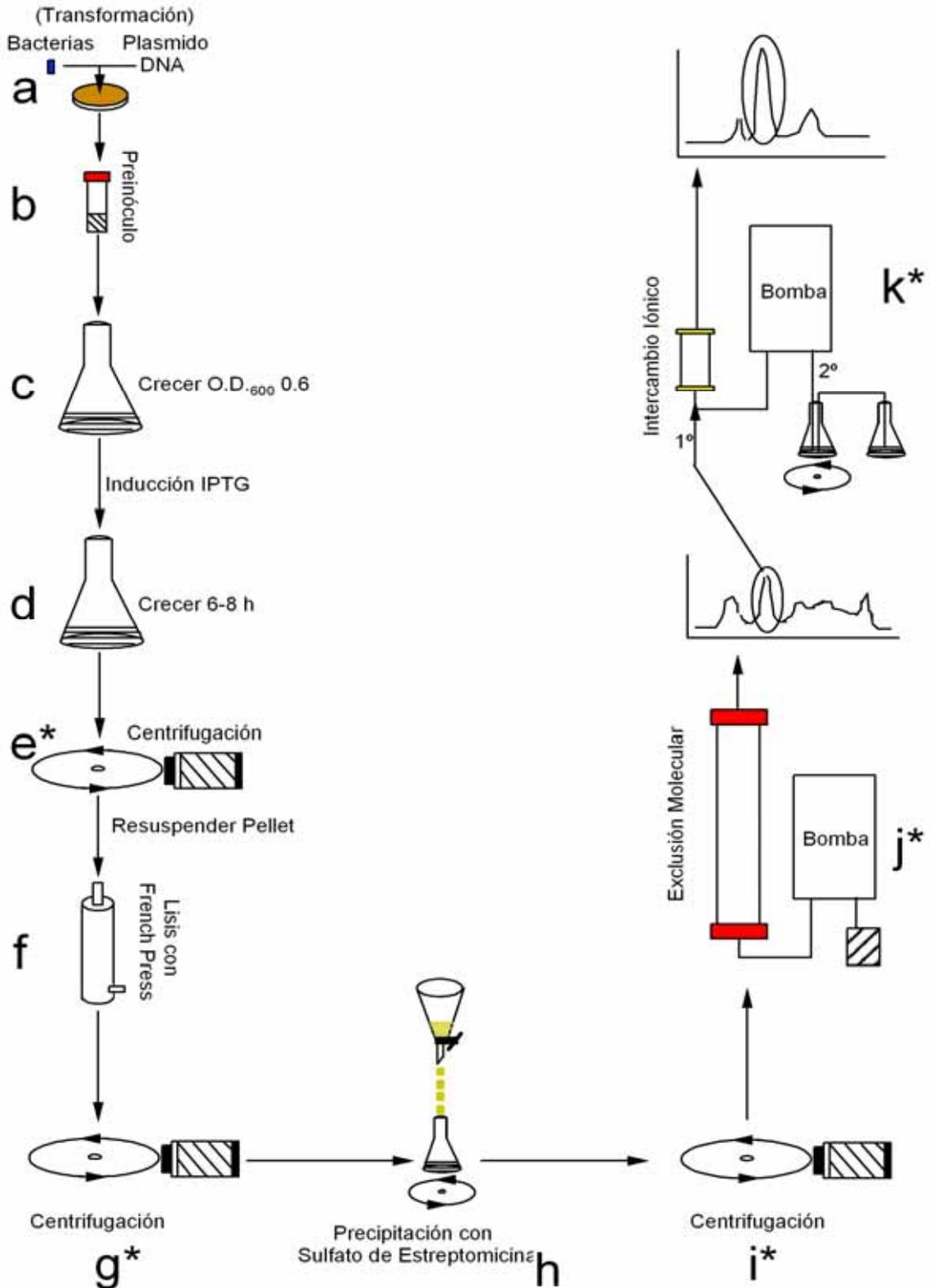
Con el fin de estudiar las propiedades biofísicas de una proteína de interés, la mayor parte de las técnicas utilizadas en este trabajo requieren de unos niveles de pureza considerables (>95 %). Es por tanto necesario llevar a cabo diversos procesos que nos permitan obtener, en principio, una cantidad apropiada de proteína con un nivel de pureza adecuado.

Si bien es posible producir pequeñas proteínas a través de síntesis orgánica en un laboratorio, dicho procedimiento resulta considerablemente caro. Por tanto, el método más frecuente para obtener nuestra proteína es directamente de un sistema biológico.

Hoy en día se usan para dicha finalidad organismos vivos (E. coli, levaduras, células de insecto,...) así como técnicas de biología molecular estándares que no se detallarán en el presente apartado. Aunque en algunos casos, donde la purificación de la proteína presenta complicaciones, se utilizan sistemas de expresión libres de células.

### **7.1- Purificación de tiorredoxina de E.coli y mutantes.**

# Purificación Trx



**Figura PUR1:** Esquema de la purificación de tiorredoxina. Los pasos están enumerados por orden de "a" a "k". El asterisco denota aquellos procesos tras los cuales es posible congelar la muestra.

El esquema general de este tipo de purificación está expuesto en la FIG PUR1. El proceso de purificación de la proteína parte del proceso de transformación de las cepas de E. coli BL21 DE3 con un vector pET 30a(+) que contiene un gen de resistencia a Kanamicina y el gen que codifica nuestra proteína de interés (en este caso tiorredoxina silvestre (wt) o bien alguna forma mutante) (FIG PUR1-a).

La transformación se realizó siguiendo el protocolo comercial, con la salvedad de que el volumen de células utilizado era de 10µl, por lo cual la cantidad de medio SOC precalentado a 42 °C tuvo que ser de 200µl. Después de que el producto de transformación se dejara incubar a 37 °C con agitación se plaqueó en LB-agar suplementado con Kanamicina (30µl/ml final).

Tras 14-16 horas a 37 °C, las colonias son aparentes y se pueden pasar a medio LB líquido (30 µl/ml final de Kanamicina) con ayuda de un asa de siembra (FIG PUR1-b). Estos preinóculos se dejan crecer hasta nuestra conveniencia a 37 °C y en agitación (con un tiempo máximo prudencial de unas 16 horas).

Los preinóculos sirven para iniciar el crecimiento a mayor escala en matraces con 500ml de LB suplementados con Kanamicina a 30µl/ml finales (FIG PUR1-c) (el volumen total de medio dependerá de cuánta proteína queramos obtener y del rendimiento que obtengamos. En términos generales, cada litro de cultivo produce aproximadamente 50-100mg de proteína). Cuando este cultivo alcanza una absorbancia de 0.6 a 600nm, se induce la expresión de la proteína de interés con IPTG a 0.4mM final (FIG PUR1-d). Se continúa este crecimiento a 37 °C con agitación durante 6-8 horas (si bien los niveles de expresión de proteína son aceptables en el rango de 4 a 16 horas).

Posteriormente se separan las bacterias del medio de crecimiento por centrifugación (7000r.p.m. durante 10 minutos en el rotor A6.9 de Herolab a 4 °C; este tiempo y esta velocidad son valores estandar, pudiéndose recortar considerablemente) (FIG PUR1-e\*). Una vez aislado el pellet de bacterias se resuspende en lo que será nuestro buffer de purificación (T.E., compuesto de Tris-HCl 30mM, EDTA 1mM y 1 g de azida cada 5 litros, pH 8.3; filtrado y desgasificado). En caso de tener bajos rendimientos en la expresión, podemos disolver una pastilla de inhibidores de proteasas (en nuestro caso Complete de Roche) en 40-60ml de T.E. y usar este volumen para resuspender las

células. Hay que tener en cuenta que los inhibidores de proteasas contienen quelantes de iones, que si bien no interfieren con ningún proceso del presente protocolo de purificación, pueden interferir con procedimientos utilizados en otras purificaciones (principalmente aquellos que utilizan cromatografías de afinidad hacia cationes divalentes) o interferir con la propia naturaleza de la proteína (aquellas que tienen iones como cofactores).

Tiorredoxina es una proteína que se produce en grandes cantidades y permanece soluble en el citoplasma celular. Por tanto será necesario llevar a cabo un procedimiento de lisis celular (FIG PUR1-f). Podremos optar por usar la French-Press o bien sonicación (pulsos de 45 segundos, 5-10 veces, cuando el volumen sea de unos 30-45ml y manteniéndolo siempre en hielo para evitar que se caliente la muestra). La sonicación en ocasiones presenta problemas de sobrecalentamiento, lo cual habrá que tener en cuenta cuando nuestra proteína no sea estable, y problemas de contaminación con ácidos nucleicos (ya que los fragmenta en una mezcla de tamaños heterogénea con los cuales tiorredoxina parece interactuar). Por experiencia, se ha observado que el método más eficaz para eliminar la contaminación por ácidos nucleicos en tiorredoxina es hacer una cromatografía de intercambio iónico (un procedimiento de esta purificación descrito en los párrafos subsiguientes). En cualquier caso la experiencia ha demostrado que la French-Press es un método seguro, limpio y rápido cuando el número de muestras no es elevado. Por tanto, es el procedimiento de lisis por defecto en esta purificación.

Tras la lisis repetimos la centrifugación descrita anteriormente para separar el lisado de cuerpos y restos celulares que no fueron lisados (FIG PUR1-g\*). El sobrenadante se deja a 4 °C en agitación muy suave mientras vamos decantando una disolución de Sulfato de Estreptomicina al 10% en T.E.. Esta decantación debe ser muy lenta para permitir la correcta homogeneización. El volumen decantado debe ser un cuarto del volumen de lisado celular (FIG PUR1-h).

Después de 6-12 horas, se centrifuga a 30.000r.p.m. en el rotor 45 Ti de Beckman durante 20 minutos a 4 °C (FIG PUR1-i\*). El sobrenadante será una mezcla de proteínas, lípidos, azúcares,... libre de ácidos nucleicos. Después procedemos a pasar la muestra por filtros de 45µm de tamaño de poro y estamos en condiciones de introducir el sobrenadante en la columna de exclusión molecular.

Dicha columna presenta una serie de características que la diferencian de las columnas de uso hoy en día en laboratorios donde se purifican proteínas rutinariamente. En primer lugar cabe resaltar su gran tamaño, lo cual implica que el volumen de carga que acepta la columna es mucho mayor (se han llegado a introducir hasta 120ml de muestra, obteniéndose una buena resolución de los picos). Por otro lado, la matriz utilizada presenta una baja capacidad de separación en comparación con las actualmente disponibles (compárese el poder resolutivo de la matriz utilizada, Sephacryl S-100, con el de otras matrices tipo Sephadex o Superdex). Sin embargo, dicho defecto se ve sobradamente compensado por la considerable longitud de la columna.

Antes de utilizar la cromatografía de exclusión molecular, es conveniente equilibrar la columna pasando a través de ella un volumen de columna (2 litros) de buffer de purificación. Este paso puede ser omitido si la columna se halla equilibrada en buffer T.E. de una purificación anterior. El óptimo de separación de la columna es de 5ml/min, flujo del cual no debemos alejarnos en ningún momento (dado que se alteraría el empaquetamiento). En la figura FIGPUR1-j\* se representa adicionalmente el cromatograma que se suele obtener bajo estas condiciones para esta proteína y se indica cuál es el pico que corresponde a tiorredoxina.

Adicionalmente, cabe indicar que es aconsejable llevar a cabo con cierta periodicidad un limpiado de la columna con Cloruro de Guanidinio 6M. Sería recomendable utilizar unos 300ml y a continuación pasar varios volúmenes de columna de agua doblemente destilada. Todos los búferes que utilicemos para la cromatografía de exclusión molecular deben estar adecuadamente filtrados (con filtros de 45µm de poro) y degasificados.

Dependiendo de los niveles de pureza que necesitemos y de la presencia de ácidos nucleicos en nuestra disolución de proteína, podemos realizar un paso adicional en la purificación. Se trata de una cromatografía de intercambio iónico que usa una resina cargada positivamente (Fractogel EMD DEAE-650). Al final de cada purificación hay que dejar la columna equilibrada en una alta concentración salina (para su óptima conservación), por lo que antes de comenzar a usarla deberemos equilibrarla con 2 volúmenes de columna (500ml aproximadamente) de buffer T.E. a fin de dejar la columna libre de

sal. Posteriormente introducimos las fracciones de la exclusión molecular que contienen nuestra proteína. La elución de los distintos componentes se realiza a través de un gradiente de T.E. de 0 a 0.5 M de NaCl a 1ml/minuto que provoca un debilitamiento gradual de las interacciones con los grupos cargados de la resina. Hay que resaltar que en esta ocasión no podemos confiar en que nuestra proteína salga siempre en el mismo volumen de elución dado que trabajamos frecuentemente con proteínas mutantes que interaccionan con distinta intensidad con la resina de intercambio. En la FIGPUR1-k\* se representa un cromatograma para este tipo de proceso (donde los picos contaminantes que en ocasiones aparecen se han exagerado en tamaño).

Por experiencia, se ha comprobado que las fracciones obtenidas tanto de la exclusión molecular como del intercambio iónico se pueden conservar durante largo período de tiempo congelándolas a -20 °C sin que ello altere las propiedades de nuestra proteína. Si bien cuando se va a realizar algún ensayo de cristalización no es recomendable que las muestras hayan sido congeladas en ningún momento.

Es interesante apuntar una optimización de tiempos en este tipo de purificación que permite obtener la proteína de interés en poco más de 24 horas. Consiste en realizar la transformación y su producto verterlo directamente sobre LB+Kanamicina para tener los preinóculos. Si realizamos esto a última hora de la tarde y guardamos los matraces para realizar el cultivo a gran escala a 37 °C, a la mañana del día siguiente podemos iniciar el crecimiento en los matraces. En tan sólo 30-45 minutos podremos inducir la expresión, de modo que podremos centrifugar, lisar y volver a centrifugar justo antes de las 14:00 para dejar incubando con el Sulfato de Estreptomicina. A última hora de la tarde podremos centrifugar para obtener el sobrenadante, que tras el filtrado se puede inyectar en la columna de exclusión molecular. Gracias al uso de un FPLC y de la reproducibilidad en los volúmenes de elución, a la mañana siguiente tendremos nuestra proteína con unos niveles de pureza superiores al 95%.

## **7.2-Purificación de bibliotecas de mutantes de Trx-hisTag.**

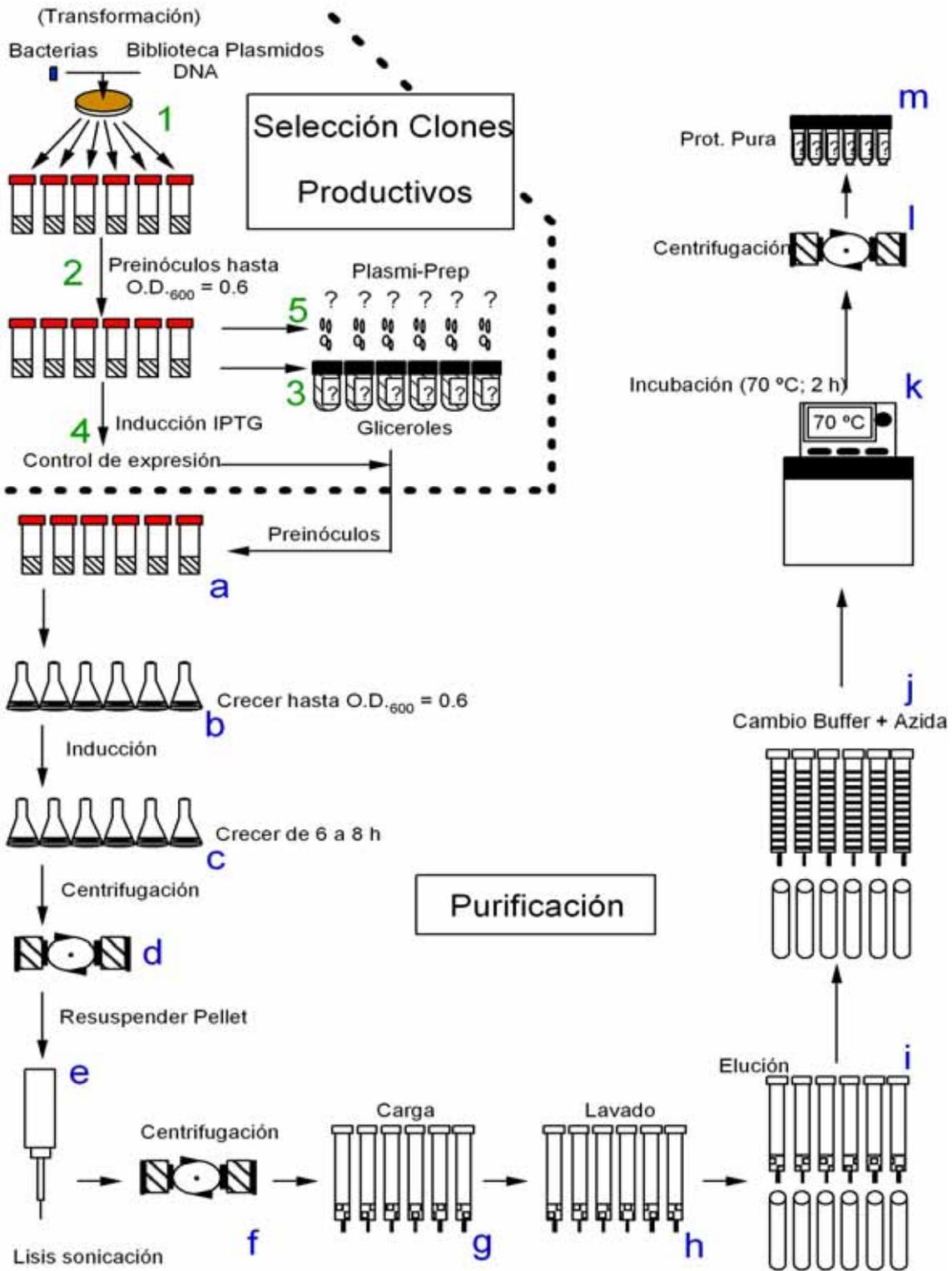
Con el fin de simplificar la purificación de numerosas formas mutantes de tiorredoxina de E.coli simultáneamente, se introdujo en la secuencia de nucleótidos que codifican la proteína, por el extremo N-

terminal, una cola de 6 histidinas (hisTag) seguida de un punto de corte para Trombina.

El grupo de histidinas que se ha introducido tiene capacidad para quelar cationes divalentes. Como consecuencia, si pasamos un lisado bacteriano por una resina con estos iones inmovilizados, nuestra proteína quedará adherida específicamente mientras que el resto del lisado será eluido de la columna.

La metodología que a continuación se describe se refiere a la purificación de proteínas de bibliotecas de mutantes. Una biblioteca de mutantes es un conjunto de formas mutantes de un gen, en este caso clonados cada uno de ellos en el mismo vector, que codifican para distintas formas mutantes de una misma proteína. Aunque la metodología para generar dicha biblioteca se encuentra descrita dentro del artículo presentado en el apéndice 1, cabe decir que las variantes se generaron a través de síntesis génica por métodos de PCR. Esencialmente consiste en diseñar los oligonucleótidos necesarios para que a la vez que se generan todas las combinaciones posibles de mutaciones que queramos estudiar, sea posible, por uso de una polimerasa, sintetizar el gen completo. Posteriormente el producto de esta reacción se purifica y se clona en nuestro vector.

# Purificación Biblioteca



**Figura PUR2:** Esquema de la purificación de librería de mutantes. Del 1 al 5, proceso de selección de clones productores de tiorredoxina. De la "a" a la "m", purificación de la proteína a partir de los clones.

En primer lugar transformaremos las bacterias BL21 DE3 con nuestra biblioteca de plásmidos (FIG PUR2-1). A diferencia del apartado anterior esta vez realizaremos la transformación tomando las cantidades indicadas por el protocolo comercial (100  $\mu$ l de células competentes para añadir posteriormente 1  $\mu$ l de plásmidos y 900  $\mu$ l de medio SOC) y distribuiremos el producto en cuantas placas de LB-agar podamos. El objeto es obtener el mayor número de colonias transformantes posible (hay que tener en cuenta que la biblioteca de plásmidos es un producto muy caro).

Dado que la efectividad en la generación de la biblioteca de mutantes ronda el 50%, antes de iniciar la purificación, como método para ahorrar trabajo y reactivos, se realizan controles de expresión de un elevado número de colonias. Para ello se pasan colonias a preinóculos enumerados y se dejan crecer hasta una absorbancia a 600nm de 0.6 (FIG PUR2-2). Es aconsejable trabajar con un número significativo de preinóculos, dado que muchos de ellos contendrán la misma forma mutante y otros de ellos no contendrán formas de interés (en el proceso de fabricación de la biblioteca aparecen muchos mutantes no buscados). Tras ello tomamos 750 $\mu$ l de cada uno de los preinóculos y los mezclamos con 300 $\mu$ l de glicerol (esterilizado) (FIG PUR2-3). Homogeneizamos la mezcla y la guardamos durante 12 horas en la cámara fría (el objeto es que la bacteria incorpore el glicerol a su medio interno y así ejerza de crioprotector para una posterior congelación). Después de las 12 horas recuperaremos las alícuotas y las guardaremos en un congelador a -80 °C. Es muy importante mantener la coherencia entre el etiquetado de los gliceroles y el de los preinóculos.

Al mismo tiempo que separamos la alícuota para el glicerol, es conveniente iniciar nuevos preinóculos a partir de los preexistentes. A partir de ellos haremos una "Plasmi-prep" (purificación de los vectores) para poder secuenciar aquellos clones que producen nuestra proteína de interés.

Para identificar qué clones (cada preinóculo contiene un clon) producen proteína de interés, induciremos con IPTG 0.4mM final cada uno de los preinóculos (FIG PUR2-4). De 6 a 12 horas después tomaremos una alícuota a la cual le haremos una electroforesis para identificar si aparece una banda correspondiente al peso molecular de Tiorredoxina (hay que tener en cuenta que la cantidad de proteína producida

"naturalmente" por la bacteria da una banda extremadamente débil en el gel). Por supuesto seguiremos prestando especial cuidado en mantener la misma nomenclatura, de modo que cada calle de la electroforesis se corresponda con un glicerol y un segundo preinóculo.

Tras determinar qué clones poseían vectores "productivos" es importante deshacerse de aquellos gliceroles correspondientes a clones no productivos. Posteriormente realizaremos la Plasmid-prep de los preinóculos correspondientes a los clones de interés (FIG PUR2-5). Para ello seguiremos el protocolo comercial. El producto se dividirá en dos alícuotas iguales, una de las cuales se desecará y se enviará para su secuenciación y la otra se guardará a  $-20\text{ }^{\circ}\text{C}$ . Es importante resaltar que este ADN no está metilado, por lo cual no puede usarse para ciclos de mutagénesis dirigida que estén basados en digestión de ADN parental o molde por DpnI. Si quisiéramos usarlo con tal fin el método más sencillo es realizar una transformación con él en células XL-1 Blue y obtener a partir de éstas el ADN (dado que estas bacterias sí que metilan el ADN).

Una vez llegados a este punto nos encontramos con una considerable cantidad de mutantes de secuencia conocida y un glicerol de cultivo para producir cada uno de ellos. Tomamos un pequeño volumen de cada glicerol, 20 $\mu\text{l}$  por ejemplo, e iniciamos un preinóculo para cada uno de los clones (FIG PUR2-a). El procedimiento para crecer las bacterias e inducir la expresión de la proteína de interés es el mismo que el descrito en el apartado anterior, con la salvedad de que usaremos sólo 250ml de medio LB para cada clón (FIG PUR2-b,c,d).

Una vez obtengamos el pellet bacteriano de cada uno de los clones, se resuspenden en "Binding buffer" (20mM Fosfato Sódico, 500mM NaCl y 20mM de Imidazol, a pH 7.4). Este buffer mantiene las histidinas con el nivel de protonación adecuado para que complejen correctamente el catión divalente, por tanto el pH es un parámetro especialmente importante. El NaCl evita interacciones electrostáticas inespecíficas que podrían darse entre distintas proteínas y los iones de la matriz. El imidazol simplemente aumenta la especificidad, dado que compete con las histidinas por quelar el catión divalente. Aumentando ligeramente la concentración de este podríamos mejorar la pureza de nuestra purificación, aunque a costa de perder una cierta cantidad de proteína de interés.

El volumen de bacterias resuspendidas que obtenemos es de unos 6ml por clon, por lo cual el uso de la French-Press es un tanto impracticable (a la vez que el elevado número de muestras lo desaconseja). Lisaremos las células a través de sonicación, manteniendo los tubos en hielo y dando 5 pulsos de 20 segundos (FIG PUR2-e). El lisado de cada uno de los clones se centrifuga en eppendorfs en una centrífuga de mesa (14.000r.p.m. durante 5 minutos) (FIG PUR2-f) y el sobrenadante se pasa a las columnas de afinidad His GraviTrap de GE Healthcare (obviamente una columna por muestra y manteniendo en todo momento el etiquetaje) (FIG PUR2-g). El volumen de estas columnas es de 1ml, lo cual permite purificar hasta un máximo de 40mg de proteína. Es recomendable que después de cada purificación se regenere la columna descargándola primero de iones para luego montarlos nuevamente (siguiendo el protocolo indicado por los fabricantes).

Después de pasar el sobrenadante a través de la columna pasamos 10ml de "Binding Buffer", que arrastrará todas las proteínas, ácidos nucleicos, azúcares,... que no son de nuestro interés (FIG PUR2-h). Idealmente, sólo quedará pegada a la columna nuestra proteína (en la realidad suele haber una pequeña cantidad de contaminantes, cantidad que irá en aumento cuanto peor haya salido la expresión de nuestra proteína). Para eluir la proteína pasamos 3ml de "Elution Buffer", que es idéntico al "Binding Buffer" pero con 0.5M de Imidazol en lugar de 20mM (FIG PUR2-i). El Imidazol competirá con nuestra proteína por interaccionar con los cationes divalentes, desplazándola y provocando su salida de la columna.

A continuación procederemos a cambiar el buffer a nuestra muestra de proteína. Para ello usaremos las columnas PD-10 de GE Healthcare, básicamente unas columnas de exclusión molecular con un tamaño de poro que permite el paso de moléculas de peso molecular inferior a 1000 Da (FIG PUR2-j). El resultado es que las proteínas avanzan en la columna a través del volumen muerto y sale con este (por ello hay que equilibrar previamente las columnas con el buffer con el que queremos que salga la proteína, en este caso Hepes 5mM pH 7), mientras que los componentes del buffer se retardan. Por ello, siguiendo los pasos indicados en el protocolo comercial, se puede cambiar el buffer de una muestra de proteína rápidamente y con muy poca dilución. Hay que resaltar que disminuyendo el volumen de muestra que introducimos y recogiendo un volumen de elución menor mejoramos la

eficacia del cambio de buffer. Finalmente añadimos azida al eluido hasta una concentración final de esta del 0.02% para mejorar la conservación de la disolución de proteína (se han observado resultados irreproducibles cuando las muestras no contenían azida).

La disolución de proteína que obteníamos no presentaba siempre unos niveles de pureza excepcionales. En el caso particular del artículo X se pudo mejorar la pureza incubando las muestras durante 2 horas a 70 °C (FIG PUR2-k), lo cual provoca la desnaturalización y agregación de la mayor parte de los contaminantes. Tras centrifugar a 14.000r.p.m. durante 5 minutos sedimentamos los agregados y nuestra proteína queda pura en el sobrenadante (FIG PUR2-l,m).

El número de purificaciones simultáneas que podemos realizar es considerablemente alto, pudiéndose llegar a purificar 24 mutantes por día. Sin embargo, el planteamiento del trabajo vendrá determinado por el tamaño de la biblioteca y el porcentaje de la misma que queramos estudiar. Así durante esta tesis doctoral se ha trabajado con varias bibliotecas de 256 mutantes. El número de mutantes distintos que queríamos estudiar era del 10%. Por tanto evaluamos 80 colonias para la expresión de formas mutantes de tiorredoxina. Aproximadamente se obtuvieron un 50% de positivos (este paso depende de la calidad de la biblioteca). Los 40 clones se enviaron a secuenciar y se acabó con 24 mutantes diferentes de tiorredoxina. Los otros 16 clones eran o bien mutantes repetidos o bien mutantes que presentaban mutaciones que no fueron elegidas por nosotros, sino que se generaron accidentalmente en el proceso de fabricación de la biblioteca.

Gracias al bajo coste de las columnas, a la eficacia que presenta nuestra proteína en su expresión y al uso de bibliotecas de mutantes, se han llegado a purificar 5-10 mg de proteína de hasta 30 variantes de Tiorredoxina en tan sólo dos días.

### **7.3-Purificación de $\gamma$ -14-3-3.**

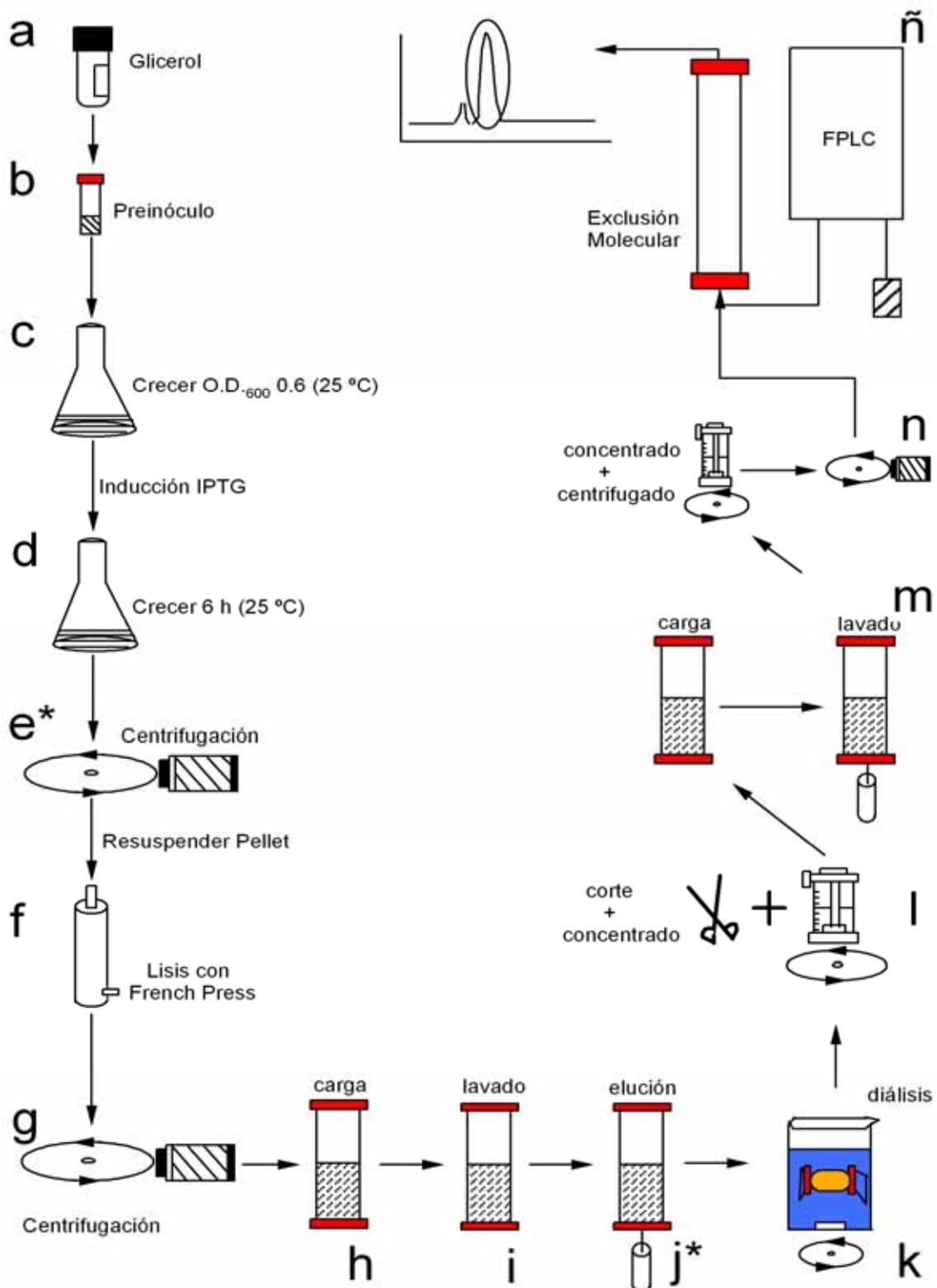
El trabajo con la proteína  $\gamma$ -14-3-3 surge de la colaboración con el laboratorio de la Dr. Aurora Martínez de la Universidad de Bergen (Noruega), quienes están especialmente interesados en conocer la estructura en solución de la proteína y desentrañar las interacciones de esta proteína implicada en señalización celular con diversas hidroxilasas. Esta proteína es hoy en día una de las más estudiadas en

Biología Molecular del cáncer, dado que parece encontrarse implicada en numerosas rutas de señalización y regulación celular. Se trata de un dímero en solución con 31900 Da por monómero. Cada uno de los monómeros posee dos zonas bien diferenciadas: Una región central cargada positivamente y que viene a ser el lugar de unión de fosfatos y a su alrededor dos alas cargadas negativamente con, posiblemente, una región de unión a iones de calcio.

Fue dicho laboratorio quien nos suministro un glicerol de bacterias BL21 DE3 transformadas con el vector y un protocolo para purificar la proteína y de este modo estudiar aquí diversos aspectos biofísicos de esta proteína. Dicho protocolo sufrió tan solo ligeras modificaciones.

En un primer momento se realizó un preinóculo a partir del glicerol con el objeto de purificar el vector con una Plasmid-prep. El plásmido purificado se guardó a -20 °C. Este constructo es un plásmido pGEX, que lleva un gen de resistencia a Ampicilina y el gen de la  $\gamma$ -14-3-3 unido por el extremo N-terminal con la proteína Glutathione S-Transferasa (GST). El nexo de unión entre ambas proteínas es un punto de corte para la Trombina, una proteasa específica de secuencia. Por otro lado, la GST es una proteína que presenta alta afinidad por el Glutathione. El Glutathione ligado a una matriz inerte permite que purifiquemos la fusión  $\gamma$ -14-3---GST a través de una cromatografía de afinidad.

# Purificación $\gamma$ -14-3-3



**Figura PUR3:** Esquema general de la purificación de  $\gamma$ -14-3-3. Se enumeran de la "a" a la "ñ" los distintos pasos. El asterisco denota aquellos procesos tras los cuales es posible congelar la muestra.

El esquema general de la purificación se encuentra representado en la figura PUR3. Se inicia a partir del glicerol de bacterias transformadas poniendo preinóculos de LB más Ampicilina a 0.1mg/ml final (FIG PUR3-a,b). Estos preinóculos se crecen en agitación a 37 °C durante la noche. De este modo a la mañana siguiente se inician los cultivos en matraces con 0.5 litros de LB mas Ampicilila a 0.1mg/ml final, que se crecen en agitación a 25 °C (FIG PUR3-c).

Cuando estos cultivos llegan a una absorbancia a 600nm de 0.6, se induce la expresión de la proteína con IPTG 0.4mM final (FIG PUR3-d). Después de 6 horas más de crecimiento, se centrifugan los medios con el objeto de separar las bacterias del medio de crecimiento a 7.000r.p.m. en el rotor A.6.9 durante 10 minutos a 4°C (FIG PUR3-e\*). El buffer para resuspender es "Binding Buffer" (140mM de NaCl, 2.7 mM de KCl, 10mM de Fosfato Disódico y 1.8mM de fosfato monopotásico, pH 7.3), que será el buffer con el cual tengamos equilibrada la columna donde se encuentra la resina Glutathione Sepharose 4B de GE Healthcare (que lleva unida glutation). Es recomendable para aumentar el rendimiento disolver una pastilla de inhibidores de proteasas en el volumen de buffer que vamos a usar para resuspender los pellets. Hay que tener en cuenta que a diferencia del apartado anterior, en este caso el ligando de afinidad no se ve afectado por la presencia de quelantes de iones.

El resuspendido es lisado, preferentemente con la French-Press para evitar un posible sobrecalentamiento en caso de usar sonicación (FIG PUR3-f). Se centrifuga nuevamente el producto en las mismas condiciones y tomamos el sobrenadante (FIG PUR3-g). Este, después de filtrarlo por 0.45µm, lo pasamos por la columna de afinidad. Dado que la cinética de unión entre GST y glutation es relativamente lenta, el flujo deberá ser lento (en este caso, basta dejar que fluya por gravedad) (FIG PUR3-h).

A continuación lavamos con "Binding Buffer" la columna (5-10 veces el volumen de resina que tengamos) (FIG PUR3-i). Esto hará que sólo se quede unido el producto de fusión  $\gamma$ -14-3-3--GST, dado que el resto de proteínas prácticamente no presentan interacciones específicas con la resina o el glutation.

La elución de la columna se produce por el paso de "Elution Buffer" (Tris-HCl 50mM con Glutation Reducido 15mM) (FIG PUR3-j\*). Se

recomienda eluir con 5 volúmenes de resina y detectar por medidas de absorbancia a 280nm en qué fracciones se encuentra la proteína, si bien por experiencia podemos indicar que prácticamente toda la proteína ha salido en los 3 primeros volúmenes (entre 0.5 y 2.5). Tan pronto como vayamos recogiendo las fracciones las pondremos en hielo para evitar que se estropee nuestra proteína.

A continuación realizaremos una diálisis con "Binding Buffer" por dos razones importantes (FIG PUR3-k). La primera de ellas es que la Trombina (una proteasa que se encargará de cortar la unión entre GST y  $\gamma$ -14-3-3, obtenida de Sigma-Aldrich) nos ha funcionado bien en este buffer. La segunda es que una vez cortadas se volverán a pasar por la columna de afinidad (en este caso para que la GST se retenga y nuestra proteína de interés fluya libremente). Se recomienda que la diálisis se haga a 4 °C y con dos cambios en 6 horas frente a un gran volumen de buffer, puesto que sólo necesitamos que la concentración de Glutathione baje de 1mM y con lo propuesto nos quedamos muy por debajo de este límite. Gracias a esto conseguiremos recortar el tiempo de purificación, un factor muy importante tratándose  $\gamma$ -14-3-3 de una proteína poco estable.

Se probaron distintas cantidades de Trombina y distintos tiempos de incubación. Se eligió realizar el corte a 4 °C con 5U de Trombina por miligramo de proteína de fusión y en agitación. Dado que en estas condiciones se obtiene un buen rendimiento de corte a las 2 horas y 30 minutos, durante la última hora de corte iniciaremos un proceso de concentración simultáneamente. Esto reducirá considerablemente el tiempo que dura la purificación, dado que un menor volumen tardará menos tiempo en entrar (y salir) de la columna de afinidad y a que en cualquier caso necesitamos concentrar la proteína para realizar el último paso de la purificación (FIG PUR3-l).

Una vez concentrado y cortado el producto de fusión tendremos una mezcla de GST,  $\gamma$ -14-3-3,  $\gamma$ -14-3-3--GST y Trombina en "Binding Buffer". Tras introducir esta disolución en la columna de afinidad, lavaremos con "Binding Buffer" hasta que la cantidad de proteína que salga esté a menos de 0.1mg/ml. Esta disolución se mantiene en todo momento en hielo. Tras este paso por columna tendremos  $\gamma$ -14-3-3 con Trombina (FIG PUR3-m).

La separación de ambos componentes se lleva a cabo a través un proceso de exclusión molecular. La columna que utilizaremos en esta ocasión no es como la que vimos en el apartado PUR1, su tamaño es considerablemente menor pero presenta una mejor resolución gracias a que la matriz que se utilizó para montarla (Superdex 75). El inconveniente es que para evitar alterar el empaquetamiento no se deben utilizar bombas peristálticas ni en su carga ni en el desarrollo de la cromatografía. Por tanto nos vemos obligados a utilizar el FPLC en todo momento y un "loop" de 5 ml para cargar la proteína. Por ello, volveremos a concentrar la proteína hasta obtener un volumen inferior a 5ml y antes de meterlo en la columna lo cetrifugaremos primero y filtraremos después para remover los agregados que frecuentemente se producen a lo largo del proceso (FIG PUR3-n).

La cromatografía de exclusión molecular que vamos a realizar la correremos con el siguiente buffer: 20mM Hepes, 200mM NaCl a pH7. Previamente a la cromatografía equilibraremos la columna en estas condiciones. El flujo óptimo de esta columna está entre 1 y 1.2 ml/min y se ha de prestar especial atención a que no aumente la presión en el interior de la columna (lo cual alteraría el empaquetamiento de la matriz). Recogemos alícuotas de 3ml y registramos la absorbancia a 280nm. En la FIG PUR3¿? se representa el cromatograma típico de este proceso. El primer pico del cromatograma es el producto de fusión que ni fue cortado y ni se unió a la segunda cromatografía de afinidad. Como resultado obtendremos nuestra proteína pura libre de Trombina (FIG PUR3-ñ). La cantidad final de proteína que consigamos viene determinado, principalmente, por la cantidad de resina de afinidad de la que dispongamos. Dado que tiene una capacidad de unión de 5 mg de proteína de fusión por ml de resina, y en nuestro laboratorio disponemos de 15 ml de resina, lo normal es acabar con, aproximadamente, 25 mg de proteína considerablemente pura (Los otros 5 mg suponen proteína que no ha sido totalmente cortada, proteína que ha agregado en el proceso y otras pequeñas pérdidas a lo largo del trabajo).

Con el fin de conservarla de un modo adecuado y optimizar su uso, es conveniente alicuotarla en tubos de criogeneización a las concentraciones a las cuales las necesitaremos con posterioridad. Una vez tengamos las alícuotas, las dejaremos en nitrógeno líquido durante 10-15 minutos y posteriormente las guardaremos a -80 °C.

## **8-BIBLIOGRAFÍA**

Aharoni, A., L. Gaidukov, et al. (2005). "The 'evolvability' of promiscuous protein functions." Nat Genet **37**(1): 73-6.

Beasty, A. M., M. R. Hurle, et al. (1986). "Effects of the phenylalanine-22----leucine, glutamic acid-49----methionine, glycine-234----aspartic acid, and glycine-234----lysine mutations on the folding and stability of the alpha subunit of tryptophan synthase from *Escherichia coli*." Biochemistry **25**(10): 2965-74.

Becktel, W. J. and J. A. Schellman (1987). "Protein stability curves." Biopolymers **26**(11): 1859-77.

Berezovsky, I. N. and E. I. Shakhnovich (2005). "Physics and evolution of thermophilic adaptation." Proc Natl Acad Sci U S A **102**(36): 12742-7.

Bloom, J. D., S. T. Labthavikul, et al. (2006). "Protein stability promotes evolvability." Proc Natl Acad Sci U S A **103**(15): 5869-74.

Bloom, J. D., Z. Lu, et al. (2007). "Evolution favors protein mutational robustness in sufficiently large populations." BMC Biol **5**: 29.

Bloom, J. D., A. Raval, et al. (2007). "Thermodynamics of neutral protein evolution." Genetics **175**(1): 255-66.

Chen, B. L., W. A. Baase, et al. (1989). "Low-temperature unfolding of a mutant of phage T4 lysozyme. 2. Kinetic investigations." Biochemistry **28**(2): 691-9.

Chen, B. L. and J. A. Schellman (1989). "Low-temperature unfolding of a mutant of phage T4 lysozyme. 1. Equilibrium studies." Biochemistry **28**(2): 685-91.

Ciccarelli, F. D., T. Doerks, et al. (2006). "Toward automatic reconstruction of a highly resolved tree of life." Science **311**(5765): 1283-7.

Conejero-Lara, F., P. L. Mateo, et al. (1991). "Effect of Zn<sup>2+</sup> on the thermal denaturation of carboxypeptidase B." Biochemistry **30**(8): 2067-72.

Cooper, A. (1976). "Thermodynamic fluctuations in protein molecules." Proc Natl Acad Sci U S A **73**(8): 2740-1.

DePristo, M. A., D. M. Weinreich, et al. (2005). "Missense meanderings in sequence space: a biophysical view of protein evolution." Nat Rev Genet **6**(9): 678-87.

Dill, K. A. (1990). "Dominant forces in protein folding." Biochemistry **29**(31): 7133-55.

Dokholyan, N. V., B. Shakhnovich, et al. (2002). "Expanding protein universe and its origin from the biological Big Bang." Proc Natl Acad Sci U S A **99**(22): 14132-6.

Drummond, D. A., J. D. Bloom, et al. (2005). "Why highly expressed proteins evolve slowly." Proc Natl Acad Sci U S A **102**(40): 14338-43.

Eaton, W. A., V. Munoz, et al. (1997). "Submillisecond kinetics of protein folding." Curr Opin Struct Biol **7**(1): 10-4.

Ellman, G. L. (1959). "Tissue sulfhydryl groups." Arch Biochem Biophys **82**(1): 70-7.

Frederick, K. K., M. S. Marlow, et al. (2007). "Conformational entropy in molecular recognition by proteins." Nature **448**(7151): 325-9.

Freire, E. (1995). "Differential scanning calorimetry." Protein stability and folding. Theory and practice (Shirley, B. A. ed.) Humana Press, Totowa, NJ: 191-218.

Freire, E. (1995). "Differential scanning calorimetry." Methods Mol Biol **40**: 191-218.

Freire, E. and R. Biltonen (1978). "Statistical mechanical deconvolution of thermal transitions in macromolecules. 1. Theory and application to homogeneous systems." Biopolymers **17**(2): 463-479.

Garcia-Mira, M. M., M. Sadqi, et al. (2002). "Experimental identification of downhill protein folding." Science **298**(5601): 2191-5.

García-Segura J.M., J. G. G., A. Martinez del Pozo, F. Montero, M. Oñaderra y F. Vivanco (2002). Técnicas instrumentales de análisis en Bioquímica, Ed. Síntesis.

Gaucher, E. A., S. Govindarajan, et al. (2008). "Palaeotemperature trend for Precambrian life inferred from resurrected proteins." Nature **451**(7179): 704-7.

Godoy-Ruiz, R., F. Ariza, et al. (2006). "Natural selection for kinetic stability is a likely origin of correlations between mutational effects on protein energetics and frequencies of amino acid occurrences in sequence alignments." J Mol Biol **362**(5): 966-78.

Guo, H. H., J. Choe, et al. (2004). "Protein tolerance to random amino acid change." Proc Natl Acad Sci U S A **101**(25): 9205-10.

Gupta, R., P. Rathi, et al. (2003). "Lipase assays for conventional and molecular screening: an overview." Biotechnol Appl Biochem **37**(Pt 1): 63-71.

Halskau, O., Jr., R. Perez-Jimenez, et al. (2008). "Large-scale modulation of thermodynamic protein folding barriers linked to electrostatics." Proc Natl Acad Sci U S A **105**(25): 8625-30.

Hammarstrom, P., R. L. Wiseman, et al. (2003). "Prevention of transthyretin amyloid disease by changing protein misfolding energetics." Science **299**(5607): 713-6.

Hilser, V. J., J. Gomez, et al. (1996). "The enthalpy change in protein folding and binding: refinement of parameters for structure-based calculations." Proteins **26**(2): 123-33.

Holmgren, A. (1979). "Thioredoxin catalyzes the reduction of insulin disulfides by dithiothreitol and dihydrolipoamide." J Biol Chem **254**(19): 9627-32.

Hurle, M. R. and C. R. Matthews (1987). "Proline isomerization and the slow folding reactions of the alpha subunit of tryptophan synthase from *Escherichia coli*." Biochim Biophys Acta **913**(2): 179-84.

Hurle, M. R., G. A. Michelotti, et al. (1987). "Characterization of a slow folding reaction for the alpha subunit of tryptophan synthase." Proteins **2**(1): 54-63.

Ibarra-Molero, B. and J. M. Sanchez-Ruiz (1996). "A model-independent, nonlinear extrapolation procedure for the characterization of protein folding energetics from solvent-denaturation data." Biochemistry **35**(47): 14689-702.

Ibarra-Molero, B. and J. M. Sanchez-Ruiz (2002). "Genetic Algorithm to Design Stabilizing Surface-Charge Distributions in Proteins." The Journal of Physical Chemistry B **106**(26): 6609-6613.

Ibarra-Molero, B. and J. M. Sanchez-Ruiz (2006). "Emerging Techniques in Biophysics, ed. J. L. Arrondo and A. Alonso, Elsevier, Amsterdam." 27-48.

Jain, R. K. and R. Ranganathan (2004). "Local complexity of amino acid interactions in a protein core." Proc Natl Acad Sci U S A **101**(1): 111-6.

Jaswal, S. S., J. L. Sohl, et al. (2002). "Energetic landscape of alpha-lytic protease optimizes longevity through kinetic stability." Nature **415**(6869): 343-6.

Jaswal, S. S., S. M. Truhlar, et al. (2005). "Comprehensive analysis of protein folding activation thermodynamics reveals a universal behavior violated by kinetically stable proteases." J Mol Biol **347**(2): 355-66.

Joh, N. H., A. Min, et al. (2008). "Modest stabilization by most hydrogen-bonded side-chain interactions in membrane proteins." Nature **453**(7199): 1266-70.

Kanehisa, M. I. and A. Ikegami (1977). "Structural changes and fluctuations of proteins. II. Analysis of the denaturation of globular proteins." Biophys Chem **6**(2): 131-49.

Kaya, H. and H. S. Chan (2000). "Polymer principles of protein calorimetric two-state cooperativity." Proteins **40**(4): 637-61.

Khersonsky, O., C. Roodveldt, et al. (2006). "Enzyme promiscuity: evolutionary and mechanistic aspects." Curr Opin Chem Biol **10**(5): 498-508.

Kiefhaber, T. and F. X. Schmid (1992). "Kinetic coupling between protein folding and prolyl isomerization. II. Folding of ribonuclease A and ribonuclease T1." J Mol Biol **224**(1): 231-40.

Kimura, M. and T. Ota (1974). "On some principles governing molecular evolution." Proc Natl Acad Sci U S A **71**(7): 2848-52.

Knott, M. and H. S. Chan (2006). "Criteria for downhill protein folding: calorimetry, chevron plot, kinetic relaxation, and single-molecule radius of gyration in chain models with subdued degrees of cooperativity." Proteins **65**(2): 373-91.

Krantz, B. A., R. S. Dothager, et al. (2004). "Discerning the structure and energy of multiple transition states in protein folding using psi-analysis." J Mol Biol **337**(2): 463-75.

- Laurent, T. C., E. C. Moore, et al. (1964). "Enzymatic Synthesis of Deoxyribonucleotides. Iv. Isolation and Characterization of Thioredoxin, the Hydrogen Donor from Escherichia Coli B." J Biol Chem **239**: 3436-44.
- Liang, J. and K. A. Dill (2001). "Are proteins well-packed?" Biophys J **81**(2): 751-66.
- Liu, Z. and H. S. Chan (2005). "Solvation and desolvation effects in protein folding: native flexibility, kinetic cooperativity and enthalpic barriers under isostability conditions." Phys Biol **2**(4): S75-85.
- Lockless, S. W. and R. Ranganathan (1999). "Evolutionarily conserved pathways of energetic connectivity in protein families." Science **286**(5438): 295-9.
- Loladze, V. V., B. Ibarra-Molero, et al. (1999). "Engineering a thermostable protein via optimization of charge-charge interactions on the protein surface." Biochemistry **38**(50): 16419-23.
- Makhatadze, G. I. (2005). "Thermodynamics Of alpha-Helix Formation." Adv Protein Chem **72**: 199-226.
- Makhatadze, G. I., V. V. Loladze, et al. (2003). "Contribution of surface salt bridges to protein stability: guidelines for protein engineering." J Mol Biol **327**(5): 1135-48.
- Matouschek, A., J. T. Kellis, Jr., et al. (1989). "Mapping the transition state and pathway of protein folding by protein engineering." Nature **340**(6229): 122-6.
- Miller, B. G. and R. T. Raines (2004). "Identifying latent enzyme activities: substrate ambiguity within modern bacterial sugar kinases." Biochemistry **43**(21): 6387-92.
- Miyazaki, K., P. L. Wintrode, et al. (2000). "Directed evolution study of temperature adaptation in a psychrophilic enzyme." J Mol Biol **297**(4): 1015-26.
- Mucke, M. and F. X. Schmid (1994). "A kinetic method to evaluate the two-state character of solvent-induced protein denaturation." Biochemistry **33**(43): 12930-5.
- Munoz, V. (2007). "Conformational dynamics and ensembles in protein folding." Annu Rev Biophys Biomol Struct **36**: 395-412.
- Munoz, V. and J. M. Sanchez-Ruiz (2004). "Exploring protein-folding ensembles: a variable-barrier model for the analysis of equilibrium unfolding experiments." Proc Natl Acad Sci U S A **101**(51): 17646-51.
- Munoz, V., P. A. Thompson, et al. (1997). "Folding dynamics and mechanism of beta-hairpin formation." Nature **390**(6656): 196-9.
- Naganathan, A. N., J. M. Sanchez-Ruiz, et al. (2005). "Direct measurement of barrier heights in protein folding." J Am Chem Soc **127**(51): 17970-1.
- Nölting, B. (2006). Physical interactions that determine the properties of proteins, Ed. Springer.

Onuchic, J. N., Z. Luthey-Schulten, et al. (1997). "Theory of protein folding: the energy landscape perspective." Annu Rev Phys Chem **48**: 545-600.

Ota, T. and M. Kimura (1971). "Amino acid composition of proteins as a product of molecular evolution." Science **174**(5): 150-3.

Perez-Jimenez, R., R. Godoy-Ruiz, et al. (2006). "A simple tool to explore the distance distribution of correlated mutations in proteins." Biophys Chem **119**(3): 240-6.

Perry, K. M., J. J. Onuffer, et al. (1989). "Long-range electrostatic interactions can influence the folding, stability, and cooperativity of dihydrofolate reductase." Biochemistry **28**(19): 7961-8.

Pey, A. L., D. Rodriguez-Larrea, et al. (2008). "Engineering proteins with tunable thermodynamic and kinetic stabilities." Proteins **71**(1): 165-74.

Plotkin, S. S. and J. N. Onuchic (2002). "Understanding protein folding with energy landscape theory. Part II: Quantitative aspects." Q Rev Biophys **35**(3): 205-86.

Plotnikov, V. V., J. M. Brandts, et al. (1997). "A new ultrasensitive scanning calorimeter." Anal Biochem **250**(2): 237-44.

Powis, G. and W. R. Montfort (2001). "Properties and biological activities of thioredoxins." Annu Rev Biophys Biomol Struct **30**: 421-55.

Privalov, P. L., V. Griko Yu, et al. (1986). "Cold denaturation of myoglobin." J Mol Biol **190**(3): 487-98.

Qian, B., S. Raman, et al. (2007). "High-resolution structure prediction and the crystallographic phase problem." Nature **450**(7167): 259-64.

Robertson, A. D. and K. P. Murphy (1997). "Protein Structure and the Energetics of Protein Stability." Chem Rev **97**(5): 1251-1268.

Rodriguez-Larrea, D., B. Ibarra-Molero, et al. (2008). "Beyond Lumry-Eyring: an unexpected pattern of operational reversibility/irreversibility in protein denaturation." Proteins **70**(1): 19-24.

Rodriguez-Larrea, D., B. Ibarra-Molero, et al. (2006). "Energetic and structural consequences of desolvation/solvation barriers to protein folding/unfolding assessed from experimental unfolding rates." Biophys J **91**(5): L48-50.

Rodriguez-Larrea, D., S. Minning, et al. (2006). "Role of solvation barriers in protein kinetic stability." J Mol Biol **360**(3): 715-24.

Rothlisberger, D., O. Khersonsky, et al. (2008). "Kemp elimination catalysts by computational enzyme design." Nature **453**(7192): 190-5.

Sanchez-Ruiz, J. M. (1992). "Theoretical-analysis of Lumry-Eyring models in differential scanning calorimetry." Biophys J **61**(4): 921-935.

Sanchez-Ruiz, J. M. (1995). "Differential scanning calorimetry of proteins." Subcell Biochem **24**: 133-76.

Sanchez-Ruiz, J. M., J. L. Lopez-Lacomba, et al. (1988). "Differential scanning calorimetry of the irreversible thermal denaturation of thermolysin." Biochemistry **27**(5): 1648-52.

Schellman, J. A. (1987). "The thermodynamic stability of proteins." Annu Rev Biophys Biophys Chem **16**: 115-37.

Schmid, F. X. (1992). Protein Folding.

Shortle, D. (2003). "Propensities, probabilities, and the Boltzmann hypothesis." Protein Sci **12**(6): 1298-302.

Slaby, I. and A. Holmgren (1975). "Reconstitution of Escherichia coli thioredoxin from complementing peptide fragments obtained by cleavage at methionine-37 or arginine-73." J Biol Chem **250**(4): 1340-7.

Sohl, J. L., S. S. Jaswal, et al. (1998). "Unfolded conformations of alpha-lytic protease are more stable than its native state." Nature **395**(6704): 817-9.

Steipe, B., B. Schiller, et al. (1994). "Sequence statistics reliably predict stabilizing mutations in a protein domain." J Mol Biol **240**(3): 188-92.

Street, T. O., D. W. Bolen, et al. (2006). "A molecular mechanism for osmolyte-induced protein stability." Proc Natl Acad Sci U S A **103**(38): 13997-4002.

Thorolfsson, M., B. Ibarra-Molero, et al. (2002). "L-phenylalanine binding and domain organization in human phenylalanine hydroxylase: a differential scanning calorimetry study." Biochemistry **41**(24): 7573-85.

Uversky, V. N. (2002). "Natively unfolded proteins: a point where biology waits for physics." Protein Sci **11**(4): 739-56.

Uversky, V. N., J. R. Gillespie, et al. (2000). "Why are "natively unfolded" proteins unstructured under physiologic conditions?" Proteins **41**(3): 415-27.

Vitetta ES, G. V. (2006). "Immunology. Considering therapeutic antibodies." Science **313**(5785): 308-9.

Wang, X., G. Minasov, et al. (2002). "Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs." J Mol Biol **320**(1): 85-95.

Weinreich, D. M., N. F. Delaney, et al. (2006). "Darwinian evolution can follow only very few mutational paths to fitter proteins." Science **312**(5770): 111-4.

Xia, Y. and M. Levitt (2004). "Simulating protein evolution in sequence and structure space." Curr Opin Struct Biol **14**(2): 202-7.

Yang, W. Y. and M. Gruebele (2003). "Folding at the speed limit." Nature **423**(6936): 193-7.

Zhou, Y., C. K. Hall, et al. (1999). "The calorimetric criterion for a two-state process revisited." Protein Sci **8**(5): 1064-74.