

UNIVERSIDAD DE GRANADA

FACULTAD DE CIENCIAS



**MODULACIÓN MUTACIONAL DE  
PROTEÍNAS: INGENIERÍA DE  
PROTEÍNAS MORFOGENÉTICAS  
DE HUESO Y ENZIMAS  
ANCESTRALES RESUCITADAS**

María Luisa Romero Romero

TESIS DOCTORAL  
PROGRAMA DE DOCTORADO EN QUÍMICA  
GRANADA 2012

Editor: Editorial de la Universidad de Granada  
Autor: María Luisa Romero Romero  
D.L.: GR 596-2013  
ISBN: 978-84-9028-407-0







UNIVERSIDAD DE GRANADA

FACULTAD DE CIENCIAS



DEPARTAMENTO DE QUÍMICA FÍSICA

MODULACIÓN MUTACIONAL DE  
PROTEÍNAS: INGENIERÍA DE PROTEÍNAS  
MORFOGENÉTICAS DE HUESO Y ENZIMAS  
ANCESTRALES RESUCITADAS

María Luisa Romero Romero

TESIS DOCTORAL  
PROGRAMA DE DOCTORADO EN QUÍMICA  
GRANADA  
2012



UNIVERSIDAD DE GRANADA  
FACULTAD DE CIENCIAS  
DEPARTAMENTO DE QUÍMICA FÍSICA

Memoria presentada para optar al Grado de Doctor por la Universidad de  
Granada

Fdo.: María Luisa Romero Romero  
Licenciada en Ciencias Químicas por la Universidad de Granada

Granada Septiembre 2012

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



Los directores de la Tesis Doctoral, José Manuel Sánchez Ruiz y Beatriz Ibarra Molero y la doctoranda, María Luisa Romero Romero, garantizamos al firmar la Tesis Doctoral, que el trabajo ha sido realizado por la doctoranda bajo la tutela de los directores y hasta donde nuestro conocimiento alcanza, en la realización del trabajo se han respetado los derechos de otros autores al ser citados cuando se han utilizado sus resultados o publicaciones.

Septiembre 2012 Granada

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

Doctoranda:

Fdo.: María Luisa Romero Romero



Yo, José Manuel Sánchez Ruiz, con DNI nº 24141354-W. Siendo coautor de la publicación “*Highly Anomalous Energetics of Protein Cold Denaturation Linked to Folding-Unfolding Kinetics*” publicada en la revista PLoS ONE con fecha 29 de julio de 2011. Declaro que:

- María Luisa Romero Romero, la también coautora de esta publicación, ha concebido, diseñado y realizado los experimentos y ha analizado los datos que se presentan en esta publicación.
- No he presentado y renuncio a presentar dicha publicación en otra Tesis Doctoral.

Septiembre 2012 Granada

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



Yo, Beatriz Ibarra Molero, con DNI nº 24256088-N. Siendo coautora de la publicación “*Highly Anomalous Energetics of Protein Cold Denaturation Linked to Folding-Unfolding Kinetics*” publicada en la revista PLoS ONE con fecha 29 de julio de 2011. Declaro que:

- María Luisa Romero Romero, la también coautora de esta publicación, ha concebido, diseñado y realizado los experimentos y ha analizado los datos que se presentan en esta publicación.
- No he presentado y renuncio a presentar dicha publicación en otra Tesis Doctoral.

Septiembre 2012 Granada

Fdo.: Beatriz Ibarra Molero

Profesora Titular

Departamento de Química-Física

Facultad de Ciencias

Universidad de Granada



Yo, Álvaro Inglés Prieto, con DNI nº 28954576-Z. Siendo coautor de la publicación publicación “*Highly Anomalous Energetics of Protein Cold Denaturation Linked to Folding-Unfolding Kinetics*” publicada en la revista PLoS ONE con fecha 29 de julio de 2011. Declaro que:

- María Luisa Romero Romero, la también coautora de esta publicación, ha concebido, diseñado y realizado los experimentos y ha analizado los datos que se presentan en esta publicación.
- No he presentado y renuncio a presentar dicha publicación en otra Tesis Doctoral.

Septiembre 2012 Granada

Fdo.: Álvaro Inglés Prieto

Contratado Predoctoral FPU

Departamento de Química-Física

Facultad de Ciencias

Universidad de Granada



*A mis padres*

*A mi hermano*

*A Héctor*



Agradecerles a mis directores, José Manuel y Beatriz, por darme la oportunidad de realizar la Tesis Doctoral bajo su tutela, también por transmitirme su entusiasmo por este “mundillo”. Sois grandes ejemplos profesionales y es una suerte aprender de vosotros.

A José Luis, por acogerme con cariño en su laboratorio de Madrid, por mostrarme el lado *in vivísimo* de las proteínas y por su constante buen humor al que sólo puedes responder con una sonrisa.

A Héctor, mi compañero de todo y mi cómplice de todo. Contigo, vivo a cada momento la espuma de los días.

A toda la gente del departamento, en especial a Mari Carmen, Asun y María del Mar, por los buenísimos momentos, las risas compartidas y por las que nos quedan por llegar. También a Valeria, por brindarme su ayuda con cariño y por los útiles consejos que me ha dado. Y a mis compañeros del laboratorio: Citlali, Valeria, Andrea, Gema, Inma, Rocío, Noel, Álvaro, Ángel, Diego, e Israel.

A mis compañeros de Madrid que me habéis enseñado, en mi estancia en la capital, a vivir con prisa y con tiempo para todo, por los buenos ratos disfrutados, dentro y fuera del laboratorio. A Mariona y Ana: sois estupendas.

A mis amigos en Granada, en Helsinki, en Madrid y en El Puerto, con vosotros todas las ciudades son bonitas y el tiempo pasa rápido.

A toda mi familia. A mi padre que me ha enseñado que si el hoy es malo, el mañana es mío. A mi madre por transmitirme la paciencia, el tesón y la ilusión con la que emprender cualquier tarea. A mi hermano, que me ha sufrido muchas “charlillas” y siempre me ha hecho reír, incluso tras los peores agobios. Y en general por todo, porque todo os lo debo.



# Índice

<b>1. Introducción.....</b>	<b>1</b>
1.1. Bibliografía.....	15
<b>2. Metodología.....</b>	<b>19</b>
2.1. Obtención de ADN de proteínas mutantes.....	21
2.2. Clonación de variantes ancestrales de tiorredoxina en el vector pET30a(+)......	24
2.3. Construcción de bibliotecas combinatoriales de variantes de tiorredoxina.....	30
2.4. Obtención de proteínas.....	31
2.4.1. Transformación de células supercompetentes BL21-DE3.....	32
2.4.2. Pruebas de expresión de proteína.....	32
2.4.3. Expresión a gran escala.....	33
2.4.4. Purificación de proteína.....	33
2.5. Preparación de las muestras.....	37
2.5.1. Preparación de tampones.....	37
2.5.2. Preparación de diálisis.....	38
2.5.3. Medida de concentración de proteína.....	38
2.5.4. Preparación de disoluciones de guanidina.....	39
2.6. Calorimetría diferencial de barrido.....	39
2.6.1. Introducción.....	39
2.6.2. Instrumentación.....	40
2.6.3. Experimento calorimétrico.....	42
2.6.4. Análisis de las trazas de DSC.....	45
2.7. Dicroísmo circular.....	59
2.7.1. Introducción .....	59

2.7.2	Dicroísmo circular de proteínas.....	63
2.7.3	Determinación del grado de estructura secundaria y terciaria de la tiorredoxina de <i>E. coli</i> seguido por dicroísmo circular.....	66
2.8.	Fluorescencia.....	67
2.8.1.	Introducción .....	67
2.8.2.	Fluorescencia de proteínas.....	69
2.8.3.	Cinéticas de desplegamiento de tiorredoxina silvestre y variantes de tiorredoxina seguidas por fluorescencia	73
2.9.	Estudio proteolítico de BMP-2 silvestre y variantes de BMP-2 por incubación con termolisina.....	75
2.10.	Ensayo de actividad reductasa de tiorredoxina silvestre y variantes de tiorredoxina.....	77
2.11.	Ensayos celulares de BMP-2 silvestre y variantes de BMP-2....	80
2.11.1	Mantenimiento de la línea celular mioblástica C2C12	80
2.11.2	Ensayo de diferenciación celular seguido por medidas de actividad fosfatasa alcalina.....	81
2.11.3	Cuantificación de la proliferación celular mediante la determinación del contenido total de proteínas por Bradford.....	84
2.12.	Bibliografía.....	87
<b>3.</b>	<b>Sección experimental: objetivos y resultados.....</b>	<b>95</b>
3.1.	Proteína Morfogenética de Hueso 2.....	97
3.1.1.	Análisis bioinformático del alineamiento de secuencias: Método de consenso.....	101
3.1.2.	Estudio de la capacidad de diferenciación ósea de BMP-2 silvestre y variantes de BMP-2.....	112
3.1.3.	Caracterización biofísica de BMP-2 silvestre y variantes de BMP-2.....	118
3.1.4	Bibliografía.....	125
3.2.	Estudio de tiorredoxinas precámbricas resucitadas.....	129

3.2.1.	Estudio mutacional dirigido a analizar los factores energéticos que determinan la estabilidad termodinámica y cinética de la tiorredoxina de LPBCA.....	137
3.2.2.	Highly Anomalous Energetics of Protein Cold Denaturation Linked to Folding-Unfolding Kinetics...	159
3.2.3.	Rediseño del centro activo de la tiorredoxina de LPBCA mediante la simulación mutacional de una carga enterrada.....	171
3.2.4.	Bibliografía.....	191
<b>4.</b>	<b>Resumen y conclusiones.....</b>	<b>197</b>



## Índice de figuras

Figura 1.1.....	6	Figura 2.25.....	83
Figura 1.2.....	9	Figura 2.26.....	84
Figura 1.3.....	10	Figura 2.27.....	85
Figura 1.4.....	12	Figura 3.1.1.....	97
Figura 2.1.....	25	Figura 3.1.2.....	98
Figura 2.2.....	28	Figura 3.1.3.....	99
Figura 2.3.....	29	Figura 3.1.4.....	104
Figura 2.4.....	31	Figura 3.1.5.....	105
Figura 2.5.....	33	Figura 3.1.6.....	107
Figura 2.6.....	37	Figura 3.1.7.....	109
Figura 2.7.....	42	Figura 3.1.8.....	110
Figura 2.8.....	44	Figura 3.1.9.....	112
Figura 2.9.....	53	Figura 3.1.10.....	114
Figura 2.10.....	59	Figura 3.1.11.....	114
Figura 2.11.....	60	Figura 3.1.12.....	115
Figura 2.12.....	61	Figura 3.1.13.....	116
Figura 2.13.....	65	Figura 3.1.14.....	116
Figura 2.14.....	68	Figura 3.1.15.....	117
Figura 2.15.....	70	Figura 3.1.16.....	119
Figura 2.16.....	71	Figura 3.1.17.....	121
Figura 2.17.....	72	Figura 3.1.18.....	122
Figura 2.18.....	74	Figura 3.1.19.....	123
Figura 2.19.....	76	Figura 3.2.1.....	130
Figura 2.20.....	77	Figura 3.2.2.....	131
Figura 2.21.....	77	Figura 3.2.3.....	132
Figura 2.22.....	78	Figura 3.2.4.....	133
Figura 2.23.....	79	Figura 3.2.5.....	138
Figura 2.24.....	81	Figura 3.2.6.....	139

Figura 3.2.7.....	141	Figura 3.2.19.....	159
Figura 3.2.8.....	142	Figura 3.2.20.....	173
Figura 3.2.9.....	142	Figura 3.2.21.....	174
Figura 3.2.10.....	144	Figura 3.2.22.....	177
Figura 3.2.11.....	146	Figura 3.2.23.....	178
Figura 3.2.12.....	147	Figura 3.2.24.....	179
Figura 3.2.13.....	148	Figura 3.2.25.....	180
Figura 3.2.14.....	149	Figura 3.2.26.....	181
Figura 3.2.15.....	150	Figura 3.2.27.....	183
Figura 3.2.16.....	153	Figura 3.2.28.....	184
Figura 3.2.17.....	155	Figura 3.2.29.....	188
Figura 3.2.18.....	156	Figura 3.2.30.....	189

## **Índice de tablas**

Tabla 3.1.1.....	111	Tabla 3.2.3.....	182
Tabla 3.2.1.....	145	Tabla 3.2.4.....	186
Tabla 3.2.2.....	176	Tabla 3.2.5.....	189

## Abreviaturas y símbolos

A	Actividad
Abs	Absorbancia
ADN	Ácido desoxirribonucleico
AECA	Ancestro común de arquea y eucariota
ARN	Ácido ribonucleico
ATCC	<i>American type culture collection</i>
b	Paso de luz
BLAST	<i>Basic local alignment search tool</i>
BMP	Proteína morfogenética de hueso
c	Concentración
CD	Dicroísmo circular
CDMP	Proteína morfogenética derivada de cartílago
cel	Células
Células CHO	Células de ovario de hámster chino
$C_p^D$	Capacidad calorífica del estado desnaturalizado
$C_p^{ex}$	Capacidad calorífica de exceso
$C_p^N$	Capacidad calorífica del estado nativo
D	Estado desplegado
$\Delta C_{app}$	Cambio de capacidad calorífica aparente
$\Delta_N^D C_p$	Cambio de capacidad calorífica de desnaturalización a presión constante
$\Delta_N^D G$	Cambio de energía libre de Gibbs de desnaturalización
$\Delta_N^D H_m$	Cambio de entalpía de desnaturalización a la $T_m$
$\Delta H^{vH}$	Cambio de entalpía de desnaturalización de van't Hoff
$\Delta_N^D S$	Cambio de entropía de desnaturalización
$\Delta_N^D S_m$	Cambio de entropía de desnaturalización a la $T_m$

DMEM	<i>Dulbecco's modified eagle medium</i>
dNTP's	Mezcla de desoxinucleótidos trifosfatos
DSC	Calorimetría diferencial de barrido
DTT	Ditiotreitol
$\epsilon$	Coefficiente de extinción molar
<i>E. coli</i>	<i>Escherichia coli</i>
$\epsilon_d$	Coefficiente de extinción molar para la luz polarizada a la derecha
EDTA	Ácido etilendiaminotetraacético
$\epsilon_i$	Coefficiente de extinción molar para la luz polarizada a la izquierda
F	Estado final
FA	Fosfatasa alcalina
gd	Guanidina
GDF	Factor de crecimiento y diferenciación
h	Constante de Plank
I	Fuerza iónica, intensidad de fluorescencia
$I_\infty$	Intensidad de fluorescencia en el equilibrio
IgG	Inmunoglobulina G
IPTG	Isopropil- $\beta$ -D-1-tiogalactopiranosido
K	Constante de equilibrio
$k_{ap}$	Constante de velocidad de primer orden para el proceso de desnaturalización irreversible
$k_f, k_{UNF}$	Constante de primer orden de desplegamiento
$k_r$	Constante de primer orden de replegamiento
LACA	Último ancestro común de arquea
LAFCA	Último ancestro común de los animales y hongos

LB	<i>Luria bertani</i>
LBCA	Último ancestro común de las bacterias
LECA	Último ancestro común de eucariotas
LPBCA	Último ancestro común de los grupos de deinnococus y thermus
LPGCA	Último ancestro común de las $\gamma$ -proteobacterias
$M_r$	Masa molecular
N	Estado nativo
$v$	Velocidad de barrido
$N_A$	Número de aminoácidos
NADPH	Nicotinamida adenina dinucleótido fosfato
NIPALS	<i>Nonlinear iterative partial least squares</i>
OP	Proteína Osteogénica
pb	Pares de bases
PCR	Reacción de la polimerasa en cadena
PLS-R	Regresión por mínimos cuadrados parciales
$P_m$	Peso molecular
R	Constante de los gases
RMN	Resonancia magnética nuclear
SDS	Dodecilsulfato sódico
SDS-PAGE	Electroforesis en gel de poliacrilamida con SDS
SOC	<i>Super optimal broth</i> con represión catabólica
T	Temperatura
$\tau$	Tiempo de relajación
$\tau_{1/2}$	Tiempo de vida media
TB	<i>Terrific broth</i>
TE	Tris EDTA
TGF- $\beta$	Factor de crecimiento transformante tipo $\beta$

$T_H$	Temperatura de inversión de la entalpía
$T_m$	Temperatura de desnaturalización por calor, temperatura a la cual la mitad de la hebra de ADN se encuentra como doble hélice
$T_m^*$	Temperatura de desnaturalización por frío
$T_{rx}$	Tiorredoxina
$T_S$	Temperatura de inversión de la entropía
$[\theta]$	Elipticidad molar
$[\theta]_{MRW}$	Elipticidad molar normalizada
UV	Ultravioleta
$x_D$	Fracción de proteína en el estado desnaturalizado
$x_N$	Fracción de proteína en el estado nativo
$^+_{+}(\neq)$	Estado de transición

## **Aminoácidos**

Ala	(A)	Alanina
Asn	(N)	Asparagina
Asp	(D)	Aspartato
Arg	(R)	Arginina
Cys	(C)	Cisteína
Gln	(Q)	Glutamina
Glu	(E)	Glutamato
Gly	(G)	Glicina
His	(H)	Histidina
Ile	(I)	Isoeucina
Leu	(L)	Leucina
Lys	(K)	Lisina
Met	(M)	Metionina

Phe	(F)	Fenilalanina
Pro	(P)	Prolina
Ser	(S)	Serina
Thr	(T)	Treonina
Trp	(W)	Triptófano
Tyr	(Y)	Tirosina
Val	(V)	Valina



# **1. Introducción**



Las proteínas son las macromoléculas biológicas más abundantes, se hallan en todas las partes de las células constituyendo el 50% de su peso. Intervienen en prácticamente todos los procesos celulares, ejerciendo una diversidad casi inagotable de funciones: catalizan numerosas reacciones químicas, transportan y almacenan otras moléculas, proporcionan apoyo mecánico y protección inmunológica, generan movimiento, transmiten impulsos nerviosos y controlan el crecimiento y la expresión génica [1, 2].

La versatilidad, eficacia y especificidad con la que desarrollan sus tareas, hace que las proteínas tengan importantes aplicaciones en biomedicina y biotecnología. Las aplicaciones terapéuticas de las proteínas están típicamente basadas en hormonas, citoquinas o factores de crecimiento [3]. Otras proteínas, como las enzimas hidrolíticas, son empleadas en múltiples procesos en la industria textil, alimentaria o farmacéutica.

La mayoría de estas aplicaciones requieren alguna modificación de la proteína de interés, para mejorarla con respecto a alguna característica fundamental. Esto se debe a que las proteínas están optimizadas para la realización de su función *in vivo*, donde operan bajo unas condiciones que rara vez coinciden con aquellas requeridas para sus aplicaciones. Por ejemplo, los anticuerpos monoclonales empleados en aplicaciones terapéuticas, son de origen animal, por lo que han de modificarse para evitar respuesta inmune en humanos [4]. Otras proteínas con aplicaciones tecnológicas suelen estar modificadas para incrementar su estabilidad, su solubilidad o su actividad bajo las condiciones de aplicación. En otros casos se ha pretendido la manipulación de los sitios activos de las proteínas de manera que se pueda modificar el reconocimiento molecular entre ligandos con diversos fines, por ejemplo, para comportarse como biosensores [5]. Un objetivo más ambicioso aun, sería el diseño de proteínas con nuevas capacidades catalíticas (*de novo*) [6-9] o incluso con nuevas estructuras no presentes en la naturaleza [10]. El diseño de

enzimas que catalicen la reacción deseada permitiría liberar las aplicaciones de las proteínas de las limitaciones asociadas a las proteínas naturales.

La “Ingeniería de Proteínas” se encarga, normalmente realizando mutaciones, de modificar las propiedades de las proteínas a fin de ser aplicadas en la industria biotecnológica y biomédica. La cuestión es cómo encontrar las mutaciones que provoquen la modulación deseada en las propiedades. Metodológicamente, se han venido aplicando dos estrategias para conseguir proteínas mejoradas con respecto a alguna propiedad:

- 1) El diseño racional o computacional se vale del conocimiento teórico de los principios básicos que rigen la relación entre estructura, estabilidad y función de la proteína, a partir del cual, se diseñan las mutaciones que presumiblemente producen el efecto deseado.

Se han usado varias estrategias para la obtención de proteínas estabilizadas bajo un enfoque racional: modificando la entropía configuracional con mutaciones de glicina a alanina o de cualquier aminoácido a prolina [11] o mediante el diseño de puentes disulfuros [12, 13], introduciendo interacciones entre residuos aromáticos [13-15], optimizando las interacciones electrostáticas de la superficie proteica, estrategia en la que nuestro grupo de investigación tiene amplia experiencia [16, 17], etc.

En otros casos, se han diseñado computacionalmente, enzimas capaces de catalizar procesos en sustratos no naturales [6], e incluso se han diseñado enzimas capaces de llevar a cabo reacciones como la eliminación de Kemp [7] o la reacción de Diels-Alder [18], capacidades catalíticas que no se han encontrado en ninguna proteína natural.

Los trabajos realizados bajo esta perspectiva, aunque notorios, son aun escasos ya que el conocimiento necesario para aplicarla no siempre está disponible.

- 2) Estas limitaciones, pueden ser solventadas con la evolución dirigida *in vitro*. Consiste en crear aleatoriamente una biblioteca de variantes que se analizan conforme a la característica deseada. Las variantes de la biblioteca seleccionadas (aquellas mejoradas con respecto a la propiedad deseada) se pueden someter a nuevos ciclos de aleatorización y selección. Esta estrategia se ha utilizado con éxito para la obtención de proteínas estabilizadas [19], para aumentar actividades catalíticas o para modular la afinidad proteína-ligando [20-22].

En principio, el éxito de los métodos de evolución dirigida *in vitro* está asegurado siempre y cuando se disponga de un sistema eficaz de cribado que permitan analizar bibliotecas muy grandes, condición que no siempre se cumple.

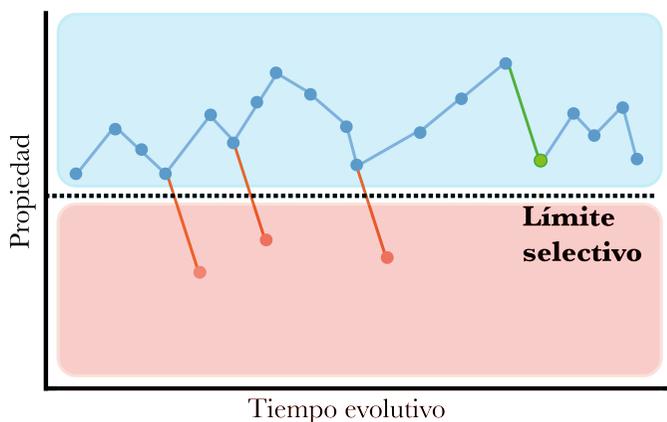
En algunos casos, las limitaciones que acarrea el diseño racional o la evolución dirigida *in vitro* pueden ser solventadas con la combinación de ambas estrategias. Por ejemplo, se ha utilizado el diseño racional para crear una actividad *de novo* con un nivel moderado, que luego se ha incrementado sustancialmente mediante evolución dirigida *in vitro* [7, 23].

Hasta ahora hemos planteado los dos enfoques que, tradicionalmente, se vienen empleado en ingeniería de proteínas. No obstante, existe un enfoque adicional que se basa en la extracción de información acerca de la historia evolutiva de una proteína, a partir de la comparación de secuencias.

Los **métodos basados en información evolutiva** se sustentan en la premisa de que las proteínas tal y como se encuentran en la naturaleza, son resultado de la selección natural, ya que a lo largo de la evolución han sufrido diversos cambios para adaptarse a las condiciones que las rodean. Entender los cambios moleculares producto de la evolución puede aportar información útil para guiar el diseño de proteínas. Dentro de esta perspectiva se encuentran, entre otros, el método de consenso [24, 25] y la resurrección de proteínas ancestrales [26-29].

El **método de consenso** consiste en realizar, sobre un alineamiento de secuencias, un recuento de frecuencias aminoacídicas en cada posición. La secuencia consenso será aquella que contenga los aminoácidos más habituales en cada posición.

A lo largo de la evolución, las secuencias de las proteínas han ido sufriendo mutaciones, que debido a la presencia de determinados umbrales fenotípicos que actúan como límites selectivos, han podido ser toleradas o rechazadas. Las secuencias de proteínas actuales han sufrido esta criba. Así que, una mutación que produzca un efecto negativo en la propiedad, será aceptada con poca frecuencia durante la evolución (sólo cuando mutaciones previas hayan aumentado significativamente el valor de la propiedad) y, como consecuencia, aparecerá con poca frecuencia en las secuencias de las proteínas actuales. De la misma manera, la mutación que provoque un efecto positivo en la propiedad estará representada con mayor frecuencia en las secuencias de las proteínas actuales. Véase fig 1.1.



*fig 1.1. Variación del valor de una propiedad de una proteína a lo largo de la evolución. Cada segmento representa el efecto de una mutación en la propiedad. La línea discontinua marca el límite selectivo por debajo del cual la proteína no es viable. Cuando una mutación rebaja el valor de la propiedad por debajo del límite selectivo, la mutación es rechazada (segmentos en rojo). Por ello, una mutación con un efecto negativo, sólo será aceptada, cuando mutaciones previas hayan aumentado significativamente el valor de la propiedad (segmento en verde). Esto hace que las mutaciones con un efecto negativo en la propiedad se aceptarán con poca frecuencia a lo largo de la evolución. Cuando una mutación tiene un efecto positivo en la propiedad, es aceptada con mayor frecuencia y, por tanto, estará más representada en las secuencias de las proteínas actuales.*

Cuando realizamos un alineamiento de la secuencia de una proteína con secuencias de la misma proteína en otros organismos, podemos pensar que mutaciones al aminoácido que aparece con más frecuencia en cada posición del alineamiento (aminoácido consenso) presumiblemente tienen un efecto beneficioso, en el sentido de aumentar la estabilidad o la función por ejemplo.

Esta metodología ha sido utilizada exitosamente en la predicción de mutaciones estabilizadoras en varias ocasiones [25, 30, 31]. Recientemente se ha demostrado que, combinando los residuos más frecuentes en un alineamiento sobre tiorredoxina, se pueden optimizar de forma independiente la estabilidad y la actividad enzimática de la tiorredoxina. Así que es posible su optimización bajo ambas propiedades de forma simultánea [25].

La posibilidad de reconstruir secuencias de **proteínas ancestrales** mediante el análisis comparativo de secuencias relacionadas en organismos actuales, fue sugerida por Linus Pauling y Emile Zuckerkandl hace unos 50 años [32]. Más tarde, el campo de la paleobioquímica que estos autores imaginaron, cobró vida, a mediados de los años noventa, gracias al crecimiento del número de secuencias conocidas, al desarrollo de procedimientos estadísticos fiables para la reconstrucción de secuencias ancestrales y a los avances en ingeniería genética, que permitieron resucitar y estudiar *in vitro* proteínas de organismos extintos.

Desde entonces, existen algunos trabajos en los que se han resucitado proteínas ancestrales, que aunque escasos en número, han proporcionado una gran información sobre el medio físico que rodeaba a los organismos extintos y cómo sus proteínas, y por tanto los organismos, se han ido adaptando a los cambios en ese entorno [33]. Por ejemplo, la resurrección de factores de elongación de organismos ancestros, y el estudio de su termoestabilidad, ha revelado información sobre condiciones ambientales del pasado [26, 27]. Este estudio apoyó la idea de que los mares primitivos eran cálidos y que se han ido enfriando a lo largo de la evolución de la Tierra, lo cual es consistente con las estimaciones de la

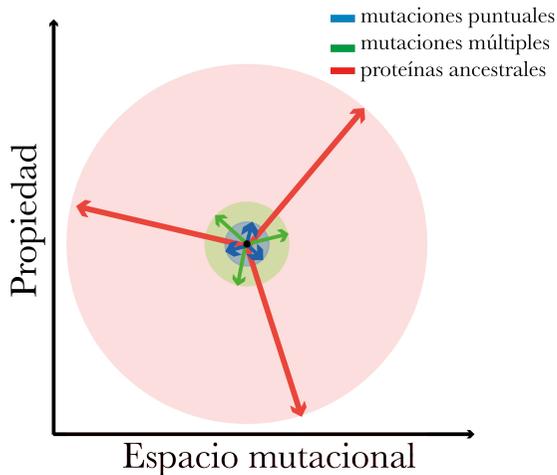
temperatura de los mares primitivos, derivadas de los estudios de deposición de isótopos de oxígeno en rocas. Igualmente, estudios sobre tiorredoxinas precámbricas resucitadas, apoyan el carácter ácido de los mares primitivos, que cabe esperar del hecho de que la atmósfera primitiva posiblemente contenía una gran proporción de CO<sub>2</sub> [29].

Otro ejemplo de interés lo proporcionan los estudios sobre los pigmentos visuales en arcosaurios. Los arcosaurios son reptiles que han existido desde hace más de 200 millones de años y que incluyen a dinosaurios, pájaros, y cocodrilos. Lo poco que se sabe de la fisiología de los arcosaurios primitivos proviene de restos fósiles, y del estudio biológico de sus descendientes vivos. Recientemente, la resurrección de un pigmento visual de arcosaurios primitivos y su posterior estudio, ha mostrado propiedades interesantes sobre la capacidad sensitiva visual de estos animales de las que se deduce su capacidad de visión nocturna [34, 35].

Se puede decir, que la historia de la vida en la Tierra es una crónica escrita sobre los estratos geológicos, el registro fósil, y los genomas de los organismos actuales. El análisis conjunto de estos datos, proporciona una mejor comprensión de cómo la vida interactúa con la biosfera y se adapta al cambio global.

Además, a medida que se han ido resucitando proteínas ancestrales, su estudio ha revelado características muy diferentes a las de las proteínas actuales, incluyendo niveles incrementados de función en algunos casos y muy alta estabilidad [29, 36-38]. La relevancia de este resultado se puede explicar teniendo en cuenta que modificaciones en la secuencia de los aminoácidos de una proteína, normalmente, dan lugar a cambios en sus propiedades. Así, mutaciones puntuales, provocan cambios modestos en las propiedades de las proteínas. Por el contrario, grandes cambios en la secuencia, provocan mayores modulaciones en sus propiedades. El estudio de proteínas ancestrales, yendo muy atrás en el tiempo, permite obtener proteínas con secuencias muy diferentes a las de sus análogas actuales (pudiéndose llegar a incluir mutaciones en más del 50% de la secuencia proteica). Esto nos

permite abarcar un amplio espacio mutacional/evolutivo que presumiblemente lleva asociado un amplio rango de modulación en la propiedad en cuestión (véase fig 1.2).



*fig 1.2. Relación entre el rango de modulación en la propiedad que se explora, en función del número de mutaciones que se realiza en la secuencia proteica. Con mutaciones puntuales se produce un efecto pequeño en la propiedad (área azul), con varias mutaciones se produce un efecto un poco mayor en la propiedad (área verde) y con grandes cambios en la secuencia de una proteína se alcanzan grandes rangos de modulación en la propiedad (área roja).*

La ingeniería de proteínas se encarga, entre otras funciones, de encontrar las mutaciones que provoquen un efecto deseado en las propiedades de las proteínas, para que puedan ser utilizadas en el campo de la biotecnología y la biomedicina. En este sentido, la resurrección de proteínas ancestrales puede convertirse en una de las principales fuentes de la próxima generación de proteínas con aplicaciones biotecnológicas y biomédicas.

En la presente Tesis Doctoral se muestran dos líneas de investigación, cada una de ellas constituyen ejemplos de metodologías útiles en ingeniería de proteínas.

En un primer bloque, hemos explorado la posibilidad de utilizar el método de consenso para obtener variantes de Proteínas Morfogenéticas de Hueso con su función biológica mejorada. Este trabajo ha sido desarrollado en colaboración con el grupo de Ingeniería Tisular liderado por J.L López Lacomba (Universidad Complutense de Madrid) y la empresa Noricum S.L..

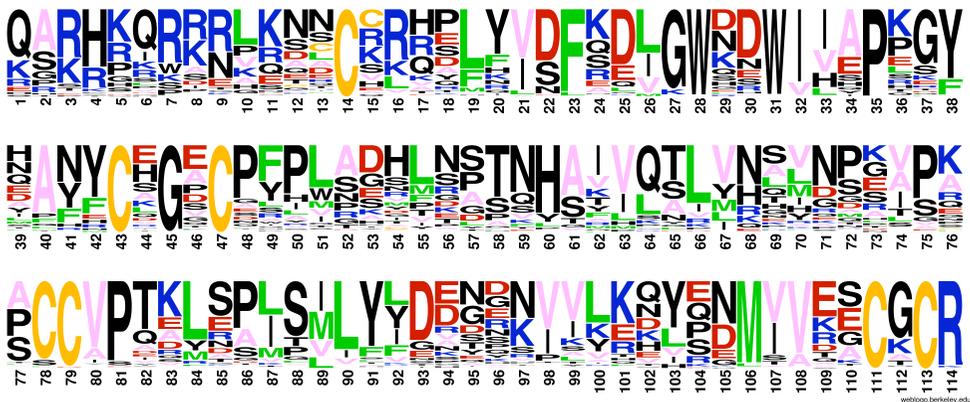
En un segundo bloque, en el que nos valemos de la resurrección de tiorredoxinas ancestrales, hemos abordado el estudio de sus propiedades inusuales y hemos

realizado lo que posiblemente sea el primer trabajo de ingeniería de proteínas que toma como punto de partida una enzima ancestral resucitada.

### **Proteína Morfogenética de Hueso 2 (BMP-2)**

La BMP-2 es un factor de crecimiento, de gran interés en el campo de la biomedicina, por su capacidad para inducir regeneración ósea. Destacan entre sus aplicaciones, las relacionadas con la fusión espinal, el tratamiento de fracturas, la cirugía ortopédica y la cirugía oral [39-42].

Por sus aplicaciones, nos planteamos explorar la posibilidad de que el enfoque de consenso pudiera conducir a variantes de BMP-2 optimizadas conforme a su actividad biológica, en concreto, conforme a su capacidad para diferenciar una línea celular mioblástica a osteoblástica. Para ello, realizamos un alineamiento de la secuencia proteica de BMP-2 con el resto de secuencias de proteínas homólogas contenidas en las bases de datos Swiss-Prot y TrEMBL. Bajo la hipótesis de que las mutaciones que aparezcan con más frecuencia en cada posición del alineamiento, pudiesen estar relacionadas con la actividad biológica de la BMP-2, realizamos un análisis de consenso (fig 1.3).



*fig 1.3. Representación del alineamiento de secuencias de BMP-2 con secuencias de proteínas homólogas contenidas en la base de datos Swiss-Prot. El tamaño con el que se representa cada aminoácido, es proporcional a la frecuencia con la que aparece en el alineamiento para cada posición. Así, la secuencia consenso del alineamiento viene dada por el aminoácido de mayor tamaño en cada posición. Obtenido de la página web <http://weblogo.berkeley.edu/>.*

Obtuvimos variantes de BMP-2 con las mutaciones propuestas por el análisis y se caracterizaron conforme a su actividad biológica. Para testar su actividad, se han llevado a cabo ensayos celulares *in vitro* basados en la detección de la actividad fosfatasa alcalina asociada al fenotipo osteoblástico inducido. Los resultados obtenidos se muestran en la sección 3.1.

### **Tiorredoxinas ancestrales resucitadas**

Como resultado de un trabajo colaborativo entre los grupos de investigación de Eric. A. Gaucher (Georgia Institute of Technology, Atlanta), de Julio M. Fernández (Universidad de Columbia, Nueva York) y de José Manuel Sánchez Ruiz (Universidad de Granada), muy recientemente, se han resucitado y caracterizado siete tiorredoxinas precámbricas [29].

A partir de la construcción de un árbol filogenético, usando 203 secuencias de tiorredoxinas actuales representativas de los tres dominios de la vida, Eric A. Gaucher (Georgia Institute of Technology, Atlanta) reconstruyó las secuencias de tiorredoxinas ancestrales asociadas a nodos extintos del árbol. En fig 1.4 se muestra un esquema del árbol utilizado que muestra el tiempo geológico en el que el organismo extinto vivió [29]. Nótese que las tiorredoxinas más antiguas datan de aproximadamente 4200 millones de años, de manera que estas enzimas operaban en tiempos cercanos al origen de la vida en la Tierra.

La resurrección experimental de estas tiorredoxinas precámbricas se realizó en nuestro laboratorio, en Granada, donde también se llevó a cabo una caracterización biofísica exhaustiva de las mismas. El estudio de la estabilidad térmica de estas proteínas ancestrales ha mostrado que son hiper-estables, propiedad que es muy interesante desde el punto de vista de la ingeniería de proteínas [29].

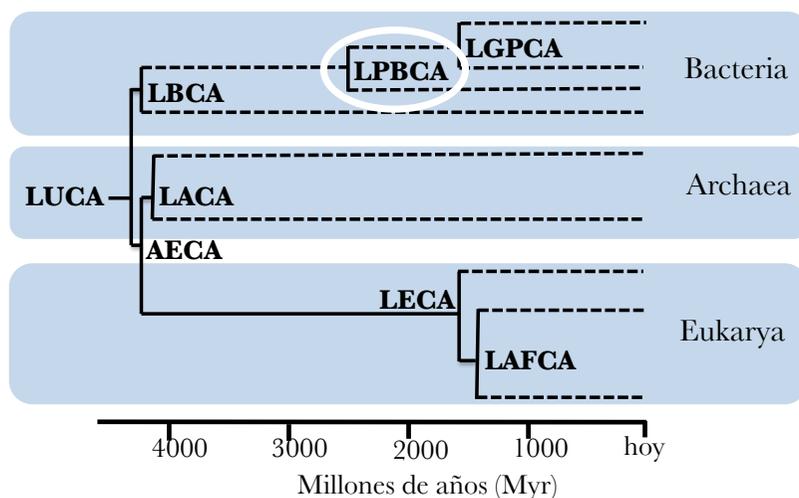


fig 1.4. Esquema del árbol filogenético que muestra el tiempo geológico en el que el organismo extinto vivió. A partir de secuencias de tiorredoxinas actuales de los tres dominios de la vida, se ha inferido la secuencia de las tiorredoxinas asociada a los nodos indicados.

En el laboratorio de Julio M. Fernández (Universidad de Columbia, Nueva York), mediante Microscopía de Fuerza Atómica (AFM) de molécula individual, se llevó a cabo el estudio de la actividad enzimática de las tiorredoxinas precámbricas. Dicho estudio reveló que las tiorredoxinas más antiguas presentan una mayor actividad en condiciones ácidas [29], lo que sugiere la posibilidad de utilizar la resurrección ancestral para modular el perfil actividad/pH de enzimas, un objetivo de claro interés biotecnológico.

Recientemente, en colaboración con José A. Gavira (perteneciente al Laboratorio de Estudios Cristalográficos del CSIC), se han resuelto las estructuras tridimensionales de las siete tiorredoxinas precámbricas, mediante cristalografía de rayos X. Este estudio reveló que todas las tiorredoxinas resucitadas tenían el mismo *fold* que sus análogas actuales.

Como consecuencia de todos estos trabajos, nos encontramos en una posición ideal para abordar el estudio de los factores moleculares, energéticos y evolutivos

responsables de las características tan inusuales que tienen estas tiorredoxinas ancestrales.

Así, como parte del segundo bloque de esta Tesis Doctoral, nos planteamos un estudio detallado de los factores que determinan la muy alta estabilidad de estas proteínas ancestrales tomando como objeto de estudio específico la tiorredoxina perteneciente al último ancestro común de cianobacterias y los grupos de deinococcus y thermus (LPBCA de sus siglas en inglés), por ser la más estable entre las tiorredoxinas precámbricas resucitadas. Los resultados de este trabajo se muestran en la sección 3.2.1.

En el transcurso de estos estudios sobre la tiorredoxina de LPBCA, obtuvimos resultados experimentales que parecían indicar una desnaturalización fría no cooperativa en esta proteína y que sugerían la posibilidad de que las tiorredoxinas ancestrales resucitadas nos pudieran permitir abordar el problema de la evolución de la cooperatividad. Esta posibilidad, sin embargo, no se confirmó ya que encontramos el mismo patrón en la desnaturalización fría de la tiorredoxina actual de *E. coli* y, además, pudimos explicarlo en base a la lentitud de la cinética de desplegamiento a baja temperatura. En cualquier caso, estos estudios son relevantes para entender la desnaturalización fría de proteínas y se describen en el apartado 3.2.2.

En el último trabajo que se presenta en esta Tesis Doctoral, hemos comenzado a explorar la posibilidad de utilizar una proteína ancestral, por su hiper-estabilidad, como punto de partida para proyectos de ingeniería de proteínas. La estabilidad es una limitación en muchos estudios de ingeniería de proteínas, en particular en aquellos enfocados al diseño *de novo*. La razón es que la introducción de mutaciones que constituyan el sitio activo producen, normalmente, una disminución de la estabilidad de la proteína. Por ejemplo, recientemente se han diseñado computacionalmente enzimas con actividades *de novo*, que catalizan reacciones como la eliminación de Kemp. Gracias a la evolución dirigida *in vitro*,

algunas de las variantes diseñadas, se optimizaron para aumentar su eficacia catalítica [7]. Sin embargo, el diseño que presentaba más eficiencia en la catálisis de la reacción de Kemp, por su baja estabilidad, no pudo ser optimizada por procedimientos de evolución dirigida *in vitro*. Por tanto, se hizo necesaria su estabilización previa, para posteriormente poder ser optimizada bajo su actividad [43]. En este sentido, las proteínas ancestrales al ser tan estables, pueden ser muy apropiadas como punto de inicio para objetivos de ingeniería de proteínas; esta es una posibilidad, que sin embargo, hasta donde sabemos no ha sido utilizada aun. En este trabajo, en concreto, hemos investigado la posibilidad de rediseñar el centro activo de la tiorredoxina de LPBCA (la más estable de entre las tiorredoxinas ancestrales que hemos resucitado), de manera que el efecto de un grupo ionizable enterrado, esencial para la catálisis enzimática, es llevado a cabo por un grupo expuesto al disolvente. Este trabajo se describe en la sección 3.2.3 de esta Tesis Doctoral.

## 1.1. Bibliografía

1. Jeremy M Berg, J.L.T., and Lubert Stryer, *Biochemistry, 5th edition*. New York: W H Freeman, 2002.
2. Christopher K. Mathews, K.G.A., K. E. Van Holde, *Bioquímica 3ª ed.* Addison Wesley, 2002.
3. Lazar, G.A., et al., *Designing proteins for therapeutic applications*. Current opinion in structural biology, 2003. **13**(4): p. 513-8.
4. Maynard, J. and G. Georgiou, *Antibody engineering*. Annual review of biomedical engineering, 2000. **2**: p. 339-76.
5. Looger, L.L., et al., *Computational design of receptor and sensor proteins with novel functions*. Nature, 2003. **423**(6936): p. 185-90.
6. Jiang, L., et al., *De novo computational design of retro-aldol enzymes*. Science, 2008. **319**(5868): p. 1387-91.
7. Rothlisberger, D., et al., *Kemp elimination catalysts by computational enzyme design*. Nature, 2008. **453**(7192): p. 190-5.
8. Suarez, M., et al., *Using multi-objective computational design to extend protein promiscuity*. Biophysical chemistry, 2010. **147**(1-2): p. 13-9.
9. Bolon, D.N. and S.L. Mayo, *Enzyme-like proteins by computational design*. Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(25): p. 14274-9.
10. Kuhlman, B., et al., *Design of a novel globular protein fold with atomic-level accuracy*. Science, 2003. **302**(5649): p. 1364-8.
11. Matthews, B.W., H. Nicholson, and W.J. Becktel, *Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding*. Proceedings of the National Academy of Sciences of the United States of America, 1987. **84**(19): p. 6663-7.
12. Matsumura, M., G. Signor, and B.W. Matthews, *Substantial increase of protein stability by multiple disulphide bonds*. Nature, 1989. **342**(6247): p. 291-3.

13. Mansfeld, J., et al., *Extreme stabilization of a thermolysin-like protease by an engineered disulfide bond*. The Journal of biological chemistry, 1997. **272**(17): p. 11152-6.
14. Serrano, L., M. Bycroft, and A.R. Fersht, *Aromatic-aromatic interactions and protein stability. Investigation by double-mutant cycles*. Journal of molecular biology, 1991. **218**(2): p. 465-75.
15. Puchkaev, A.V., L.S. Koo, and P.R. Ortiz de Montellano, *Aromatic stacking as a determinant of the thermal stability of CYP119 from Sulfolobus solfataricus*. Archives of biochemistry and biophysics, 2003. **409**(1): p. 52-8.
16. Sanchez-Ruiz, J.M. and G.I. Makhatadze, *To charge or not to charge?* Trends in biotechnology, 2001. **19**(4): p. 132-5.
17. Ibarra-Molero, B. and J.M. Sanchez-Ruiz, *Genetic Algorithm to Design Stabilizing Surface-Charge Distributions in Proteins*. The Journal of Physical Chemistry B, 2002. **106**(26): p. 6609-6613.
18. Siegel, J.B., et al., *Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction*. Science, 2010. **329**(5989): p. 309-13.
19. Eijssink, V.G., et al., *Directed evolution of enzyme stability*. Biomolecular engineering, 2005. **22**(1-3): p. 21-30.
20. Dougherty, M.J. and F.H. Arnold, *Directed evolution: new parts and optimized function*. Current opinion in biotechnology, 2009. **20**(4): p. 486-91.
21. Cobb, R.E., N. Sun, and H. Zhao, *Directed evolution as a powerful synthetic biology tool*. Methods, 2012.
22. Wang, M., T. Si, and H. Zhao, *Biocatalyst development by directed evolution*. Bioresource technology, 2012. **115**: p. 117-25.
23. Huang, J., et al., *Design of protein function leaps by directed domain interface evolution*. Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(18): p. 6578-83.
24. Steipe, B., et al., *Sequence statistics reliably predict stabilizing mutations in a protein domain*. Journal of molecular biology, 1994. **240**(3): p. 188-92.

25. Rodríguez-Larrea, D., et al., *Role of conservative mutations in protein multi-property adaptation*. The Biochemical journal, 2010. **429**(2): p. 243-9.
26. Gaucher, E.A., S. Govindarajan, and O.K. Ganesh, *Palaeotemperature trend for Precambrian life inferred from resurrected proteins*. Nature, 2008. **451**(7179): p. 704-7.
27. Gaucher, E.A., et al., *Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins*. Nature, 2003. **425**(6955): p. 285-8.
28. Jermann, T.M., et al., *Reconstructing the evolutionary history of the artiodactyl ribonuclease superfamily*. Nature, 1995. **374**(6517): p. 57-9.
29. Perez-Jimenez, R., et al., *Single-molecule paleoenzymology probes the chemistry of resurrected enzymes*. Nature structural & molecular biology, 2011. **18**(5): p. 592-6.
30. Lehmann, M., et al., *The consensus concept for thermostability engineering of proteins: further proof of concept*. Protein engineering, 2002. **15**(5): p. 403-11.
31. Lehmann, M., et al., *The consensus concept for thermostability engineering of proteins*. Biochimica et biophysica acta, 2000. **1543**(2): p. 408-415.
32. Pauling L, Z.E., *Chemical paleogenetics. Molecular "restoration studies" of extinct forms of life*. Acta Chemica Scandinava, 1963(17): p. S9-S16.
33. Benner, S.A., S.O. Sassi, and E.A. Gaucher, *Molecular paleoscience: systems biology from the past*. Advances in enzymology and related areas of molecular biology, 2007. **75**: p. 1-132, xi.
34. Chang, B.S., M.A. Kazmi, and T.P. Sakmar, *Synthetic gene technology: applications to ancestral gene reconstruction and structure-function studies of receptors*. Methods in enzymology, 2002. **343**: p. 274-94.
35. Chang, B.S., *Ancestral gene reconstruction and synthesis of ancient rhodopsins in the laboratory*. Integrative and comparative biology, 2003. **43**(4): p. 500-7.

36. Miyazaki, J., et al., *Ancestral residues stabilizing 3-isopropylmalate dehydrogenase of an extreme thermophile: experimental evidence supporting the thermophilic common ancestor hypothesis*. Journal of biochemistry, 2001. **129**(5): p. 777-82.
37. Fukami-Kobayashi, K., D.R. Schreiber, and S.A. Benner, *Detecting compensatory covariation signals in protein evolution using reconstructed ancestral sequences*. Journal of molecular biology, 2002. **319**(3): p. 729-43.
38. Shimizu, H., et al., *Extremely thermophilic translation system in the common ancestor commonote: ancestral mutants of Glycyl-tRNA synthetase from the extreme thermophile Thermus thermophilus*. Journal of molecular biology, 2007. **369**(4): p. 1060-9.
39. Cochran, D.L., et al., *Evaluation of recombinant human bone morphogenetic protein-2 in oral applications including the use of endosseous implants: 3-year results of a pilot study in humans*. Journal of periodontology, 2000. **71**(8): p. 1241-57.
40. Govender, S., et al., *Recombinant human bone morphogenetic protein-2 for treatment of open tibial fractures: a prospective, controlled, randomized study of four hundred and fifty patients*. The Journal of bone and joint surgery. American volume, 2002. **84-A**(12): p. 2123-34.
41. Einhorn, T.A., et al., *A single percutaneous injection of recombinant human bone morphogenetic protein-2 accelerates fracture repair*. The Journal of bone and joint surgery. American volume, 2003. **85-A**(8): p. 1425-35.
42. Granjeiro, J.M., et al., *Bone morphogenetic proteins: from structure to clinical use*. Brazilian journal of medical and biological research = Revista brasileira de pesquisas medicas e biologicas / Sociedade Brasileira de Biofisica ... [et al.], 2005. **38**(10): p. 1463-73.
43. Khersonsky, O., et al., *Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59*. Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(26): p. 10358-63.

## **2. Metodología**



## 2.1. Obtención de ADN de proteínas mutantes

Se utilizó el método de mutagénesis dirigida quikchange™ de STRATAGENE, que permite realizar mutaciones en un plásmido de cadena doble (ADN-cd). El procedimiento básico se puede resumir en:

- 1) Diseño de oligonucleótidos cebadores (oligos).

Los oligos se diseñaron siguiendo las siguientes recomendaciones:

- Los oligos deben ser complementarios entre sí y cada uno de ellos a una de las cadenas del plásmido.
- Si definimos la temperatura de fusión del oligo  $T_m$  como la temperatura a la cual la mitad del oligo no se encuentra unido al vector de ADN, se deben diseñar de manera que la  $T_m$  sea superior a  $75^\circ\text{C}$ . La  $T_m$  de los oligos se puede estimar con la siguiente fórmula:

$$T_m = 81.5 + 0.41(\%GC) - \frac{675}{N} - \% \text{bases mutadas en el oligo} \quad \text{ec.2.1}$$

Donde  $N$  es el número total de bases del oligo y  $\%GC$  es el tanto por ciento de Guanidinas/Citosinas.

- Deben contener entre 25-45 pares de bases.
- La mutación debe encontrarse centrada.
- El  $\%GC$  debe ser superior al 40% y la secuencia del oligo debe termina en uno o varios G/C.

- 2) Condiciones de la PCR.

Las condiciones utilizadas fueron las indicadas en el manual del QuikChange Site-Directed Mutagenesis Kit:

La mezcla de reacción consta de los siguientes reactivos:

- 5  $\mu\text{L}$  de tampón 10x
- x  $\mu\text{L}$  (5-50ng) de ADN parenteral
- 1.25  $\mu\text{L}$  de primer directo 5'→3' (100 ng/ $\mu\text{L}$ )
- 1.25  $\mu\text{L}$  de primer reverso 3'→5' (100 ng/ $\mu\text{L}$ )
- 1  $\mu\text{L}$  de mezcla de nucleótidos dNTPs (10 mM)
- añadir agua Milli-Q estéril hasta un volumen final de 50  $\mu\text{L}$
- 1  $\mu\text{L}$  de Pfu Turbo ADN Polimerasa (2.5 U/ $\mu\text{L}$ )

El programa que se utilizó en el termociclador *Mastercycler gradient* de eppendorf para realizar la PCR es el siguiente:

Número de ciclos	Temperatura ( $^{\circ}\text{C}$ )	Duración (seg)
1	95	30
16	95	30
	55	60
	68	70

Una vez finalizado el programa del termociclador se dejó enfriar cada tubo de PCR manteniéndolo 2 min en hielo hasta  $T \leq 37^{\circ}\text{C}$ , se le añadió 1 $\mu\text{L}$  de enzima de restricción Dpn I (10 U/ $\mu\text{L}$ ), se agitó suavemente y se incubó durante 1h a  $37^{\circ}\text{C}$ . La enzima Dpn I es específica para ADN metilado y hemimetilado, de esta forma se digiere el ADN-cd parental no mutado y queda libre el ADN-cd sintetizado por la PCR que es de estructura lineal y que contiene la mutación deseada.

- 3) Transformación de las células supercompetentes XL-1 Blue.

Las células XL1-Blue se encuentran congeladas a  $-80^{\circ}\text{C}$  en alícuotas de  $200\ \mu\text{L}$ . Una vez descongeladas, se hicieron alícuotas de  $50\ \mu\text{L}$  de células en los tubos Falcon® 2059 de polipropileno que previamente se habían enfriado en hielo. El protocolo que se siguió es el siguiente:

- Se añade  $1\ \mu\text{L}$  de producto de PCR (previamente tratado con Dpn I) a cada alícuota de células y se agita suavemente.
- Se incuba  $30\ \text{min}$  en hielo.
- Se introduce en un baño termostatzado a  $42^{\circ}\text{C}$  durante  $45\ \text{segundos}$ . El pulso de calor es la etapa más crítica de la transformación.
- Se incuba en hielo durante  $2\ \text{minutos}$ .
- Se añade  $0.5\ \text{mL}$  de medio de cultivo NZY<sup>+</sup> previamente calentado a  $42^{\circ}\text{C}$  y se deja incubar en un orbital a  $37^{\circ}\text{C}$  durante  $1\ \text{h}$ .
- Finalmente se plaquean  $100\text{-}200\ \mu\text{L}$  de mezcla de transformación en placas con LB-Agar a una concentración adecuada de antibiótico ( $100\ \mu\text{L}/\text{mL}$  de ampicilina en el caso del vector pQE - 80L y pT7.7 o  $30\ \text{ng}/\text{mL}$  de kanamicina en el caso del vector pET30a (+) y pET24 (+)) y se incuba en una estufa termostatzada a  $37^{\circ}\text{C}$  al menos  $16\ \text{h}$ . Al cabo de este periodo de tiempo crecerán solo las colonias de células que contengan el plásmido con el gen que ofrece resistencia al antibiótico en cuestión.

#### 4) Preparación de ADN-cd mutado.

De las placas crecidas durante toda la noche, se eligieron  $5\ \text{colonias}$  y se dejaron crecer en un preinóculo de medio de cultivo LB con antibiótico. El ADN - cd se extrajo de dicha células utilizando el kit QIAprep®. Por último, la secuenciación de ADN y la síntesis de oligonucleótidos fueron llevadas a cabo por la empresa Eurosfins MWG Operon (<http://www.eurofinsdna.com/home.html>).

## 2.2. Clonación de variantes ancestrales de tiorredoxina en el vector pET30a (+)

Las variantes de tiorredoxina que se encuentran clonadas en el vector pQE - 80L, se expresan con una cola de 6 histidinas en la zona amino terminal de la proteína. Esto hace que el proceso de purificación sea sencillo y permite el estudio de un gran número de proteínas en poco tiempo. Sin embargo la cola de histidina puede repercutir en la estabilidad de la proteína, por ello las variantes que resultaron más interesantes se clonaron en el vector pET30a (+), se purificaron sin la cola de histidina y se estudiaron de nuevo. El proceso de clonación se puede resumir en:

### 1) Diseño de los oligos.

Se diseñaron los oligos de manera que introducen los sitios de restricción *NdeI* y *XhoI* respectivamente en 5' y 3' de la secuencia codificante:

Directo: 5'- AGCTACGACCATATGTCTGTAATTGAAGTCAC- 3'

Reverso: 5'- TCAGATCTCGAGTTACAGATGTTTATCGATGC - 3'

### 2) Amplificación del inserto mediante PCR.

Se utilizó una polimerasa de alta fidelidad. La mezcla de reacción consta de los siguientes reactivos de STRATAGENE:

- 5 µL de tampón 10x
- 1 µL (50ng) de ADN parental
- 1.5 µL de primer directo 5'→3' (100 ng/µL)
- 1.5 µL de primer reverso 3'→5' (100 ng/µL)
- 1 µL de mezcla de nucleótidos dNTP (10 mM)
- 39 µL de agua Milli-Q estéril
- 1 µL de Pfu Turbo ADN Polimerasa (2.5 U/µL)

El programa que se utilizó en el termociclador Mastercycler gradient de eppendorf para realizar la PCR es el siguiente:

Número de ciclos	Temperatura (°C)	Duración (seg)
1	95	30
12	95	30
	55	60
	68	60
1	68	600
1	4	∞

Se comprobó mediante electroforesis en gel de agarosa (0.8% agarosa) que se había amplificado el inserto (fig.2.1).

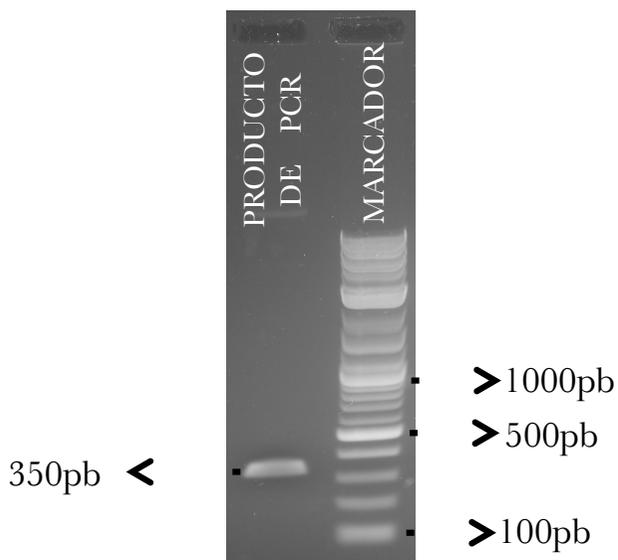


fig.2.1. Gel de agarosa al 0.8% que muestra la amplificación del inserto de 350 pares de bases.

## 3) Purificación del producto de PCR.

Se utilizó el kit QIAquick® siguiéndose exactamente el protocolo que le acompaña. Finalmente se eluyó el inserto en 50 µL con agua Milli-Q estéril.

## 4) Digestión del inserto y del vector pET30a (+).

La mezcla de digestión del inserto consta de los siguientes reactivos:

- 2.5 µL de tampón 4
- 20.25 µL de inserto (100 ng)
- 0.25 µL de BSA de NEW ENGLAND *BioLabs* (100 mg/mL)
- 1 µL *XhoI* de NEW ENGLAND *BioLabs*
- 1 µL *NdeI* de NEW ENGLAND *BioLabs*

y la mezcla de digestión del vector:

- 5 µL de tampón 4
- 40.5 µL de vector pET30a(+) (1µg de ADN)
- 0.5 µL de BSA de NEW ENGLAND *BioLabs* (100 mg/mL)
- 2 µL *XhoI* de NEW ENGLAND *BioLabs*
- 2 µL *NdeI* de NEW ENGLAND *BioLabs*

En ambos casos se incubaron las mezclas de reacción durante 2h a 37°C

## 5) Purificación del producto de digestión.

De nuevo, se utilizó el kit QIAquick® y la mezcla de digestión del inserto y del vector se eluyeron con 50 µL de agua Milli-Q.

## 6) Ligación del inserto en el vector pET30a (+).

La mezcla de ligación consta de los siguientes reactivos:

- 16  $\mu\text{L}$  de vector (1.5 ng/ $\mu\text{L}$ )
- 0.5  $\mu\text{L}$  de inserto (10 ng/ $\mu\text{L}$ )
- 2  $\mu\text{L}$  de tampón 10x
- 1  $\mu\text{L}$  de ligasa T4 de NEW ENGLAND de *BioLabs*

Se incubó durante 15h a 16°C.

- 7) Transformación de células XL1-Blue competentes con el producto de ligación (tal y como se describe en el apartado 2.1.3).
- 8) Selección de clones con inserto mediante PCR de colonia y secuenciación.

Con objeto de seleccionar aquellos clones que contengan el inserto en el vector pET30a (+), de las placas crecidas durante toda la noche, se eligieron 11 colonias al azar y se volvieron a plaquear las colonias en placas individuales. Tras plaquear, el asa se introdujo en 20  $\mu\text{L}$  de agua estéril, se calentó 10 min a 100°C y se centrifugó a 13000 rpm durante 10 min. Se tomaron 15.3  $\mu\text{L}$  del sobrenadante para comprobar mediante PCR la amplificación del inserto.

La mezcla de PCR contiene los siguientes reactivos:

- 15.3  $\mu\text{L}$  de sobrenadante
- 2  $\mu\text{L}$  de tampón 10x que acompaña a la enzima
- 0.5  $\mu\text{L}$  de mezcla de nucleótidos dNTP (100 mM) de STRATAGENE
- 1  $\mu\text{L}$  de primer directo 5'→3' (100 ng/ $\mu\text{L}$ )
- 1  $\mu\text{L}$  de primer reverso 3'→5' (100 ng/ $\mu\text{L}$ )
- 0.2  $\mu\text{L}$  de DFS-Taq DNA polimerasa (5 U/ $\mu\text{L}$ ) de BIORON

El programa que se utilizó en el termociclador *Mastercycler gradient* de eppendorf para realizar la PCR es el siguiente:

Número de ciclos	Temperatura (°C)	Duración (seg)
1	94	120
30	94	20
	55	20
	72	60
1	72	600
1	4	∞

Tras la PCR se comprobó mediante electroforesis en gel de agarosa (fig.2.2) aquellos clones que contenían el inserto. Los clones seleccionados (col 2 y col 7) se inocularon en medio de cultivo LB con 30 ng/mL de kanamicina y se incubaron durante 16h a 37°C. El ADN plasmídico se extrajo utilizando el kit QIAprep® y se envió a secuenciar tal y como se describe en el apartado 2.1.4.

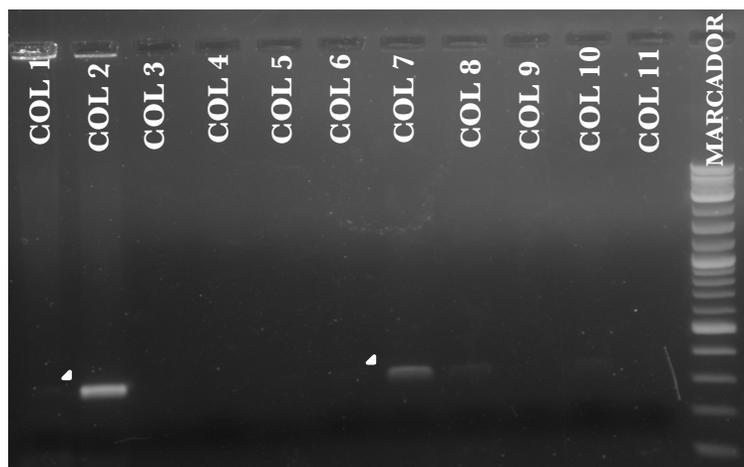


fig.2.2. Gel de agarosa al 0.8% que muestra la amplificación del inserto de 350 pb en los clones 2 y 7 que posteriormente se secuenciaron.

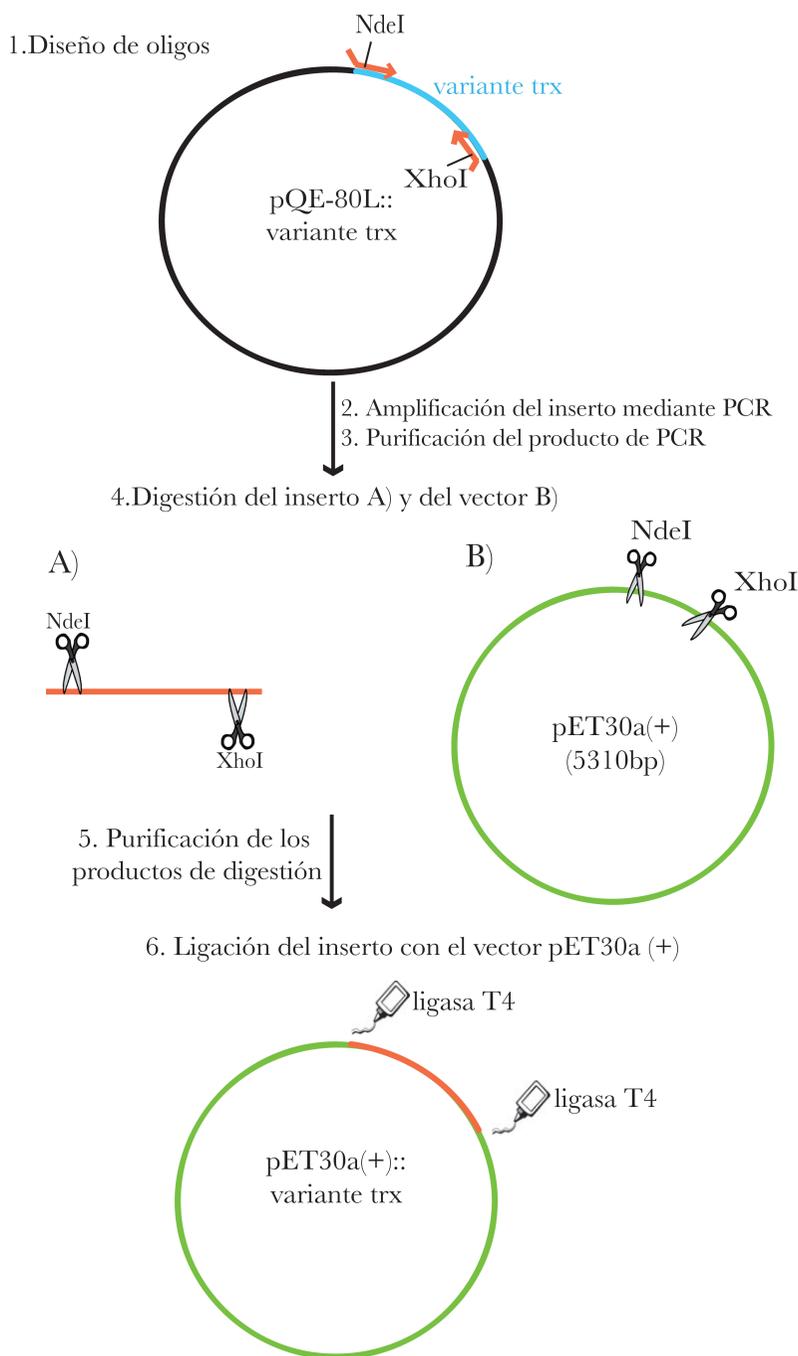


fig.2.3. Esquema resumido de las etapas realizadas para subclonar la secuencia codificante de una variante ancestral de tiorredoxina del vector pQE-80L al vector pET30a(+). En azul se muestra la secuencia codificante de la variante de tiorredoxina, en rojo se muestra el inserto que se amplifica en la PCR, en negro el vector pQE-80L y en verde el vector pET30a(+).

### **2.3. Construcción de bibliotecas combinatoriales de variantes de tiorredoxina**

Llamamos biblioteca combinatorial de mutantes al conjunto de variantes de un gen que contenga todas las combinaciones posibles de una serie de mutaciones elegidas *a priori*. La construcción de las bibliotecas combinatoriales de variantes de tiorredoxina se realizaron basándose en el método “*Gene Assembly Mutagenesis*” [1], que se basa en el hecho de que se puedan construir genes enteros a partir de pequeños oligonucleótidos solapantes que pueden ser ensamblados por la ADN polimerasa. Brevemente se puede resumir en:

- Diseño de oligos.

Se sintetizan los oligos necesarios para construir todas las variantes del gen que codifiquen tanto la hebra positiva como la negativa. Deben tener una longitud de unos 40 pares de bases y han de codificar las dos hebras del ADN. En los oligos de los extremos, se incluyeron los sitios de anclaje de las enzimas de restricción para un posterior clonaje.

- Ensamblaje de los oligonucleótidos mediante PCR.
- Amplificación del producto de PCR.

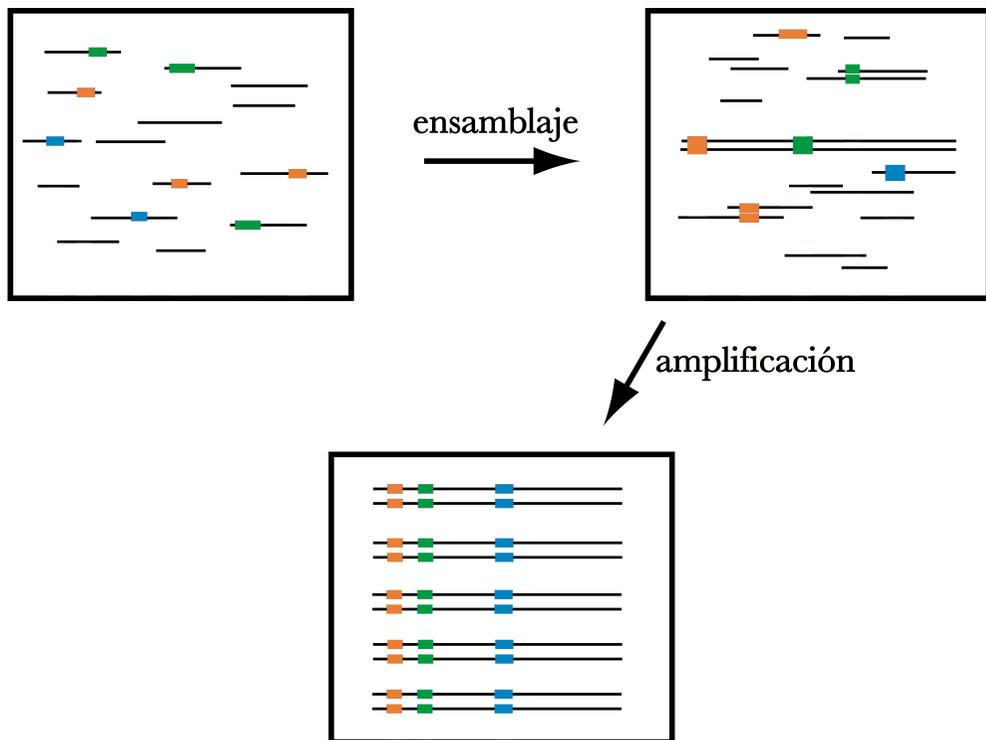
Los cebadores que se utilizaron para la amplificación son los oligonucleótidos del extremo 5' de cada hebra.

- Clonaje.

Se clonó el gen completo (el gen que codifica la proteína con una cola de histidina en el amino terminal) en el vector pQE-80L que contiene el gen de resistencia a ampicilina o en el vector pET24(+) que contiene el gen de resistencia a kanamicina. La expresión de proteína en ambos vectores está bajo el control del operón Lac, por lo tanto se puede inducir su expresión con IPTG.

- Transformación de las células DH10B.

Se transformó el producto del clonaje en las células DH10B y del total de colonias obtenidas se secuenciaron 8 al azar para realizar los test de calidad pertinentes.



*fig.2.4. Esquema del método seguido para la construcción de bibliotecas. Los oligos que codifican ambas hebras del gen completo se ensamblan mediante PCR. En color se muestran las posiciones específicas que contienen las mutaciones. Tras la PCR de montaje, las variantes de longitud completa se amplifican con cebadores exteriores.*

## 2.4. Obtención de proteínas

En el siguiente apartado se describen las etapas necesarias para expresar y purificar tanto variantes únicas de proteínas como bibliotecas combinatoriales de variantes de tiorredoxina.

### **2.4.1. Transformación de células supercompetentes BL21-DE3**

Se utilizaron las células BL21-DE3 de STRATAGENE (Catalogo #200131). El protocolo a seguir es similar al usado para las células supercompetentes XL1-Blue con algunas excepciones: las alícuotas de células a transformar deben ser de 10  $\mu\text{L}$  (un plásmido con una única variante de proteína) o de 50  $\mu\text{L}$  (una genoteca con diferentes variantes de tiorredoxina), se aconseja añadir  $\beta$ -mercaptoetanol hasta una concentración final de 25 mM en cada alícuota de células y tras el choque térmico se añadieron 0.9 mL de medio de cultivo SOC.

### **2.4.2. Pruebas de expresión de proteína**

Las pruebas de expresión de proteínas se realizaron solamente en el caso de trabajar con bibliotecas combinatoriales de mutantes, ya que debido al método de construcción de las bibliotecas (apartado 2.3) pueden darse clones que contengan solo fragmentos del gen que codifica la proteína. De las placas crecidas durante toda la noche se eligieron unas 150 colonias y se inocularon cada una de ellas en 5 mL de medio líquido LB más el antibiótico correspondiente. Se dejaron crecer con agitación a 37°C hasta que su absorbancia a 660 nm fue de 0.8. Se cogieron 700  $\mu\text{L}$  de cultivo crecido, se mezclaron con otros 700  $\mu\text{L}$  de glicerol al 50 % en agua (previamente esterilizado) y se almacenaron a -80°C en tubos de congelación debidamente etiquetados. En el cultivo restante se indujo la expresión de proteína añadiendo IPTG (hasta una concentración final de 40  $\mu\text{M}$ ) y se dejó con agitación a 37°C, tras 9-12h se tomaron 20  $\mu\text{L}$  de cultivo inducido para procesar por electroforesis PAGE-SDS al 15 % de poliacrilamida (fig.2.5).

Los gliceroles de aquellos cultivos que no produjeron proteína se desecharon y de aquellos cultivos que sobre-expresaron proteína con un tamaño similar a la tiorredoxina se extrajo ADN para secuenciarlo (tal y como se describe en el apartado 2.1.4).

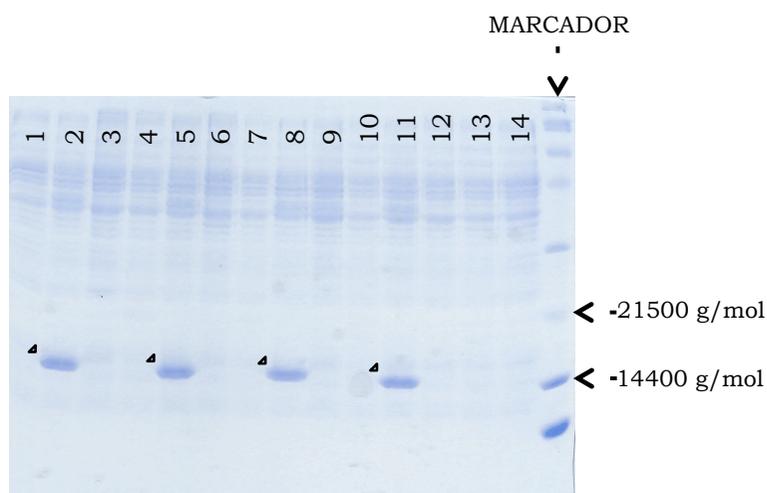


fig.2.5. Gel de poliacrilamida en presencia de SDS que muestra la sobreexpresión de variantes de tiorredoxina ( $P_m = 11990$  g/mol) en las calles 2, 5, 8 y 11.

### 2.4.3. Expresión a gran escala

El inóculo se realizó a partir de una colonia o de un glicerol en 5 mL de LB más el antibiótico correspondiente. Se dejó crecer con agitación a 37°C durante 2-4 h hasta que se alcanzó la fase exponencial del crecimiento bacteriano, aproximadamente una absorbancia de 0.6 a 600 nm. Se transfirieron 5 mL de cultivo a 500 mL de LB (suplementado con el antibiótico correspondiente) en el caso de tiorredoxina, y en el caso de la BMP-2 a 500 mL de TB con 100 µg/mL de ampicilina. En ambos casos se dejó crecer hasta que se alcanzó una absorbancia de 0.6. A dicha absorbancia, se indujo la expresión añadiendo IPTG hasta una concentración de 40 µM y se incubó en agitación a 37°C de 9-12h.

### 2.4.4. Purificación de proteína

Son necesarios varios pasos hasta obtener la proteína a la pureza requerida por las técnicas que posteriormente se utilizarán para su estudio. Se siguieron diferentes protocolos para la purificación de variantes de tiorredoxina dependiendo de que la proteína se expresase con una cola de 6 histidinas en el amino terminal o sin

dicha cola. La purificación de variantes BMP-2 es objeto de patente de la empresa Noricum S.L. por lo que se incluyen solamente las líneas generales del proceso de purificación.

#### **2.4.4.1. Purificación de tiorredoxina de *E. coli* y variantes de tiorredoxina sin cola de histidina**

- Lisis celular.

Los cultivos crecidos durante toda la noche se centrifugaron a 7000 r.p.m. a 4°C durante 15 min (centrífuga Hettich). El sobrenadante se eliminó y el precipitado se resuspendió en 5 mL de tampón 30 mM Tris, 1mM EDTA, pH 8.3 (TE). La disolución resultante se sonicó con objeto de lisar las células. Se mantuvieron los tubos en hielo y se sometieron a 10 pulsos de sonicación durante 20 segundos intercalándose con periodos de 50 segundos de descanso. Para eliminar los restos de paredes celulares, se centrifugó el homogeneizado a 8000 r.p.m. durante 10 minutos a 4°C, en este caso se aprovechó el sobrenadante y se desechó el precipitado.

- Eliminación de ácidos nucleicos

Para eliminar el ADN y ARN presentes se trató el sobrenadante con una disolución de sulfato de estreptomicina. La cantidad de sulfato de estreptomicina a preparar se calculó con la siguiente regla:

$$X=0.25 \times S \qquad \text{ec.2.2}$$

Donde X representa el volumen en mL de disolución de sulfato de estreptomicina y S representa el volumen en mL de disolución de sobrenadante. La disolución de sulfato de estreptomicina al 10% preparada se transfirió a un embudo de decantación de manera que se permitió un goteo constante y lento en el vaso que contenía el sobrenadante. Se mantuvo con agitación a una temperatura de 4°C,

tras 9-12h se centrifugó la disolución a 20000 r.p.m. durante 30 minutos a 4°C. Se descartó el precipitado blanquecino y nos quedamos con el sobrenadante que se filtró (0.45 µm) para eliminar pequeñas partículas.

- Cromatografía de filtración en gel.

La cromatografía de filtración en gel o de exclusión molecular se utiliza para separar la tiorredoxina de otras proteínas de diferente peso molecular y de otros componentes tales como lípidos y polisacáridos presentes en la disolución a tratar. Como soporte se utilizó la resina Sephacryl S-100 de AMERSHAM PHARMACIA BIOTECH AB. Es una matriz de polímeros (poliacrilamida/dextrano) fuertemente hidrofílicos, que al ser puestos en contacto con el agua adquiere el aspecto de un gel. La columna se encuentra acoplada a un FPLC. Una vez equilibrada la columna en TE, se introdujo la muestra a procesar manteniendo un flujo de 5 mL/min, se siguió circulando tampón TE hasta un volumen de 4.5 L aproximadamente. Los componentes que salen de la columna se registraron por detección ultravioleta y se recolectaron en fracciones de 10 mL. La pureza de las fracciones correspondientes al pico de tiorredoxina se evaluaron mediante la realización de SDS-PAGE (*Polyacrylamide Gel Electrophoresis SDS*).

- Eliminación de ácidos nucleicos

Registrando el espectro de absorbancia de la disolución de proteína se puede saber si la muestra está contaminada con restos de ADN ya que los ácidos nucleicos presentan un espectro de absorbancia con un máximo a 260 nm y el espectro de absorbancia de la tiorredoxina presenta un máximo a 280 nm. Se consideró que la muestra de proteína estaba libre de ácidos nucleicos cuando  $A_{280}/A_{260} \geq 1.7$ . Algunas variantes de tiorredoxina, tras la exclusión molecular no cumplían dicha condición. Solo en estos casos se realizó esta última etapa que consiste en dializar la disolución de proteína frente a grandes volúmenes de tampón 20 mM glicina pH=2.0. A bajos valores de pH, el ADN se hidroliza y va

saliendo lentamente por los poros de la bolsa de diálisis, por ello los cambios se realizaron cada largos periodos de tiempo (1-2 días) y frente a grandes volúmenes de tampón (5 L).

#### **2.4.4.2. Purificación de variantes de tiorredoxina con cola de histidina**

- Lisis celular

El proceso es similar al descrito en el apartado 2.4.4.1, solo que en este caso el tampón que se utilizó para resuspender las células es 20 mM Fosfato Sódico, 500 mM NaCl y 20 mM imidazol a pH 7.4 (*binding buffer*). Por cada 50 mL del tampón previamente se disolvió una pastilla de cóctel de antiproteasas libre de EDTA.

- Cromatografía de afinidad

Se utilizaron columnas de níquel His GraviTrap de GE Healthcare. La proteína se expresa con una cola de 6 histidinas en el amino terminal. Al pasar el sobrenadante del lisado celular por la columna, previamente equilibrada con *binding buffer*, la proteína queda unida a la resina por la formación de un complejo entre el Níquel y las histidinas. Se lavó la resina con la misma disolución tampón y finalmente se eluyó la muestra con 20 mM Fosfato Sódico, 500 mM NaCl y 500 mM Imidazol a pH 7.4 (*elution buffer*). El imidazol al unirse al níquel con mayor afinidad que las histidinas desplaza a la proteína. La pureza de las fracciones se evaluó mediante la realización de SDS-PAGE.

#### **2.4.4.3. Purificación de BMP-2 silvestre y variantes de BMP-2**

Tras la lisis celular la proteína se encontraba formando cuerpos de inclusión que se resuspendieron a altas concentraciones de urea. Una vez disueltos, se centrifugaron y el sobrenadante se procesó por cromatografía de intercambio catiónico con la resina SP Sepharose® Fast Flow de Pharmacia Biotech. Tras la

elución podemos observar mediante la realización de SDS - PAGE que las primeras fracciones eluidas contienen proteína monomérica, las siguientes, proteína dimérica y las últimas agregados de alto peso molecular (fig.2.6).

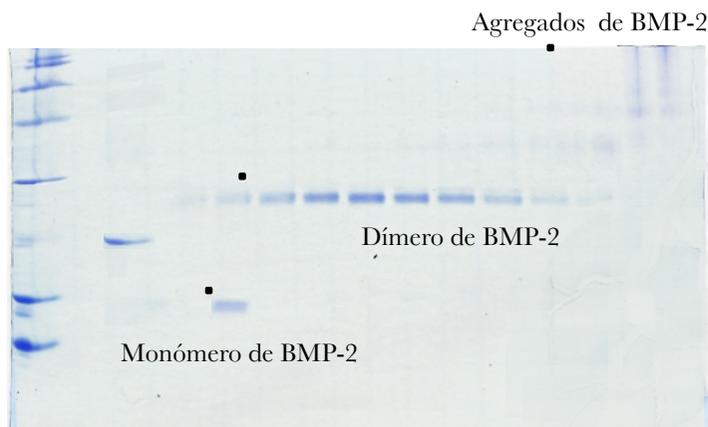


fig.2.6. Gel de poliacrilamida en presencia de SDS que muestra las eluciones de la cromatografía de intercambio catiónico. Las fracciones dimericas fueron las utilizadas para la siguiente fase de replegamiento *in vitro*.

Esto ocurre porque la proteína BMP-2 no se pliega correctamente en un sistema de expresión bacteriano y necesita de una etapa de replegamiento *in vitro*. Se descartaron las primeras y las últimas fracciones y se sometió a replegamiento solamente la fracción dimerica de proteína. Finalmente, se purificó la proteína correctamente plegada con varias etapas de cromatografía; una primera cromatografía de interacción hidrofóbica (con la resina Phenyl Sepharosa 6 Fast Flow de GE Healthcare) y una segunda cromatografía de afinidad por heparina (con la resina Heparin Sepharose™ 6 Fast Flow de GE Healthcare).

## 2.5. Preparación de las muestras

### 2.5.1. Preparación de tampones

La preparación de los tampones se realizó siguiendo el procedimiento general que consiste en preparar dos disoluciones de igual concentración de tampón en su

forma ácida y básica. Ambas disoluciones se mezclaron hasta alcanzar el pH deseado. En el caso de disponer solamente una de las formas del tampón, se preparó la disolución a la concentración deseada y se ajustó el pH con HCl o NaOH concentrado.

### **2.5.2. Preparación de diálisis**

Una vez que teníamos la proteína disuelta en el tampón correspondiente a la última etapa del proceso de purificación se le cambió el tampón mediante diálisis. Así, se tomó la cantidad de proteína deseada y con ayuda de una micropipeta se introdujo en la bolsa de diálisis previamente hidratada. La bolsa se cerró con pinzas de diálisis y se introdujo en un vaso de precipitado que contenía un volumen de tampón al menos 600 veces mayor que el de la muestra. Tras 8h a 4°C con agitación suave, se cambió el tampón que bañaba la bolsa de diálisis por tampón fresco, este procedimiento se repitió hasta tres veces. Con objeto de eliminar cualquier residuo, una vez se extrajo la disolución de proteína en el nuevo tampón, se centrifugó a 15000 r.p.m. a 4°C durante 15 minutos en una centrifuga de mesa.

### **2.5.3. Medida de concentración de proteína**

La concentración de proteína en disolución se determinó espectrofotométricamente registrando el espectro ultravioleta-visible y midiendo la absorbancia a 280 nm. Atendiendo a la ley de Lambert-Beer:

$$A = \epsilon \cdot b \cdot c \quad \text{ec.2.3}$$

Siendo A la Absorbancia,  $\epsilon$  el coeficiente de extinción molar de la proteína, b el paso de luz (en centímetros) y c la concentración molar de la sustancia absorbente en el medio.

El coeficiente de extinción molar que se utilizó para la tiorredoxina silvestre de *E. coli* es de  $14000 \text{ M}^{-1} \text{ cm}^{-1}$  a 280 nm [2]. Se recalculó el coeficiente de extinción molar para cada variante de tiorredoxina mediante la herramienta web: <http://www.expasy.ch/tools/protparam.html>.

Para la BMP-2 silvestre el coeficiente de extinción molar que se usó es  $37775 \text{ M}^{-1} \text{ cm}^{-1}$  a 280 nm, para las formas mutantes se recalculó el coeficiente de extinción molar con la herramienta web anteriormente nombrada.

#### **2.5.4. Preparación de disoluciones de guanidina**

Se prepararon dos disoluciones con alta concentración de desnaturalizante ( $\approx 8\text{M}$ ) e igual concentración de tampón en su forma ácida y básica cada una. Ambas disoluciones se mezclaron hasta alcanzar la lectura de pH deseada y así obtuvimos una disolución madre a alta concentración de guanidina. Disoluciones a concentraciones menores de guanidina se obtuvieron mezclando esta disolución madre con una disolución tampón sin guanidina. La concentración de guanidina en disolución se determinó mediante medidas de índice de refracción [3]. La diferencia entre el índice de refracción de la disolución de guanidina y el correspondiente tampón acuoso es  $\Delta n$ , que está relacionado con la concentración de guanidina (M) por la siguiente expresión[4]:

$$[\text{GdnClH}] = 57.147 \cdot (\Delta n) + 38.68 \cdot (\Delta n)^2 - 91.60 \cdot (\Delta n)^3 \quad \text{ec.2.4}$$

## **2.6. Calorimetría diferencial de barrido**

### **2.6.1. Introducción**

La calorimetría diferencial de barrido (DSC, *Differential Scanning Calorimetry*) es una técnica muy poderosa para la caracterización energética de los cambios conformacionales inducidos por la temperatura en proteínas y en otras

macromoléculas biológicas [5-15]. En un experimento de calorimetría se calienta la proteína en disolución que se encuentra en la célula de muestra a una velocidad constante, al mismo tiempo se registra la energía que hay que aplicar a la célula en relación con la requerida para aumentar la temperatura de la célula de referencia (contiene el mismo tampón que la disolución de la proteína, sin proteína) a la misma velocidad. Es decir, se mide el calor absorbido o emitido por una proteína en disolución a medida que se va calentando. Así, podemos obtener la capacidad calorífica molar aparente de una proteína como función de la temperatura. El posterior análisis de esta dependencia nos dará una caracterización termodinámica completa de la transición que ha tenido lugar por efecto de la temperatura. Si estudiamos dicha dependencia a diferentes velocidades de calentamiento (velocidad de barrido) podemos obtener también información cinética sobre el proceso ocurrido inducido por la temperatura.

En general, si analizamos una traza de DSC según la termodinámica de equilibrio podemos obtener:

- La capacidad calorífica parcial absoluta de una proteína
- El conjunto de parámetros termodinámicos (cambio de entalpía  $\Delta H$ , cambio de entropía  $\Delta S$ , cambio en la capacidad calorífica ( $\Delta C_p$ ) asociados a la transición.
- La función de partición y por lo tanto la población de estados intermedios y sus parámetros termodinámicos.

### **2.6.2. Instrumentación**

Los experimentos se realizaron con un microcalorímetro VP capilar de *microcal*. Las características más notables de este equipo son: la inyección de muestra está totalmente automatizada, alcanza velocidades de barrido de hasta 250 K/h y opera en un amplio rango de temperaturas (desde 0°C hasta 140°C). Para evitar

la ebullición de las muestras, el equipo está conectado a una salida de nitrógeno que suministra sobrepresión. En fig.2.7 se muestra un esquema del núcleo térmico del calorímetro [16]. Las células están construidas de tantalio y se sitúan dentro de una camisa adiabática de plata. Tienen forma helicoidal (con un volumen de trabajo de 125  $\mu\text{L}$  y un diámetro interior de 1.5 mm) y los orificios de entrada y salida de muestra son diferentes, de manera que las células se limpian por desplazamiento. Los tubos de entrada y salida de ambas células tienen un diámetro interior de 0.8 mm y pasan a través de la chaqueta adiabática hasta la válvula de distribución que está fuera del calorímetro. Las células están en contacto con una serie de sensores de temperatura (de telurio de bismuto) que recogen la diferencia de temperatura de ambas células y proporcionan una señal de voltaje  $\Delta T - 1$ . Esta señal amplificada sirve como alimentación para los calentadores auxiliares (situados en la célula de referencia o en la de muestra) que neutralizan la diferencia de temperatura entre ambas células. Entre la chaqueta adiabática y las células hay una termopila que produce una señal de voltaje  $\Delta T - 2$ . De nuevo, esta señal amplificada sirve para alimentar a un dispositivo Peltier que enfriará o calentará la chaqueta para anular la diferencia de temperatura. Estos circuitos de retroalimentación están continuamente funcionando, de manera que siempre se mantienen a la misma temperatura ambas células y también las células y la chaqueta. A medida que las células aumentan la temperatura según la velocidad de barrido, la chaqueta va calentándose al mismo ritmo controlado por el circuito de retroalimentación  $\Delta T - 2$ . La diferencia de voltaje  $\Delta T - 1$  se va registrando a cada temperatura, siendo esta la señal experimental de salida del instrumento.

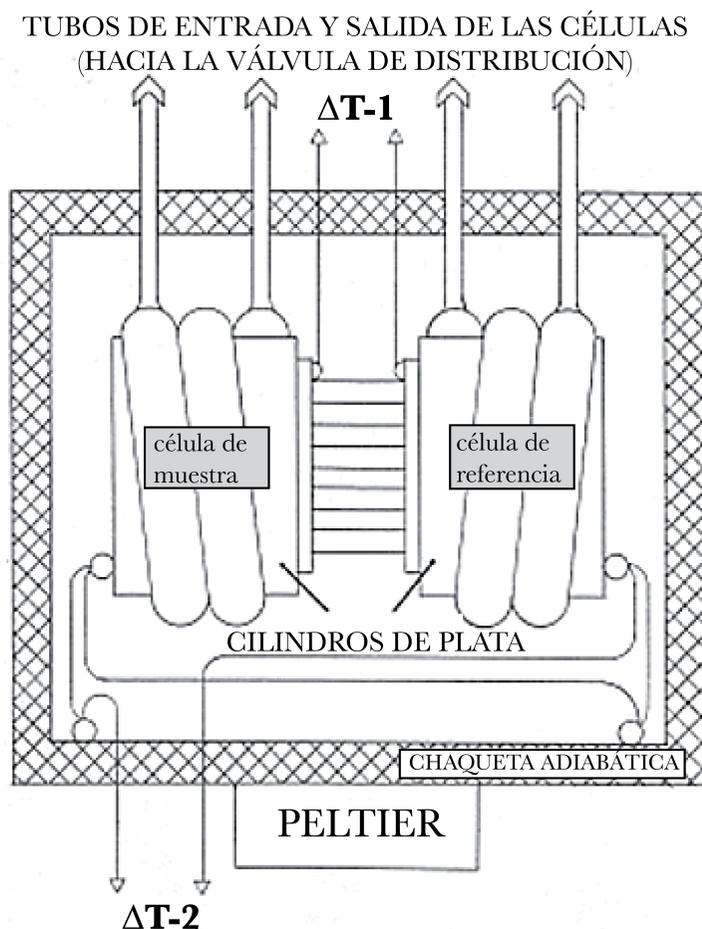


fig.2.7. Diagrama esquemático del núcleo térmico del calorímetro capilar VP que muestra las células helicoidales con los tubos de entrada y salida, los cilindros de plata, la chaqueta adiabática, el dispositivo Peltier que sirve para calentar o refrigerar la camisa y los sensores de temperatura entre ambas células que proporcionan una señal  $\Delta T-1$ .

### 2.6.3. Experimento calorimétrico

#### 2.6.3.1. Preparación de la muestra

La muestra se prepara tal y como se describe en el apartado 2.5. Se debe ser riguroso en el trato de la muestra para que se pueda asegurar que el tampón con el que rellenar la célula de referencia es exactamente el mismo que el tampón en

el que está disuelta la proteína, que la pureza de la muestra es  $\geq 90\%$  y que la concentración se ha determinado de manera precisa.

### **2.6.3.2. La línea base instrumental**

El objeto de registrar un barrido con ambas células (la de referencia y la de muestra) llenas de tampón, es determinar y corregir las diferencias instrumentales entre ambas células. Si idealmente ambas células fueran iguales no haría falta realizar este barrido. Para obtener una línea base precisa es necesario hacer varios barridos tampón-tampón hasta que sean reproducibles, es decir, hasta que diferencia entre líneas sucesivas sea menor a  $0.5 \mu\text{cal}/^\circ\text{C}$ . Esta línea base se le restará al barrido proteína-tampón para obtener el calor que se pone en juego debido exclusivamente al cambio conformacional inducido por la temperatura en la proteína y eliminar la contribución debido a las diferencias instrumentales de ambas células.

### **2.6.3.3. Barrido de la muestra**

Una vez equilibrado el instrumento (obtenemos líneas bases reproducibles), pasamos a registrar el termograma correspondiente a la proteína en estudio. Para ello, se llena la célula de muestra con la disolución de proteína y la de referencia con tampón fresco (en el microcalorímetro capilar, este proceso está completamente automatizado). En nuestro caso, un perfil típico de DSC muestra un pico o transición, que se atribuye a la absorción de calor asociada a la desnaturalización de la proteína, mientras que los valores de la señal antes y después de la transición, reflejan la capacidad calorífica parcial de los estados nativo y desnaturalizado respectivamente. Es posible obtener la capacidad calorífica parcial absoluta de la proteína a partir de la diferencia de capacidad calorífica entre muestra y referencia y teniendo el volumen molar parcial de la proteína[7]. En fig.2.8 se indican alguno de los parámetros más significativos en una transición calorimétrica.

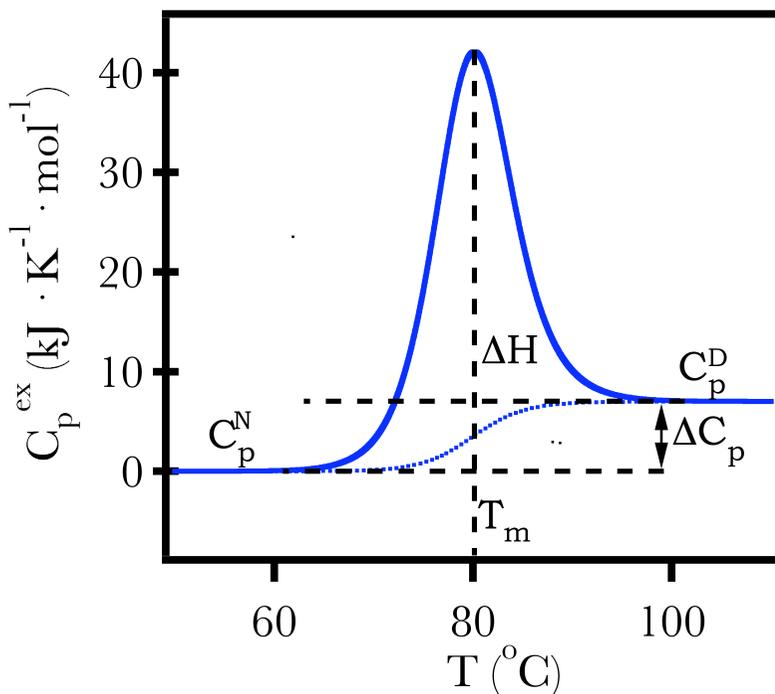


fig.2.8. Curva de capacidad calorífica (en azul con trazo continuo) en función de la temperatura calculada para una hipotética proteína que despliega según un modelo de equilibrio de dos estados.  $C_p^N$  representa la capacidad calorífica del estado nativo y  $C_p^D$  la del estado desplegado. En azul y trazo punteado se representa la línea base química (línea que conecta de un modo suave los valores de capacidad calorífica pre- y post-transicionales). Al restar la línea base química a los valores de capacidad calorífica obtenemos la capacidad calorífica de exceso. El área encerrada entre los valores de capacidad calorífica y la línea base química representa la entalpía de desnaturalización  $\Delta H$ . Los parámetros utilizados para la simulación son:  $\Delta C_p = C_p^D - C_p^N = 7 \text{ kJ} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$ ,  $\Delta H = 400 \text{ kJ} \cdot \text{mol}^{-1}$  y  $T_m = 80^\circ\text{C}$ .

#### 2.6.3.4. Segundo barrido de la muestra

Generalmente es necesario someter a un segundo barrido la proteína para comprobar si la señal calorimétrica se recupera significativamente. Este segundo barrido, nos indica la reversibilidad del proceso, información necesaria a la hora de analizar e interpretar los datos.

#### 2.6.4. Análisis de las trazas de DSC

Como se ha indicado en el apartado 2.6.2, la señal de registro del instrumento es la diferencia de voltaje que se aplica a ambas células calorimétricas al ir calentándose, en función de la temperatura. Los ficheros tal y como los presenta el programa MicroCal Origin para DSC ( $\text{cal min}^{-1}$ ) requieren de una manipulación para poder ser tratados. El mismo programa ofrece la posibilidad de realizar el tratamiento de forma automática, resumidamente:

- corrección de la línea base instrumental

Al barrido proteína-tampón se le resta la línea base instrumental (el barrido tampón-tampón).

- normalización y conversión de unidades

Para ello usa el factor de conversión f:

$$f = \frac{4.1868 \left(\frac{\text{J}}{\text{cal}}\right) \cdot 10^{-3} (\text{kJ/J})}{V_{\text{cel}} (\text{mL}) \cdot c (\text{mol/L}) \cdot v (\text{K/min}) \cdot 10^{-3} (\text{L/mL})} \quad \text{ec.2.5}$$

De esta manera los datos se presentan finalmente con unidades de  $\text{kJ} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$ .

- corrección dinámica

Como todo instrumento de medida, el calorímetro tiene un tiempo de respuesta y por ello las curvas obtenidas se encuentran ligeramente distorsionadas. Esta distorsión será tanto mayor cuanto mayor sea la velocidad a la que se calienta la muestra (velocidad de barrido). Para corregir los termogramas se realiza la siguiente operación:

$$C_p(T) = C_p^*(T) + \tau \cdot v \cdot \frac{dC_p^*(t)}{dt} \quad \text{ec.2.6}$$

donde  $C_p(T)$  es la señal real producida por el sistema en estudio,  $C_p^*(T)$  la obtenida experimentalmente y que se encuentra distorsionada por la cinética del instrumento,  $v$  la velocidad de barrido y  $\tau$  es el tiempo de respuesta del instrumento que previamente se ha determinado en el proceso de caracterización dinámica del mismo [17].

El siguiente paso es el análisis de los datos conforme a un modelo teórico propuesto *a priori*. El modelo más adecuado será el más simple que explique correctamente los datos experimentales. Para elegir dicho modelo hay que analizar los siguientes criterios:

- Criterio de reversibilidad. Bajo ciertas condiciones experimentales, la desnaturalización de la proteína es reversible, recuperándose en gran medida la señal calorimétrica de la proteína nativa tras enfriar la disolución [5, 6]. En estos casos, suponiendo que en todo momento existe equilibrio entre los diferentes estados poblados de la proteína, se puede aplicar la formulación basada en la termodinámica de equilibrio [8, 18].
- Análisis del efecto de concentración de proteína. Resulta indicativo de procesos de oligomerización de alguno de los estados que se pueblan durante el desplegamiento.
- Efecto de velocidad de barrido. Indican la presencia de procesos cinéticos.

Una vez dispongamos de esta información experimental, será posible elegir un modelo teórico para el análisis de los datos como los que se describen a continuación.

#### **2.6.4.1. Análisis según la termodinámica de equilibrio**

Una vez comprobado que un sistema es caloriméricamente reversible se intenta ajustar al modelo termodinámico más sencillo posible: el modelo de equilibrio de dos estados.

### 2.6.4.1.1. Modelo de equilibrio de dos estados

Es aquel en el que solo se encuentran poblados de forma significativa los estados nativo (N) y desplegado o desnaturalizado (D) de la proteína:



Las proporciones relativas de dichos estados a una temperatura dada están determinadas por el valor de la constante de equilibrio (K) a esa temperatura:

$$K = \frac{[D]}{[N]} \quad \text{ec.2.8}$$

En general, los cambio en un parámetro termodinámico producidos por la desnaturalización, se definen como las diferencias entre los valores de dicho parámetro en el estado desnaturalizado y el nativo.

$$\Delta_N^D J = J(D) - J(N) \quad \text{ec.2.9}$$

Donde J puede ser la energía de Gibbs (G), entalpía (H), entropía (S), capacidad calorífica a presión constante ( $C_p$ ), etc. Esta definición implica que se está tomando como referencia el estado nativo. Así, la entalpía de exceso y la capacidad calorífica de exceso vienen dadas por:

$$\langle \Delta H \rangle = x_D \cdot \Delta_N^D H = \frac{K}{1+K} \cdot \Delta_N^D H \quad \text{ec.2.10}$$

$$C_p^{\text{ex}} = \frac{\partial \langle \Delta H \rangle}{\partial T} = \frac{(\Delta_N^D H)^2}{RT^2} \cdot \frac{K}{(1+K)^2} + x_D \cdot \Delta_N^D C_p \quad \text{ec.2.11}$$

Donde  $x_D = \frac{K}{1+K}$  es la fracción de proteína que se encuentra en el estado desnaturalizado. Los corchetes en las ecuaciones anteriores indican el valor promedio de todos los estados de la proteína a este parámetro termodinámico.

Para expresar la dependencia con la temperatura de  $K$  y de  $\Delta_N^D H$  se han utilizado las ecuaciones de van't Hoff y de Kirchoff:

$$\frac{\partial \ln K}{\partial T} = \frac{\Delta_N^D H}{RT^2} \quad \text{ec.2.12}$$

$$\frac{\partial \Delta_N^D H}{\partial T} = \Delta_N^D C_p \quad \text{ec.2.13}$$

La temperatura de desnaturalización ( $T_m$ ) se define como la temperatura a la cual  $x_D=0.5$  y por lo tanto  $K=1$  y  $\Delta_N^D G=0$ . Las transiciones predichas son prácticamente simétricas, de manera que el valor de  $T_m$  está próximo al de la temperatura correspondiente al máximo en capacidad calorífica.

En la ecuación ec.2.11 vemos que la función capacidad calorífica de exceso consta de dos términos. El primero de ellos, se debe al cambio inducido por la temperatura en el equilibrio de desnaturalización (forma de pico), mientras que el segundo término representa la capacidad calorífica promedio de la proteína y corresponde a la línea base química (con forma sigmoidal). El área encerrada entre la transición y la línea base química es igual a cambio total de entalpía en el proceso de desnaturalización y se conoce como entalpía calorimétrica  $\Delta_N^D H$  (fig.2.8). Esta depende de la temperatura según la ecuación de Kirchoff. Si la transición es simétrica y estrecha se puede despreciar esta dependencia dentro del rango estrecho de la transición y se asigna dicha entalpía a la  $T_m$ , denominándose  $\Delta H_m$ . Como a esta temperatura el cambio de energía de Gibbs es cero, el cambio de entropía a la  $T_m$  se puede calcular con:

$$\Delta S_m = \frac{\Delta H_m}{T_m} \quad \text{ec.2.14}$$

En principio, el cambio de capacidad calorífica en el proceso de desnaturalización también se puede determinar desde el perfil de DSC, y por lo tanto los valores de  $\Delta_N^D G$ ,  $\Delta_N^D H$ , y  $\Delta_N^D S$  a cualquier temperatura  $T$  se pueden calcular con las ecuaciones termodinámicas estándar:

$$\Delta_N^D H(T) = \Delta_N^D H(T_m) + \int_{T_m}^T \Delta_N^D C_p \cdot dT \quad \text{ec.2.15}$$

$$\Delta_N^D S(T) = \frac{\Delta_N^D H(T_m)}{T_m} + \int_{T_m}^T \frac{\Delta_N^D C_p}{T} \cdot dT \quad \text{ec.2.16}$$

$$\Delta_N^D G(T) = \Delta_N^D H(T) - T \cdot \Delta_N^D S(T) \quad \text{ec.2.17}$$

Por supuesto todo este formalismo solo es aplicable en el caso de que el modelo de equilibrio de dos estados describa adecuadamente el proceso de desnaturalización en estudio. La aplicabilidad de dicho modelo se puede determinar a partir de un análisis directo de la transición de DSC. Si realizamos simulaciones dando diferentes valores a la entalpía de desnaturalización en la ecuación ec.2.11 veremos que esta determina el área bajo la transición, así como su anchura. La entalpía de desnaturalización calculada a partir de la anchura de la transición se conoce como entalpía aparente de van't Hoff ( $\Delta H^{vH}$ ). Podemos obtener una expresión para  $\Delta H^{vH}$ , considerando  $T=T_m$  (y por tanto  $K=1$ ) en la ecuación ec.2.11, teniendo en cuenta solo el término debido al desplazamiento del equilibrio de desnaturalización:

$$\Delta H^{vH} = 4RT^2 \cdot \left[ \frac{\Delta C_{p,m}}{\Delta_N^D H_m} \right] \quad \text{ec.2.18}$$

Donde  $\Delta C_{p,m}$  es la capacidad calorífica de exceso a la temperatura  $T_m$  medida sobre la línea base química. El cociente  $\Delta C_{p,m} / \Delta_N^D H_m$  no depende de la cantidad

de proteína presente en la célula del calorímetro, pero sí es sensible a la anchura de la transición. Como no puede haber dos cambios de entalpía asociados al mismo proceso y la ecuación ec.2.18 se ha obtenido suponiendo el modelo de dos estados, en el caso de que se cumpla, las entalpías calorimétricas y de van't Hoff deben ser iguales. Por tanto, si existen diferencias significativas entre  $\Delta H^{vH}$  y  $\Delta_N^D H_m$  se concluye que el proceso de desnaturalización no sigue el modelo de dos estados. Se define el parámetro r:

$$r = \frac{\Delta_N^D H_m}{\Delta H^{vH}} \quad \text{ec.2.19}$$

Si  $r=1$  se puede considerar una baja proporción de estados intermedios y por lo tanto, que se cumple el modelo de dos estados. Privalov [5] y colaboradores han encontrado una relación  $r=1.05 \pm 0.03$  para diversas proteínas globulares pequeñas, indicando una buena concordancia con el modelo de dos estados

Si  $r > 1$  se debe considerar la presencia de estados intermedios en la desnaturalización térmica de la proteína en estudio. En este caso la transición de DSC será más ancha de lo predicha por el modelo de dos estados ya que  $\Delta H^{vH} < \Delta_N^D H_m$ . Un mecanismo general para la desnaturalización térmica de proteínas en equilibrio fue propuesto por Freire y Biltonen [19], sugiriendo que el estado nativo ( $I_0$ ) pasa al estado desnaturalizado ( $I_n$ ) a través de  $n-1$  estados intermedios.

Si  $r < 1$  indica que existe cooperatividad intermolecular, es decir, la unidad cooperativa no es un monómero sino un dímero u otro estado de agregación superior al que se suponía inicialmente. Un modelo de dos estados en el que una proteína multimérica ( $N\mu$ ) desnaturaliza concomitantemente con un proceso de disociación en monómeros ha sido analizado en la literatura [20]. La transición de

DSC predicha por este modelo es asimétrica y se desplaza a mayor temperatura a medida que aumenta la concentración de la proteína.

#### **2.6.4.1.2. Estabilidad termodinámica de proteínas. Curva de estabilidad**

La estabilidad de una proteína se define con frecuencia como la diferencia entre la energía de Gibbs del estado nativo y la del estado desnaturalizado, la dependencia de  $\Delta G_N^D$  con la temperatura se conoce como curva de estabilidad. Si el proceso de desplegamiento sigue un modelo de equilibrio de dos estados, la curva de estabilidad puede calcularse a partir de la traza de DSC usando las ecuaciones ec.2.15, ec.2.16 y ec.2.17. Para calcular la estabilidad de una proteína se requiere un valor fiable de  $\Delta_N^D C_p$ , que se puede calcular directamente a partir de la traza de DSC (fig.2.8) [21, 22] o a partir de la pendiente resultante al representar  $\Delta_N^D H(T_m)$  versus  $T_m$  para varias transiciones de DSC con diferentes valores de  $T_m$ . Para ello, normalmente se registran varios perfiles de DSC de la proteína disuelta en tampones a diferentes valores de pH y se asume que los valores de  $\Delta_N^D H$  y  $\Delta_N^D C_p$ , no dependen del pH. Independientemente del método que se use para calcular  $\Delta_N^D C_p$ , este valor resulta ser positivo y dependiente de la temperatura. Valores de  $\Delta_N^D C_p$  obtenidos a partir de trazas de DSC en un amplio rango de temperaturas, revelan que su valor disminuye a altas temperaturas y se predice su anulación (por extrapolación) a temperaturas muy superiores a 100°C [23-26]. Sin embargo, no se comete mucho error si se ignora la dependencia con la temperatura entre valores de 0 y 80°C [23, 24].

Si se considera el cambio de capacidad calorífica constante con la temperatura las ecuaciones ec.2.15, ec.2.16 y ec.2.17 pasarían a ser:

$$\Delta_N^D H(T) = \Delta H_m(T_m) + \Delta_N^D C_p \cdot (T - T_m) \quad \text{ec.2.20}$$

$$\Delta_N^D S(T) = \frac{\Delta H_m(T_m)}{T_m} + \Delta_N^D C_p \cdot \ln\left(\frac{T}{T_m}\right) \quad \text{ec.2.21}$$

$$\Delta_N^D G(T) = \Delta H_m(T_m) \cdot \left[1 - \frac{T}{T_m}\right] + \Delta_N^D C_p \cdot \left[T - T_m - T \cdot \ln\left(\frac{T}{T_m}\right)\right] \quad \text{ec.2.22}$$

En fig.2.9 [8] se muestra un ejemplo de curva de estabilidad y a continuación se comentan las propiedades fundamentales más significativas [21, 27]:

- La curvatura viene dada por  $\partial^2 \Delta_N^D G / \partial T^2 = -\Delta_N^D C_p / T$ , que tiene siempre valores negativos ya que el cambio de capacidad calorífica de desnaturalización siempre es positivo.
- La pendiente de la curva es  $-\partial \Delta_N^D G / \partial T = \Delta_N^D S$  y presenta un único máximo a la temperatura  $T_S$ , a la que  $\Delta_N^D S = 0$  [8].
- La temperatura  $T_H$  a la que la entalpía de desnaturalización es cero, es ligeramente inferior a  $T_S$ .
- La proteína es estable en el rango de temperatura en que  $\Delta_N^D G > 0$ . Según se puede ver en fig.2.9 la curva de estabilidad cruza la línea de  $\Delta_N^D G = 0$  en dos puntos diferentes, siendo el punto de corte a más alta temperatura la correspondiente a la temperatura de desnaturalización  $T_m$ , usada en el cálculo de la curva de estabilidad. El punto de corte a más baja temperatura es el resultado de la extrapolación de la curva e indica que es posible desnaturalizar la proteína también por frío. Esta idea fue predicha por Brandts [28] y demostrada experimentalmente con posteriores estudios [29]. En la mayoría de los casos la desnaturalización fría se predice para temperaturas menores o cercanas a  $0^\circ\text{C}$  y se detecta experimentalmente si se sobreenfría la disolución de proteína o por adición de urea o guanidina, con objeto de desplazar dicha temperatura hacia

valores superiores. Estudios de DSC [24, 30-35] sugieren que la desnaturalización fría es una propiedad común a las proteínas globulares, confirmando así las características ya mencionadas de la curva de estabilidad.

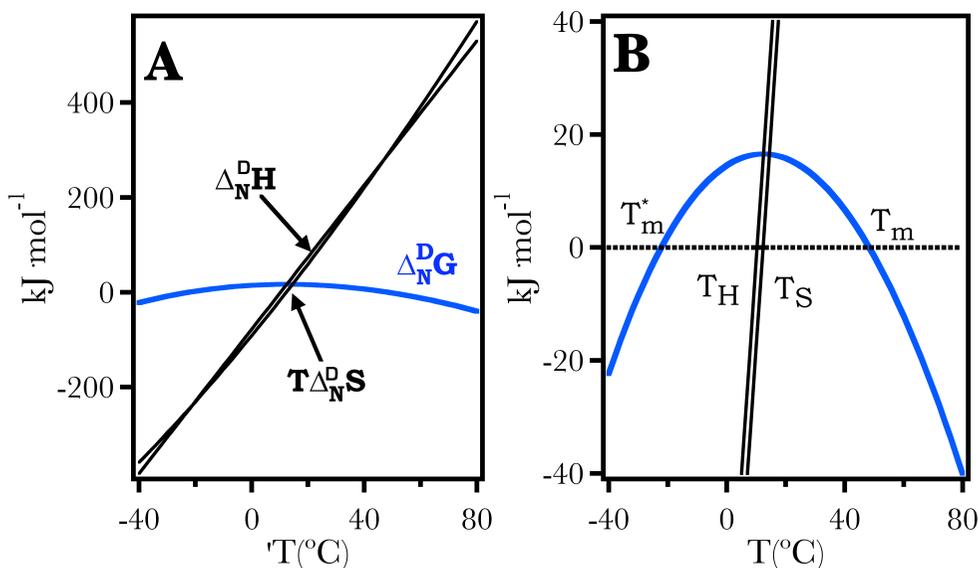


fig.2.9. A) Efecto de la temperatura sobre los cambios de entalpía, entropía y energía de Gibbs para la desnaturalización de una proteína hipotética. B) Dependencia de  $\Delta G_N^D$  con la temperatura que muestra las principales características de la curva de estabilidad de la proteína. Se indican las temperaturas correspondientes a la desnaturalización caliente ( $T_m$ ) y fría ( $T_m^*$ ), así como las temperaturas de inversión de la entalpía ( $T_H$ ) y entropía ( $T_S$ ) a las que  $\Delta H$  y  $\Delta S$  se hacen cero, respectivamente. Los valores de  $\Delta H$ ,  $\Delta S$  y  $\Delta G$  se calcularon usando las ecuaciones ec.2.20, ec.2.21 y ec.2.22 y los valores de  $\Delta C_p = 7.6 \text{ kJ} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$ ;  $\Delta H = 290 \text{ kJ} \cdot \text{mol}^{-1}$  y  $T_m = 48.3^{\circ}\text{C}$ .

#### 2.6.4.2. Análisis cinético de las trazas de DSC

Si la cinética de plegamiento-desplegamiento es lenta o si el desplegamiento ocurre de manera irreversible, las concentraciones de los estados significativamente poblados a cada temperatura dependen del tiempo que se tarda en alcanzar dicha temperatura. Como consecuencia, la transición calorimétrica depende de la velocidad de barrido a la que se realice el experimento.

### 2.6.4.2.1. El efecto del equilibrio lento

En algunos casos, las trazas de DSC reversibles muestran un marcado efecto de velocidad de barrido [10, 17, 36, 37]. Esta dependencia con la velocidad de barrido se atribuye al hecho de que la cinética del proceso de desnaturalización es lenta y por ello la concentración de los estados significativamente poblados durante el desplegamiento se aleja de sus valores de equilibrio. Para analizar estos casos, se utiliza un enfoque análogo al descrito para realizar la corrección dinámica de las trazas de DSC (apartado 2.6.4.). Así, la capacidad calorífica de exceso para el equilibrio ( $C_p^{ex}(eq)$ ) y la capacidad calorífica de exceso ( $C_p^{ex}$ ) determinada a una velocidad de barrido ( $v$ ), se relacionan mediante la ecuación[17]:

$$C_p^{ex}(eq) = C_p^{ex} + \tau_c \cdot \left( \frac{dC_p^{ex}}{dt} \right) = C_p^{ex} + v \tau_c \cdot \left( \frac{dC_p^{ex}}{dT} \right) \quad \text{ec.2.23}$$

donde  $\tau_c$  es el tiempo de relajación (análogo al tiempo de respuesta en la ecuación de la corrección dinámica),  $C_p^{ex}$  es un dato experimental y  $dC_p^{ex}/dT$  se obtiene a partir de dos o más termogramas realizados a diferentes velocidad de barrido. Nótese que  $C_p^{ex}(eq)$  resulta de la extrapolación a velocidad de barrido nula y que se puede obtener también el perfil de  $\tau_c$  a cada temperatura mediante este mismo procedimiento. Más tarde [10] se redefinió un nuevo tiempo de relajación ( $\tau_H$ ) para la función de la entalpía:

$$\langle \Delta H \rangle_{eq} = \langle \Delta H \rangle + \tau_H \cdot \left( \frac{d\langle \Delta H \rangle}{dt} \right) = \langle \Delta H \rangle + v \cdot \tau_H \cdot C_p^{ex} \quad \text{ec.2.24}$$

donde  $\langle \Delta H \rangle_{eq}$ , es la entalpía de exceso del equilibrio y  $\langle \Delta H \rangle$  es la entalpía de exceso calculada experimentalmente a una velocidad de barrido  $v$ . Nótese que el tiempo de relajación definido para la ecuación ec.2.23 ( $\tau_c$ ) y el definido para la

ecuación ec.2.24 ( $\tau_H$ ), no son iguales, y que la ecuación ec.2.24 no es el resultado de derivar con respecto a la temperatura la ecuación ec.2.23, ya que los tiempos de relajación dependen de la temperatura. Una cuestión importante, que no se ha abordado aún, es la relación entre ambos tiempos de relajación y la constante cinética de la desnaturalización. A continuación se describe dicha relación basada en el tiempo de relajación en función de la entalpía:  $\tau_H$ . Un mecanismo cinético simple, consistente con el modelo de dos estados, se puede describir como:



donde  $k_f$  y  $k_r$  son las constantes de primer orden para los mecanismos directos e inversos respectivamente. La constante de desnaturalización de equilibrio viene dada por  $K = k_f/k_r$  y por lo tanto la ecuación de velocidad para el mecanismo de desnaturalización será:

$$\frac{dx_D}{dt} = k_f x_N - k_r x_D \quad \text{ec.2.26}$$

como  $x_N + x_D = 1$  y  $K = k_f/k_r$ , la ecuación se puede describir como:

$$\frac{dx_D}{dt} = (k_f + k_r) \cdot (x_D^{eq} - x_D) \quad \text{ec.2.27}$$

donde  $x_D^{eq} = K/(K+1)$  es la fracción de proteína desnaturalizada cuando se alcanza el equilibrio a una temperatura dada (el perfil de  $x_D^{eq}$  versus  $T$ , sería el observado en un experimento de DSC a una velocidad de barrido infinitamente lenta). Si se considera un experimento llevado a cabo a una temperatura constante la ecuación ec.2.27 se puede reescribir de manera que incluya la relación de  $x_D$  con el tiempo, resultando una ecuación exponencial con un tiempo de relajación

de  $(k_f+k_r)^{-1}$ . Como en un experimento de DSC la temperatura varía con el tiempo de acuerdo con la velocidad de barrido, la ecuación finalmente quedaría:

$$\frac{dx_D}{dT} = \left(\frac{1}{U}\right) \cdot (k_f+k_r) \cdot (x_D^{eq}-x_D) \quad \text{ec.2.28}$$

Siendo  $k_f$ ,  $k_r$  y  $x_D^{eq}$  dependientes de la temperatura. Si se mide el exceso de capacidad calorífica desde la línea base química, entonces:  $x_D = \langle \Delta H \rangle / \Delta_D^N H$ ,  $x_D^{eq} = \langle \Delta H \rangle_{eq} / \Delta_D^N H$  y  $dx_D/dT = C_p^{ex} / \Delta_D^N H$ . Consecuentemente la ecuación ec.2.28 se reescribe como:

$$\langle \Delta H \rangle_{eq} = \langle \Delta H \rangle + \frac{U}{k_f+k_r} \cdot C_p^{ex} \quad \text{ec.2.29}$$

Si comparamos las ecuaciones ec.2.29 y ec.2.24, podemos ver que ambas son idénticas donde:

$$\tau_H = (k_f+k_r)^{-1} \quad \text{ec.2.30}$$

De manera que  $\tau_H$  (y no  $\tau_c$ ) resulta ser el tiempo de relajación del proceso de desnaturalización que estamos considerando.

#### 2.6.4.2.2. El efecto de la irreversibilidad

Muchas proteínas desnaturalizan de manera irreversible, mostrando un segundo barrido de DSC sin efecto térmico. Generalmente, esta irreversibilidad se atribuye a alteraciones (como autólisis, agregación o alteraciones químicas de los residuos) que bloquean a la proteína en un estado final incapacitándola para recuperar su estructura nativa. Con frecuencia transiciones irreversibles se analizan aplicando modelos de equilibrio termodinámico ya que a veces se puede considerar que estas alteraciones irreversibles ocurren tan lentamente que apenas se apreciaban en el corto periodo de tiempo en el que ocurre la transición. Otras veces, también se

pueden aplicar modelos de equilibrio termodinámico porque las alteraciones irreversibles ocurren a mayor temperatura que el pico de la transición, por tanto, aunque el segundo barrido muestre una transición sin efecto térmico, a temperaturas cercanas al pico de la transición los estados más poblados no se alejan de sus valores de equilibrio. Sin embargo, otros trabajos muestran transiciones de DSC distorsionadas por la cinética de alteraciones irreversibles. En algunos casos se muestran trazas que resultan irreversibles y con un gran efecto de velocidad de barrido [9, 38-43] y en otros casos se muestra la correlación entre el efecto térmico medido por DSC y la existencia de desnaturalización irreversible medida con otras técnicas [40, 41, 44]. De manera que en estos casos, las trazas de DSC irreversibles se pueden analizar según modelos cinéticos como:

- El modelo de dos estados irreversibles que considera que la proteína despliega según un proceso cinético de primer orden y que los dos estados significativamente poblados durante el desplegamiento son el estado nativo (N) y el estado desplegado (D), además supone que este proceso ocurre de manera irreversible de acuerdo con el esquema:



Donde  $k_{ap}$  es la constante de velocidad de primer orden para el proceso de desnaturalización irreversible.

- El modelo de Lumry-Eyring [45]. Es un modelo menos simple, pero más realista que el modelo de dos estados irreversibles. Incluiría dos etapas: etapa 1, desnaturalización reversible del estado nativo (N) para llegar a otro estado (D) desnaturalizado o parcialmente desnaturalizado y la etapa 2, alteración irreversible del estado (D) para dar lugar a un estado final (F). Este modelo se esquematiza como:

$$N \overset{K}{\leftrightarrow} D \overset{k}{\rightarrow} F \quad \text{ec.2.32}$$

### 2.6.4.3. Determinación de las capacidades caloríficas absolutas de los estados nativo y desnaturalizado a partir de los termogramas experimentales

En un experimento de DSC, al restar al barrido de proteína la línea base tampón-tampón, obtenemos la capacidad calorífica aparente  $\Delta C_{\text{app}}$ . La capacidad calorífica del agua es mayor que la de la proteína, con lo cual a mayor concentración de proteína menor será el valor de  $\Delta C_{\text{app}}$ . Siguiendo el proceso descrito por Kholodenko y Freire [46], es posible determinar valores de capacidad calorífica absoluta a partir de la dependencia de  $\Delta C_{\text{app}}$  con la concentración de proteína:

$$\Delta C_{\text{app}} = C \cdot V_0 \cdot 10^{-6} \cdot \left( C_{p.P} - \frac{V_P}{V_W} \cdot C_{p.W} \right) \quad \text{ec.2.33}$$

donde  $C$  es la concentración de proteína expresada en mM,  $V_0$ , es el volumen de la célula del calorímetro en mL,  $C_{p.P}$  y  $C_{p.W}$  son las capacidades caloríficas de la proteína y el agua (tampón), respectivamente y,  $V_P$  y  $V_W$  son los volúmenes molares de la proteína y el agua [47]. El volumen específico de la proteína se puede calcular de acuerdo a Makhatadze y Privalov [48]. Representando  $\Delta C_{\text{app}}$  frente a la concentración de proteína a una temperatura dada (de acuerdo a la ec.2.33), de la pendiente del ajuste lineal forzando a la recta a pasar por el origen podemos obtener el valor de la capacidad calorífica absoluta a cada temperatura. Este procedimiento requiere la realización de varios termogramas a varias concentraciones de proteína, con lo cual se minimiza el error asociado al valor de capacidad calorífica y además nos permite una estimación de dicho error.

## 2.7. Dicroísmo circular

### 2.7.1. Introducción

El fenómeno en el que la absorción de luz es diferente para diferentes direcciones de polarización se denomina dicroísmo. En la luz polarizada linealmente, la dirección del vector eléctrico es constante y la magnitud es modulada; por contra, en la luz polarizada circularmente, la magnitud del vector eléctrico es constante y la dirección es modulada (fig.2.10).

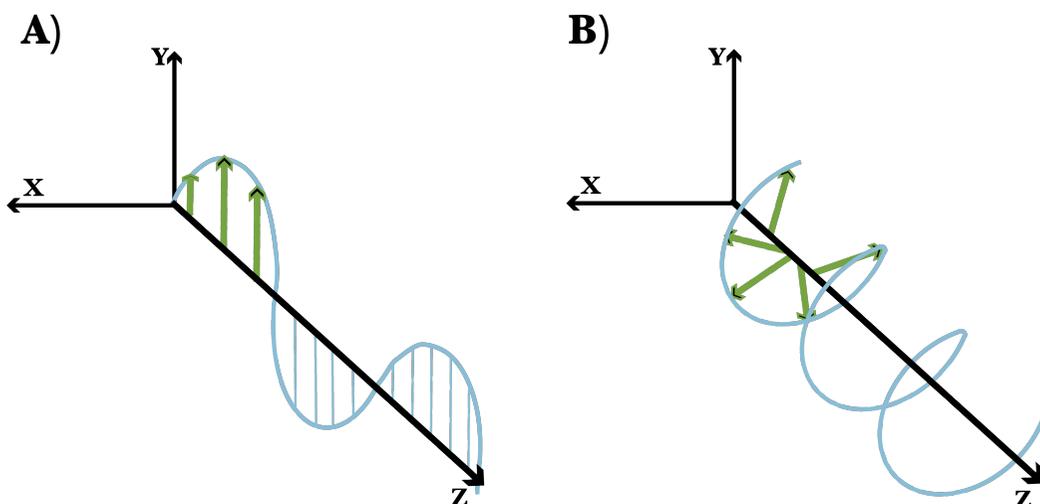


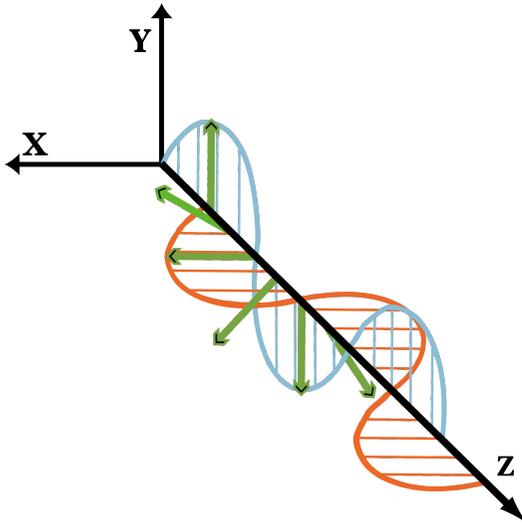
fig.2.10. Vector eléctrico en la luz polarizada linealmente (A) y en la luz polarizada circularmente (B).

La luz polarizada circularmente se consigue mediante la interferencia de dos haces de luz ortogonales, polarizados linealmente y desfasados entre sí  $90^\circ$  (fig.2.11). El vector eléctrico de la luz polarizada circularmente se puede expresar matemáticamente como [49, 50]:

$$E_{\pm} = E_0(i \pm j) \exp(2\pi i \cdot (vt - z/\lambda)) \quad \text{ec.2.34}$$

Donde + indica luz circular polarizada a la derecha y - a la izquierda,  $E_0$  es la amplitud de la onda,  $i$  y  $j$  son los vectores unidad en las coordenadas  $x$  e  $y$

respectivamente de un sistema de coordenadas cartesiano hacia la derecha en el que la  $+z$  es la dirección de propagación,  $i$  es  $(-1)^{1/2}$  y  $\nu$  y  $\lambda$  son la frecuencia y la longitud de onda de la luz.



*fig.2.11. En verde se muestra el vector eléctrico de la luz polarizada circularmente que resulta de la interferencia de dos haces de luz ortogonales, polarizados linealmente y desfasados entre sí  $90^\circ$ . En azul se muestra el haz polarizado en la dirección del eje  $y$  y en rojo el haz de luz polarizado en la dirección del eje  $x$ .*

La hélice que describe el vector eléctrico de la luz polarizada circularmente puede girar hacia la derecha (lpcd) o hacia la izquierda (lpci). El dicroísmo circular (CD, *Circular Dichroism*) se basa en la propiedad que tienen las moléculas quirales de interaccionar de diferente manera con la luz polarizada circularmente hacia la derecha y hacia la izquierda en ausencia de campos magnéticos [51] y se define como:

$$\Delta A = A_i - A_d \quad \text{ec.2.35}$$

Siendo  $A_i$  la absorción de la luz polarizada circularmente hacia la izquierda:

$$A_i = \log_{10} \left( \frac{I_i^0}{I_i} \right) = \epsilon_i c l \quad \text{ec.2.36}$$

Y  $A_d$  hacia la derecha:

$$A_d = \log_{10} \left( \frac{I_d^0}{I_d} \right) = \epsilon_d c l \quad \text{ec.2.37}$$

Donde  $I^0$  e  $I$  son las intensidades de luz polarizada que inciden y que salen de la muestra respectivamente,  $l$  es la distancia que recorre la luz a través de la muestra,  $c$  es la concentración de soluto quiral y  $\epsilon$  es el coeficiente de extinción molar del soluto para la luz polarizada hacia el lado correspondiente. En consecuencia, según la definición de dicroísmo circular (ec.2.35):

$$\Delta A = A_i - A_d = \epsilon_i c l - \epsilon_d c l = (\epsilon_i - \epsilon_d) \cdot c l = \Delta \epsilon \cdot c l \quad \text{ec.2.38}$$

Donde las unidades de  $\Delta \epsilon$  son  $M^{-1} \cdot \text{cm}^{-1}$ . Además de expresar el dicroísmo circular en forma de  $\Delta \epsilon$ , en bioquímica generalmente se utiliza la elipticidad ( $\theta$ ), que es el método de medida original propuesto por Lowry [52]. Este se basa en que cuando una luz polarizada circularmente pasa a través de un medio circularmente dicroico, al producirse la absorción diferencial de las dos componentes circulares el haz de salida resulta elípticamente polarizado (fig.2.12).

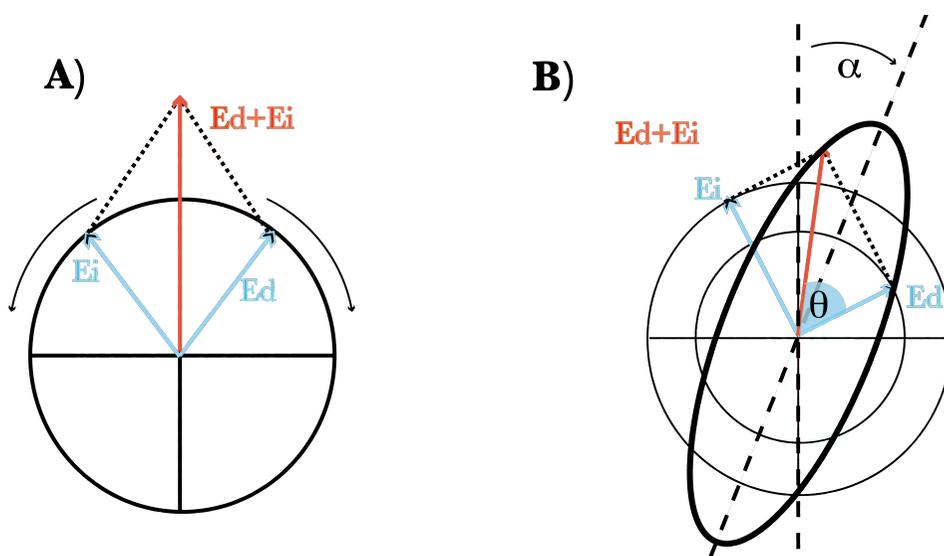


fig.2.12. A) La luz polarizada circularmente, se obtiene por la interferencia de dos haces de luz polarizados linealmente, cuyos vectores se representan en azul:  $E_i$  y  $E_d$ . B) Tras la interacción con la muestra quiral, al producirse una absorción diferencial de la luz polarizada a izquierdas y a derechas, el haz de luz resulta polarizado elípticamente y el plano de polarización rota un ángulo  $\alpha$ .

Se denomina elipticidad a la relación entre los ejes mayor y menor de la elipse, que es la tangente del ángulo  $\theta$  (fig.2.12). Cuando los vectores eléctricos de las componentes circulares van en el mismo sentido, la suma de sus componentes definen el eje mayor de la elipse y cuando van en sentidos opuestos, la resta de sus magnitudes definen el valor del eje menor de la elipse :

$$\theta(\text{rad}) = \tan(\theta) = \frac{|E_i| - |E_d|}{|E_i| + |E_d|} = \frac{\exp(-A_i/2) - \exp(-A_d/2)}{\exp(-A_i/2) + \exp(-A_d/2)} \quad \text{ec.2.39}$$

Expandiendo exponenciales, teniendo en cuenta el valor despreciable de  $\Delta A$  frente a la unidad y pasando a grados, se obtiene que la elipticidad es proporcional al dicroísmo circular.:

$$\theta(\text{grad}) = \frac{180 \cdot \ln 10 \cdot \Delta A}{4\pi} = 32,98 \cdot \Delta A \quad \text{ec.2.40}$$

La elipticidad molar se define como:

$$[\theta] = \frac{\theta}{10 \cdot c \cdot l} \quad \text{ec.2.41}$$

Y la elipticidad molar normalizada como:

$$[\theta]_{\text{MRW}} = \frac{\theta \cdot M_r}{10 \cdot c \cdot l \cdot N_A} \quad \text{ec.2.42}$$

Donde  $\theta$  se introduce en mrad (datos crudos que proporciona el espectropolarímetro),  $M_r$  es la masa molecular y  $N_A$  el número de aminoácidos;  $[\theta]$  y  $[\theta]_{\text{MRW}}$  se expresan en  $\text{grados} \cdot \text{dmol}^{-1} \cdot \text{cm}^2$ . La concentración se introduce en  $\text{mg/mL}$ .

### 2.7.2. Dicroísmo circular de proteínas

Mediante espectroscopía de dicroísmo circular se puede obtener mucha información acerca de las propiedades conformacionales de proteínas, más concretamente:

- Determinar si una proteína se encuentra plegada y en tal caso, se puede caracterizar su estructura secundaria (ultravioleta lejano) y su estructura terciaria (ultravioleta lejano).
- Comparar la estructura de diferentes mutantes de una misma proteína.
- Estudiar la estabilidad conformacional de una proteína a diferentes temperaturas, en diferentes condiciones de pH o en presencia de agentes desnaturalizantes.
- Determinar los cambios conformacionales asociados a interacciones entre moléculas asimétricas, ya sean interacciones proteína-proteína, proteína-ligando, proteína-DNA o DNA-ligando.

#### 2.7.2.1. Dicroísmo circular y la estructura secundaria de las proteínas

La estructura secundaria de las proteínas se puede determinar mediante CD en el ultravioleta lejano (190-250 nm). A estas longitudes de onda el principal cromóforo es el enlace peptídico y aunque el enlace por sí solo no es una estructura quiral, la orientación espacial de los grupos peptídicos adyacentes confieren asimetría al enlace y por tanto, hace que presente actividad óptica. Cada elemento principal de la estructura secundaria de las proteínas (hélices  $\alpha$ , lámina  $\beta$  y estructura no regular) presenta un espectro de dicroísmo circular característico (fig.2.13).

- El espectro de CD de una proteína con estructura de hélice  $\alpha$  presenta un doble mínimo a 222 y a 208-210 nm y un máximo más intenso a 191-193 nm (fig.2.13). La intensidad de las tres bandas, que depende de los ángulos diedros  $\phi$  y  $\Psi$ , da una idea de la cantidad de contenido helicoidal que tiene la proteína.
- Las proteínas con estructura  $\beta$  suelen presentar un único pico negativo y un único pico positivo cuyas intensidades son menores que las de las hélices. El mínimo se encuentra normalmente entre 210 y 225 nm (fig.2.13) aunque se han descrito algunos ejemplos [53, 54] donde proteínas, que por RMN se han determinado que son todo lámina  $\beta$ , presentan un banda negativa intensa en torno a los 200 nm similar a la de los péptidos con estructura no regular.
- Los péptido y proteínas con estructura no regular (bucles de las conformaciones nativas y, en general, conformaciones desplegadas) presentan habitualmente una banda negativa fuerte cerca de 200 nm (fig.2.13) y algunas bandas débiles entre 220 y 230 nm que pueden ser tanto positivas como negativas.

En el UV lejano también absorben los aminoácidos aromáticos y, aunque sus cadenas laterales no son quirales, se comportan como tales cuando así lo propicia su entorno. Con esta salvedad, la actividad óptica de los enlaces peptídicos es de gran utilidad para conocer el contenido de estructura secundaria de las proteínas, teniendo cada proteína un espectro de UV lejano que es aproximadamente una combinación lineal de los espectros de las estructuras secundarias y no regulares que contienen.

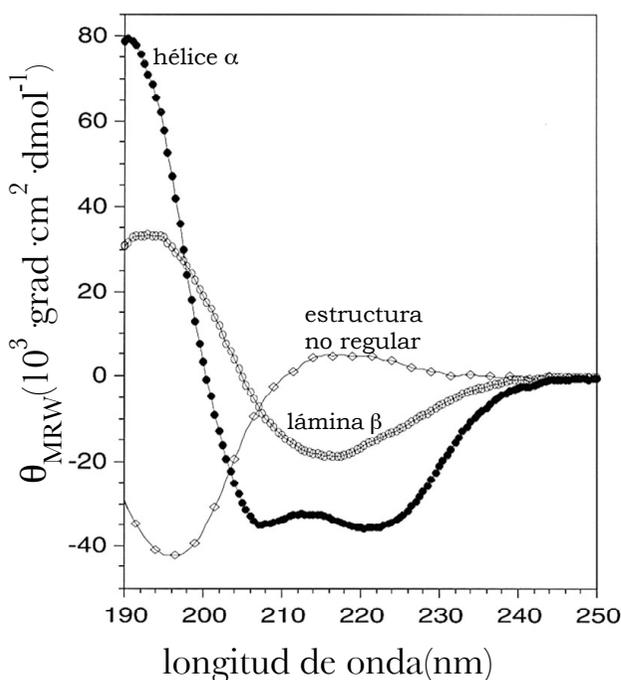


fig.2.13. Espectros de ultravioleta lejano de CD de diferentes conformaciones de un péptido de poli-lisina: en círculos rellenos con estructura de  $\alpha$ -hélice, en círculos con estructura de lámina  $\beta$  y en rombos con estructura no regular [55].

### 2.7.2.2. Dicroísmo circular y estructura terciaria de las proteínas

El espectro ultravioleta cercano (250-350 nm) de una proteína es sensible a cambios en la estructura terciaria. A estas longitudes de onda los cromóforos son los aminoácidos aromáticos y los enlaces disulfuro cuando se encuentran inmovilizados en un entorno asimétrico, lo que normalmente ocurre en la estructura nativa.

- La contribución de la fenilalanina al espectro de dicroísmo circular es pequeña debido a su bajo coeficiente de extinción, además es poco sensible a las alteraciones de su ambiente. Presenta una banda de absorción en la región 250-270 nm.

- La tirosina tiene una banda de absorción más intensa a 276 nm y un hombro a 283 nm; esta banda puede desplazarse a mayores longitudes de onda por la formación de puentes de hidrógeno con el grupo fenólico.
- El triptófano presenta una banda de absorción más intensa centrada a 282 nm, que también puede desplazarse a mayores longitudes de onda por la formación de un enlace de hidrógeno con el grupo indol.

### **2.7.3. Determinación del grado de estructura secundaria y terciaria de la tiorredoxina de *E. coli* seguido por dicroísmo circular**

Los experimentos se llevaron a cabo en un espectropolarímetro JASCO modelo J-715 equipado con un sistema de calentamiento PTC-348WI. La proteína se dializó en tampón HEPES 50 mM, pH 7.0 y se prepararon alícuotas a 50 mM HEPES, pH 7.0 y diferentes concentraciones de guanidina (descrito en el apartado 2.5.4). Una vez preparada la muestra como se describe en el apartado 2.5, se registró el espectro de la proteína en disolución y el espectro del tampón en que está disuelta la proteína. El blanco se substrajo a cada espectro correspondiente con el fin de eliminar las contribuciones instrumentales y del disolvente al espectro de CD.

Se registraron los espectros de CD en el UV lejano y cercano de tiorredoxina silvestre a 2°C y a 98°C con objeto de observar las diferencias estructurales a ambas temperaturas. Los experimentos en el ultravioleta lejano se registraron desde 260 hasta 200 nm, se utilizaron cubetas de 1 mm de paso de luz y concentraciones de muestras de 0.5 mg/mL. En el ultravioleta cercano se registraron desde 325 hasta 260 nm, se utilizaron cubetas de 5 mm de paso de luz y concentraciones de muestra de 1 mg/mL.

La máxima diferencia de señal entre los espectros realizados a 2°C y 98°C se registró a 222 nm (ultravioleta lejano) y 280 nm (ultravioleta cercano). Se realizaron barridos de temperatura en un rango de 2°C a 98°C a una velocidad de calentamiento de 90°C/h a 222 nm (ultravioleta lejano) y a 280 (ultravioleta cercano).

## **2.8. Fluorescencia**

### **2.8.1. Introducción**

La luminiscencia es el fenómeno mediante el cual los electrones que se encuentran en un estado excitado se desactivan emitiendo luz. Dependiendo de la naturaleza del estado excitado podemos hablar de fluorescencia o fosforescencia [56]. Fluorescencia es la emisión de luz desde un estado excitado singlete hasta el estado fundamental en el que la relajación se produce sin cambio en la multiplicidad del electrón. La velocidad de emisión de fluorescencia suele ser de  $10^8 \text{ s}^{-1}$  con lo cual, el tiempo de vida media ( $\tau$ ) de la fluorescencia suele ser del orden de 10 nanosegundos.

Fosforescencia es la emisión de luz desde un estado excitado triplete, en el que la relajación implica cambio de multiplicidad del electrón. La fosforescencia es un fenómeno más lento que la fluorescencia, con una velocidad de emisión de  $10^3$  a  $10^0 \text{ s}^{-1}$ , con lo que el tiempo de vida media de la fosforescencia suele ser del orden de los milisegundos (fig.2.14). La lentitud del proceso hace que otros fenómenos de desactivación compitan con la fosforescencia y por ello no se suele observar.

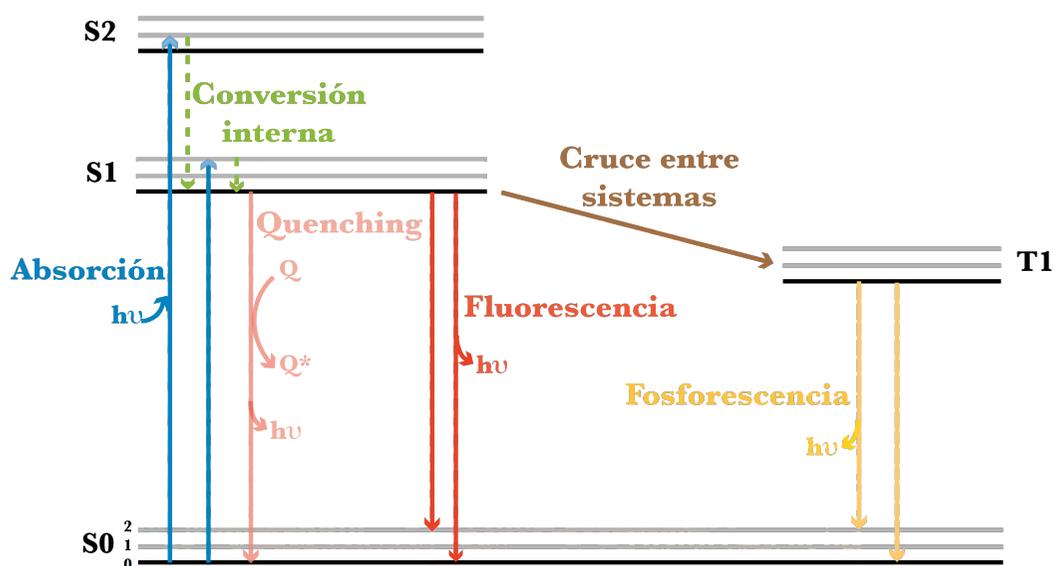


fig.2.14. Diagrama de Jablonski. Los estados singlete fundamental, el primer estado singlete excitado y el segundo estado singlete excitado se representan como  $S_0$ ,  $S_1$  y  $S_2$ , en cada uno de estos niveles están representados los niveles vibracionales 0, 1 y 2. Tras la absorción de luz generalmente ocurren varios procesos. Un fluoróforo normalmente es excitado a un estado vibracional alto de  $S_1$  o  $S_2$ , rápidamente se produce la relajación al nivel vibracional más bajo de  $S_1$  mediante conversión interna. La fluorescencia se produce con la relajación desde  $S_1$  hasta  $S_0$  emitiéndose luz. La Intensidad de fluorescencia puede decaer por la colisión del fluoróforo excitado con otras moléculas presentes en disolución (quenching). Las moléculas en el estado  $S_1$  también pueden pasar al primer nivel triplete  $T_1$ , mediante cruce de sistemas, proceso por el cual el electrón cambia su multiplicidad. La fosforescencia se produce con la relajación desde  $T_1$  hasta  $S_0$  emitiéndose luz de mayor longitud de onda de la absorbida (menos energética).

Analizando el diagrama de Jablonski (fig.2.14) se pueden observar algunas de las características de la fluorescencia:

- Desplazamiento de Stokes. La emisión de fluorescencia es generalmente menos energética que la absorción, con lo cual hay un desplazamiento hacia mayores longitudes de onda del espectro de emisión con respecto al espectro de absorción.
- La emisión de fluorescencia es generalmente independiente de la longitud de onda de excitación. Porque aunque la molécula se excite hasta niveles

vibracionales altos o hasta un nivel singlete  $S_2$ , rápidamente se produce la relajación hasta el nivel singlete  $S_1$ .

- Rendimiento cuántico. El rendimiento cuántico es el número de fotones que se emiten en relación al número de fotones absorbidos y podrá adquirir valores de 0 a 1. 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.
- *Quenching* de la fluorescencia. La disminución de la intensidad de fluorescencia se denomina *quenching*; generalmente, se debe a la desactivación del fluoróforo excitado por colisión con otras moléculas presentes en disolución. A estas moléculas se les denomina *quencher* y el oxígeno molecular disuelto en el tampón suele comportarse como tal.

### 2.8.2. Fluorescencia de proteínas

En proteínas, tres aminoácidos aromáticos presentan fluorescencia: fenilalanina, tirosina y triptófano (fig.2.15). Además, la intensidad de fluorescencia de estos aminoácidos es muy dependiente del entorno en el que se encuentren, resultando la fluorescencia muy sensible a cambios conformacionales, asociaciones, uniones a ligandos o la desnaturalización de la proteína.

La tiorredoxina silvestre de *E. coli* presenta dos triptófanos en las posiciones 28 y 31, dos tirosinas en las posiciones 49 y 70 y cuatro fenilalaninas en las posiciones 12, 27, 81 y 102. Los espectros de excitación de la tiorredoxina nativa y desplegada presentan un máximo centrado a 276 nm y un hombro alrededor de 290 nm (fig.2.16 A). Los espectros de emisión de fluorescencia de tiorredoxina nativa y desplegada, se han registrado excitando a 276 nm y presentan un máximo muy ancho a 344 nm característico de los triptófanos (fig.2.16 B). Sin embargo, a altas concentraciones de agente desnaturalizante el máximo de

emisión se desplaza de 344 a 351 nm; estos cambios reflejan en gran medida la separación estérica del triptófano 28 del puente disulfuro existente entre las cisteínas 32 y 35, razón por la cual disminuye el efecto *quenching* que existe entre ambos y aumenta la intensidad de fluorescencia [57, 58]. Este análisis ha permitido observar que la máxima diferencia entre el espectro de emisión de tiorredoxina nativa y desplegada ocurre a 350 nm (fig.2.16 B), por lo que se ha fijado como la longitud de onda a la que se registran las medidas de emisión de fluorescencia.

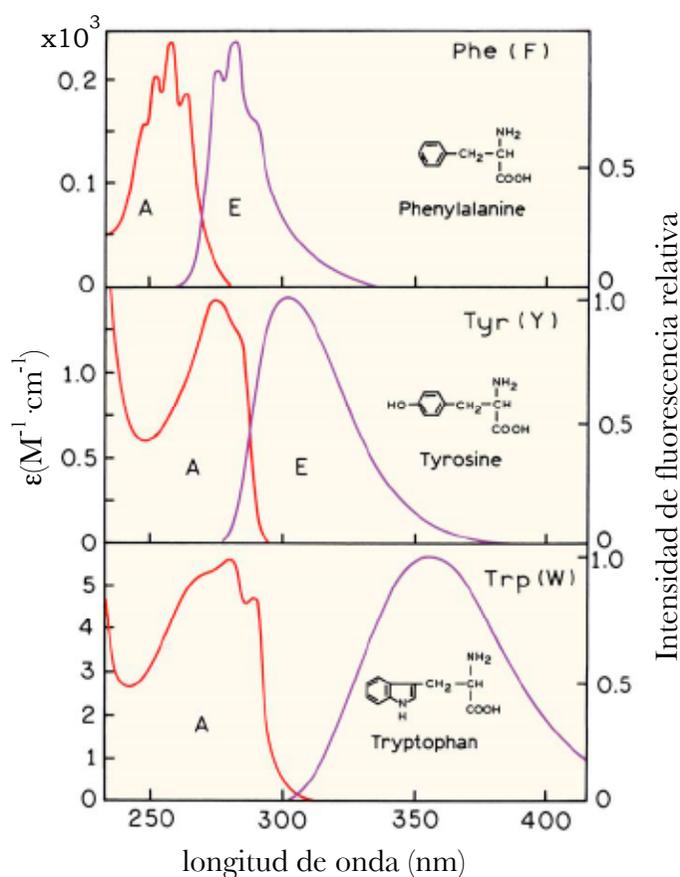


fig.2.15. Espectros de absorción (A) y emisión (E) de los aminoácidos aromáticos a pH 7 en disolución acuosa [56].

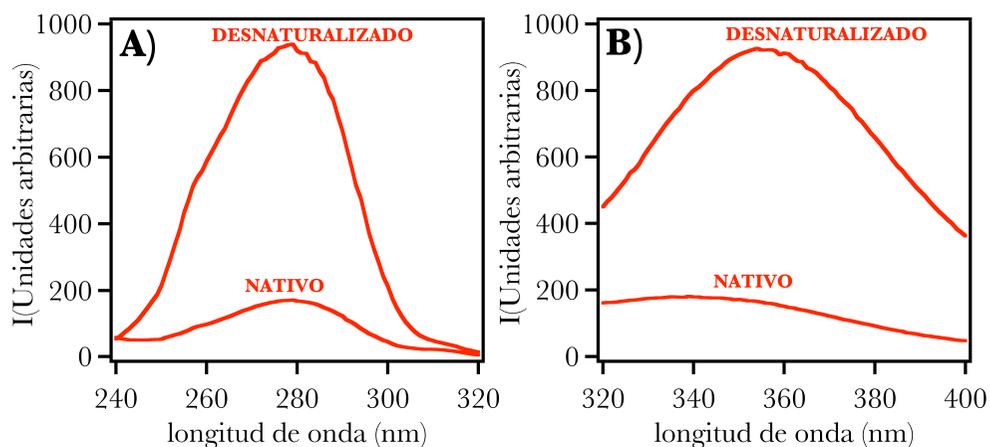


fig.2.16. A) Espectro de excitación de fluorescencia a 25°C de la tiorredoxina de *E. coli* en su estado nativo y desnaturalizado. B) Espectro de emisión de fluorescencia a 25°C de tiorredoxina de *E. coli* en su estado nativo y desnaturalizado. La proteína nativa está disuelta en tampón HEPES 50 mM pH 7, mientras que la proteína desplegada se encuentra en guanidina 7M, HEPES 50 mM pH 7. La concentración de proteína es de 0.03 mg/mL en ambos casos. La velocidad de barrido seleccionada es de 600 nm/min y las rendijas de excitación y emisión son de 10 nm. Se ha utilizado una cubeta de cuarzo de 1 cm de paso de luz. La longitud de onda de excitación seleccionada ha sido de 276 nm; se puede observar como la máxima diferencia entre los dos espectros correspondientes a la proteína nativa y desplegada se produce a 350 nm.

Es importante establecer el rango en el que la intensidad de emisión de fluorescencia cambia linealmente con la concentración de proteína en disolución. Este aspecto es importante ya que interesa que la relación señal/ruido sea máxima en un rango lineal para evitar efectos no deseados como el de filtro interno (parte de las moléculas pueden absorber a las longitudes de onda en las que se emite radiación). Para la tiorredoxina silvestre de *E. coli* se ha encontrado una dependencia lineal en el rango de 0.01 a 0.06 mg/mL. También se deben fijar los valores óptimos de la anchura de las rendijas de excitación y emisión; es deseable, si la calidad de la señal lo permite, que la rendija de excitación sea estrecha para evitar la fotodescomposición de la muestra.

Para la tiorredoxina ancestral de LPBCA, los espectros de excitación y emisión (fig.2.17) son similares a los de la tiorredoxina de *E. coli*. La tiorredoxina de

LPBCA posee la misma cantidad de triptófanos y tirosinas que la tiorredoxina de *E. coli* en posiciones equivalentes, en cuanto a la fenilalanina posee tres en las posiciones 12, 27 y 81 (una menos que la tiorredoxina de *E. coli*). Sin embargo, al ser la emisión del triptófano mucho más intensa que la de la tirosina y esta a su vez más intensa que la de la fenilalanina, el espectro de fluorescencia de la tiorredoxina de *E. coli* y de LPBCA está dominado por la emisión de los triptófanos. Razón por la cual ambas proteínas poseen espectros de emisión de fluorescencia muy parecidos en su estado nativo y desnaturalizado. Se fijaron, por tanto, para la tiorredoxina de LPBCA las mismas condiciones de longitud de onda de excitación (276 nm), de apertura de rendijas de excitación y emisión (10 nm) y rango de concentración lineal (0.01-0.6 mg/mL) que para la tiorredoxina de *E. coli*.

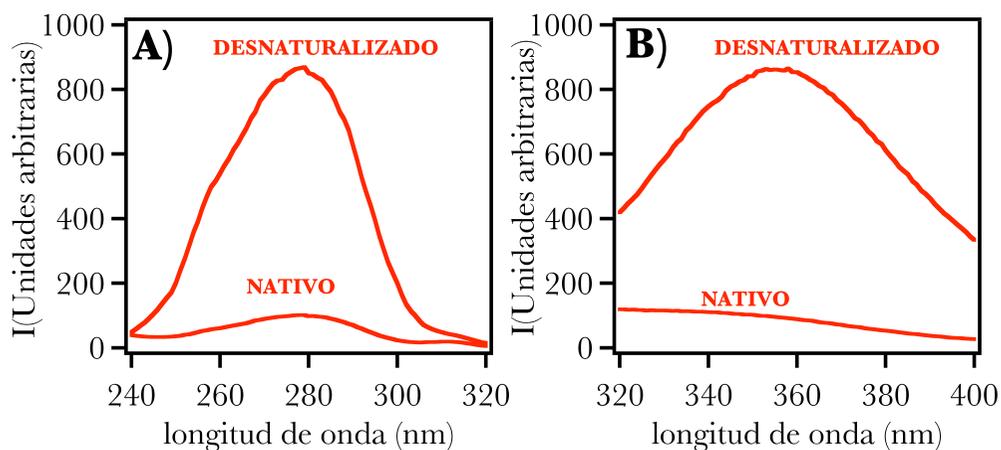


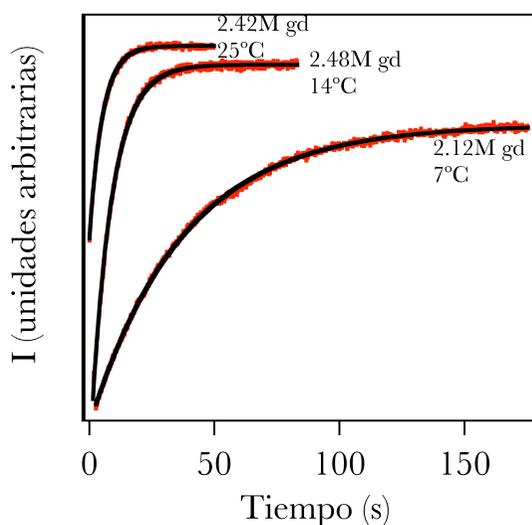
fig.2.17. A) Espectro de excitación de fluorescencia a 25°C de la tiorredoxina de LPBCA en su estado nativo y desnaturalizado. B) Espectro de emisión de fluorescencia a 25°C de la tiorredoxina de LPBCA en su estado nativo y desnaturalizado. La proteína nativa está disuelta en tampón HEPES 50 mM pH 7, mientras que la proteína desplegada se encuentra en guanidina 7M, HEPES 50 mM pH 7. La concentración de proteína es de 0.03 mg/mL en ambos casos. La velocidad de barrido seleccionada es de 600 nm/min y las rendijas de excitación y emisión son de 10 nm. Se ha utilizado una cubeta de cuarzo de 1 cm de paso de luz. La longitud de onda de excitación seleccionada ha sido de 276 nm; se puede observar como la máxima diferencia entre los dos espectros correspondientes a la proteína nativa y desplegada se produce a 350 nm.

### **2.8.3. Cinéticas de desplegamiento de tiorredoxina silvestre y variantes de tiorredoxina seguidas por fluorescencia**

Las cinéticas de desplegamiento de la tiorredoxina de *E. coli* se llevaron a cabo con un fluorímetro  $\lambda 55$  de Perkin Elmer y las cinéticas de desplegamiento de la tiorredoxina de LPBCA se llevaron a cabo con un fluorímetro Cary Eclipse de Agilent Technologies. Ambos equipos llevan acoplado un dispositivo multicubeta que permite la medida de un máximo de cuatro muestras simultáneamente. El dispositivo multicubeta del equipo  $\lambda 55$  de Perkin Elmer es termostatzado con un baño MH de Julabo, que permite trabajar en un rango de temperaturas de 7°C a 68°C. El dispositivo multicubeta del equipo Cary Eclipse de Agilent Technologies está controlado térmicamente por una Peltier permitiendo trabajar en un mayor rango de temperaturas (2-98°C). Para evitar posibles efectos de fotodescomposición de la muestra, el fluorímetro constan de una lámpara pulsante que irradia la muestra intermitentemente. Se han caracterizado las cinéticas de desplegamiento inducido por guanidina de la tiorredoxina de *E. coli* y de LPBCA a diferentes temperaturas (a 7°C, 14°C, 25°C, 30°C y 37°C para la tiorredoxina de *E. coli* y a 6°C, 14°C, 20°C, 25°C, 30°C, 37°C, 44°C y 50°C para la tiorredoxina de LPBCA), excitando a 276 nm y siguiendo la emisión de fluorescencia a 350 nm. En fig.2.18 se muestran, a modo de ejemplo, algunos perfiles cinéticos de desplegamiento de la tiorredoxina de *E. coli*. Las rendijas de emisión y de excitación se mantuvieron en 10 nm. La proteína se dializó en tampón HEPES 50 mM, pH 7.0 y las disoluciones concentradas de guanidina se prepararon según se ha indicado en el apartado 2.5.4.

Se partió de una disolución de proteína nativa a 5 mg/mL en tampón acuoso y se transfirieron 20  $\mu$ L a la cubeta del fluorímetro que contenía 1980  $\mu$ L de una disolución conocida de guanidina termostatzada. Una vez se añadió la proteína a la cubeta, se agitó la disolución con una pipeta pasteur y rápidamente se comenzó a medir el aumento en la intensidad de fluorescencia en función del tiempo. Una

vez finalizada la cinética, se midió el índice de refracción de la disolución para conocer exactamente la concentración de guanidina (ec.2.4), ya que se ha producido una dilución a causa de la adición de la proteína. Así se midieron las cinéticas de desplegamiento en presencia de desnaturante a varias temperaturas. La temperatura máxima a la que se registraron las cinéticas (37°C para la tiorredoxina silvestre de *E. coli* y 50°C para la tiorredoxina de LPBCA) viene determinada por la limitación instrumental que supone no poder caracterizar cinéticas cuyos tiempos de relajación sean comparables al tiempo de respuesta del fluorímetro y al tiempo muerto debido a la homogeneización manual de la muestra.



*fig.2.18. Perfiles cinéticos de desplegamiento de tiorredoxina silvestre, registrados a 7°C, 14°C y 25°C después de transferir una alícuota de tiorredoxina nativa hasta disoluciones de una concentración de guanidina igual a 2.12 M, 2.48M y 2.42M respectivamente. En rojo se muestran los datos experimentales y en negro el ajuste a una ecuación monoexponencial (ec.2.43).*

Las cinéticas se ajustaron a perfiles monoexponenciales. Para una cinética de primer orden, la dependencia de la intensidad de fluorescencia con el tiempo viene dada por la ecuación de velocidad:

$$I=I_{\infty}-\Delta I \cdot e^{(-t/\tau)} \quad \text{ec.2.43}$$

Donde  $I_{\infty}$  es la intensidad de fluorescencia a tiempo infinito,  $\Delta I$  la amplitud de la exponencial y  $\tau$  el tiempo de relajación que se define como el tiempo requerido para que la concentración de reactivo disminuya  $1/e$  de su valor inicial.  $\tau$  viene dado por la inversa de la constante de velocidad:

$$\tau = \frac{1}{k} \quad \text{ec.2.44}$$

## 2.9. Estudio proteolítico de BMP-2 silvestre y variantes de BMP-2 por incubación con termolisina

La susceptibilidad de una proteína a ser digerida por proteasas viene condicionada por la exposición al medio de las zonas de corte que frecuentemente son secuencias enterradas en la estructura de la proteína [59]. Esto hace que proteínas con estructura muy compacta sean malos sustratos para las proteasas, mientras que proteínas con estructuras flexibles permitan ser digeridas más fácilmente [60, 61]. Varios trabajos en la literatura describen cómo sometiendo una proteína a incubación con proteasas y mediante el análisis por SDS-PAGE del lisado final se puede obtener información de la estabilidad termodinámica o cinética de la proteína en cuestión [62-64].

En cuanto al procedimiento experimental, tanto la termolisina de *Bacillus thermoproteolyticus rokko* (de SIGMA de referencia T-7902) como las muestras de variantes de BMP-2 se dializaron frente a 100 mM Acético, 10 mM  $\text{CaCl}_2$  a pH=5.5 (tampón de proteólisis). En un eppendorf termostaticado a 25°C se preparó la mezcla de reacción que contenía:

- variante de BMP-2 a una concentración aproximada de 0.25 mg/mL
- termolisina a diferentes concentraciones en un rango de 0.01-0.1 mg/mL
- tampón de proteólisis hasta un volumen total de 180  $\mu\text{L}$

Tras la adición de termolisina se agitó la mezcla de reacción mediante pipeteo y se tomaron muestras de 10  $\mu\text{L}$  a diferentes tiempos. Inmediatamente se adicionó 5  $\mu\text{L}$  de disolución 100 mM EDTA a cada muestra tomada, con objeto de frenar la digestión. El EDTA forma complejo con los iones  $\text{Ca}^{2+}$ , al no tener disponibles la termolisina los iones  $\text{Ca}^{2+}$  no puede digerir la variante de BMP-2. Las muestras se procesaron por SDS-PAGE y se analizaron con el programa Image J, obtenido de la página web <http://rsbweb.nih.gov/ij/>. La intensidad de la banda correspondiente a la proteína sometida a digestión se representó frente a tiempo de incubación (fig.2.20) y los perfiles resultantes se ajustaron a una ecuación monoexponencial (ec.2.45) o biexponencial (ec.2.46).

$$I = I_{\infty} + \Delta I \cdot e^{(-t/\tau)} \quad \text{ec.2.45}$$

$$I = I_{\infty} + \Delta I_1 \cdot e^{(-t/\tau_1)} + \Delta I_2 \cdot e^{(-t/\tau_2)} \quad \text{ec.2.46}$$

Siendo  $I$  la intensidad de la banda electroforética que disminuye con el tiempo de incubación,  $I_{\infty}$  es la intensidad de la banda a tiempo infinito que tiende a 0 y  $\tau$  es el tiempo al que se digiere la mitad de la BMP-2.

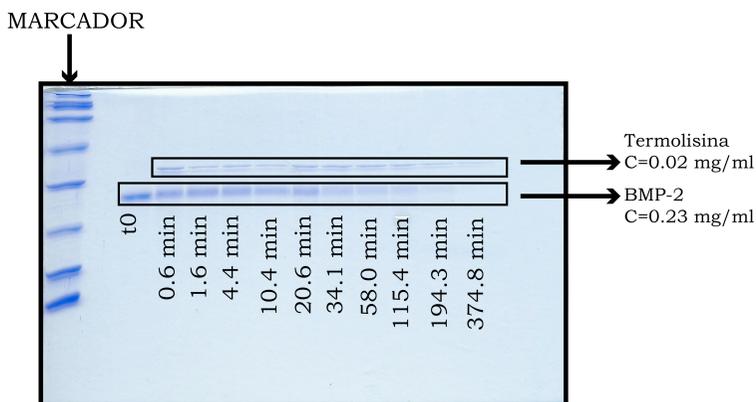


fig.2.19. Gel de electroforesis SDS-PAGE donde se observa la disminución de intensidad de la banda correspondiente a la BMP-2 silvestre a medida que aumentan los tiempo de incubación con termolisina. La disolución de 0.23 mg/mL de BMP-2 silvestre en 100 mM Acético, 10 mM  $\text{CaCl}_2$  a pH=5.5, se sometió a digestión con 0.02 mg/mL de Termolisina a 25°C.

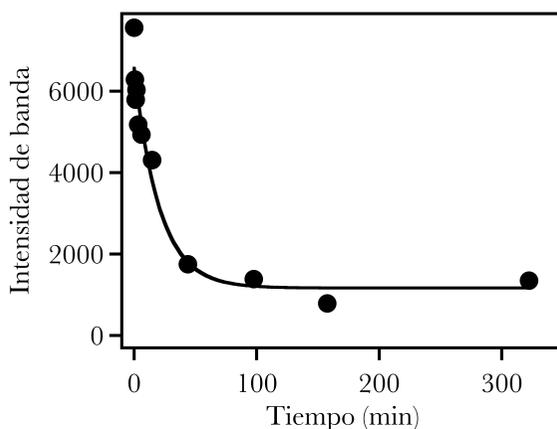


fig.2.20. Perfil de intensidad de banda electroforética correspondiente a la BMP-2 silvestre sometida a digestión con termolisina frente al tiempo de incubación. En círculos negros se representan los datos experimentales y la línea continua el ajuste a una ecuación monoexponencial.

## 2.10. Ensayo de actividad reductasa de tiorredoxina silvestre y variantes de tiorredoxina

La tiorredoxina contiene un puente disulfuro con actividad reductasa [65, 66] que puede reaccionar con NADPH y reducirse hasta la forma ditiol (tiorredoxina-SH<sub>2</sub>). La forma reducida de la tiorredoxina es capaz, a su vez de reducir puentes disulfuros en otras proteínas [67, 68] (fig.2.21).

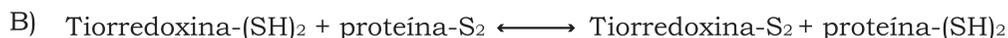
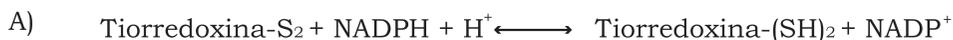
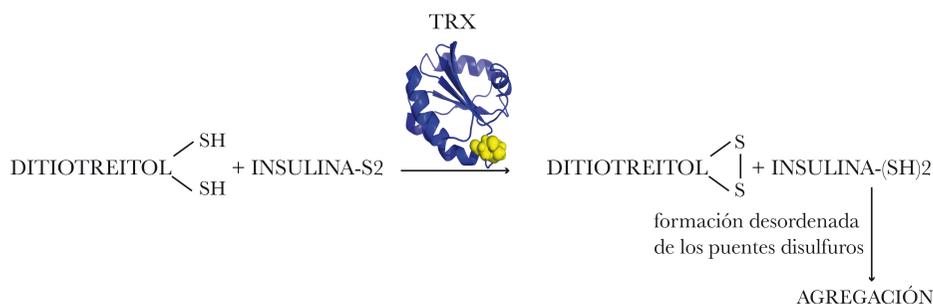


fig.2.21. Reacción de reducción del enlace disulfuro de la tiorredoxina (A) y de oxidación del grupo ditiol de la tiorredoxina en la forma reducida (B).

El ensayo utilizado en el presente trabajo [69] se basa en la capacidad de la tiorredoxina para catalizar la reducción de los puentes disulfuros de la insulina por el ditiotreitól (DTT). La insulina es una proteína compuesta de dos cadenas polipeptídicas unidas entre sí por dos enlaces disulfuros. Tras la reducción de los

enlaces disulfuros por parte del ditioneitol y catalizada por la tiorredoxina, estos se vuelven a formar aleatoriamente lo que produce la agregación de la insulina (fig.2.22).



*fig.2.22. Reacción de reducción de la insulina por DTT catalizada por la tiorredoxina. El aumento de turbidez debido a la agregación de la insulina por la formación desordenada de sus puentes disulfuro, se sigue por medidas de absorbancia a 650 nm.*

La velocidad de agregación de la insulina se registra con la medida de la turbidez a 650 nm en función del tiempo (fig.2.23). Representando la velocidad máxima de agregación de insulina en función de la concentración de tiorredoxina en disolución, obtenemos una relación lineal entre ambos valores a concentraciones iniciales de tiorredoxina. La pendiente resultante del ajuste a una línea recta es un buen parámetro para cuantificar la actividad reductasa de la tiorredoxina.

Las medidas de absorbancia a 650 nm se han llevado a cabo en un espectrofotómetro Agilent modelo 8453 con dispositivo multicubeta que permite la medida de 8 muestras simultáneamente y que está acoplado a un baño termostático de PolyScience. Para la realización del ensayo, se necesitan los siguientes reactivos:

- Tampón 0.1M fosfato potásico, 2 mM EDTA a pH 6.5.
- Disolución 0.1 M de DTT. Debe prepararse fresca cada día.
- Disolución de insulina a 10 mg/mL. Con el fin de eliminar cualquier posible microagregado, la disolución se sometió a un salto de pH. Así, se

pesaron 50 mg de insulina bovina de SIGMA y se disolvieron en 4 mL de Tris-HCl 50 mM pH 8.3. Por adición de HCL 1M se bajó el pH a 3.0, posteriormente, se vuelve a subir el pH hasta 8.0 mediante adición de NaOH 1M y se enrasa a 5 mL con agua.

- Disolución de tiorredoxina silvestre o de variante de tiorredoxina a 0.5 mg/mL aproximadamente.

Con las cubetas situadas en el espectrofotómetro termostatzado a 37°C, se mezclaron 784  $\mu\text{L}$  de tampón 0.1M fosfato potásico, 2 mM EDTA a pH 6.5 con diferentes volúmenes de disolución de tiorredoxina (10, 20, 40 y 80  $\mu\text{L}$ ) y se llevó hasta un volumen total de 950  $\mu\text{L}$  con agua destilada. Se añadió 50  $\mu\text{L}$  de insulina y 10  $\mu\text{L}$  de DTT 0.1M, se agitó rápidamente con una pipeta pasteur y se comenzó a registrar la absorbancia a 650 nm en función del tiempo.

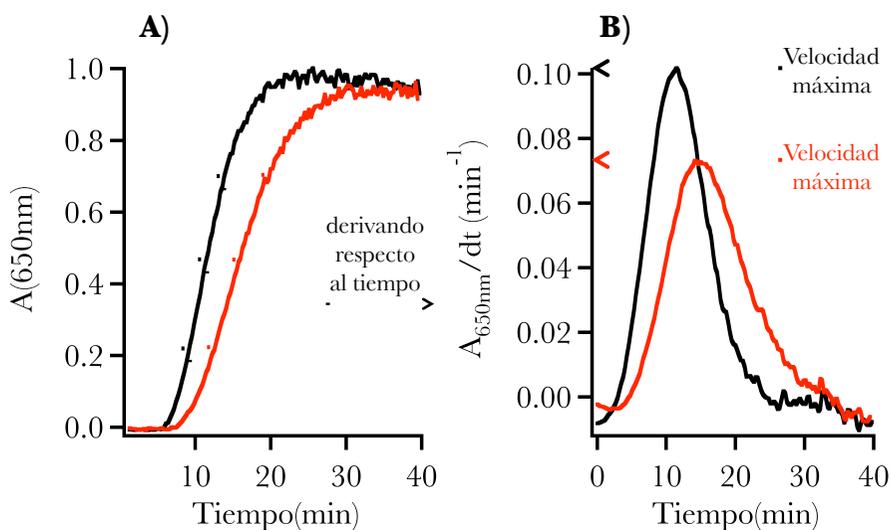


fig.2.23. A) Curva de reducción de la insulina, registrada por medidas de absorbancia a 650 nm, catalizada por la tiorredoxina de *E. coli* silvestre a una concentración de 1.49  $\mu\text{M}$  (negro) y de 0.99  $\mu\text{M}$  (rojo). B) Velocidad de agregación de la insulina (derivada de la absorbancia registrada a 650 nm respecto al tiempo); el máximo señalado en la gráfica se corresponde con la máxima velocidad de formación de agregados, que es el parámetro utilizado para cuantificar la actividad de la proteína catalizadora.

## **2.11. Ensayos celulares de BMP-2 silvestre y variantes de BMP-2**

### **2.11.1. Mantenimiento de la línea celular mioblástica C2C12**

Los ensayos celulares se realizaron sobre la línea celular mioblástica C2C12 [70] de ATCC (*American Type Culture Collection*). Las células C2C12 fueron aisladas en 1977 del tejido muscular del muslo de un ratón de dos meses de edad por Yaffe y Saxel [71] y en 1983 fueron cariotipadas y subclonadas por Blau [70]. Estas se mantienen como mioblastos no diferenciados en medio DMEM suplementado con 20% de suero fetal bovino y 0.5% de extracto de embrión de pollo; para inducir la fusión de mioblastos y la formación de fibras multinucleadas, el medio DMEM se suplementa con 2% de suero de caballo .

En el presente trabajo las células se conservaron en matraces de 25 cm<sup>2</sup> de superficie con medio DMEM enriquecido (10% de suero fetal bovino, 100 U/mL de penicilina y 100 µg/mL de sulfato de estreptomicina) a 37°C y en atmósfera de 5% de CO<sub>2</sub>. Tras 3-4 días, cuando las células se encontraban en confluencia (cubrían el 80-90% de la superficie del matraz), el cultivo se sometió a propagación celular. Para ello se eliminó el medio de cultivo, se lavaron las células con PBS (20 mM fosfato, 150 mM NaCl, pH 7.4), se despegaron con tripsina y se pasaron a un tubo falcon con medio DMEM enriquecido. El número de células obtenidas se midió con una cámara de Neubauer y se sembraron en un nuevo matraz a una concentración de 2500 células por cm<sup>2</sup> de superficie de cultivo en medio DMEM enriquecido. Todos los ensayos celulares se realizaron por triplicado, en células con menos de 10 pases de propagación celular y en placas de cultivo de 48 pocillos de Corning® Costar®.

### 2.11.2. Ensayo de diferenciación celular seguido por medidas de actividad fosfatasa alcalina

La proteína BMP-2 convierte la vía de diferenciación celular mioblástica de las células C2C12 hacia el linaje osteoblástico, siendo esta conversión no hereditaria. De manera que incubando las células con BMP-2, se inhibe la formación de miotubos multicelulares y se induce la expresión de numerosas fosfatasas alcalinas [72]. La enzima fosfatasa alcalina es un marcador temprano de la diferenciación ósea. Es una enzima hidrolasa responsable de la defosforilación de varios tipos de moléculas tales como nucleótidos, proteínas y alcaloides. La fosfatasa alcalina es más efectiva a pH básico. La actividad de esta enzima puede ser medida por un método colorimétrico ya que cataliza la defosforilación del p-nitrofenilfosfato para producir p-nitrofenol que es un producto de color amarillo que absorbe a 405 nm (fig.2.24).

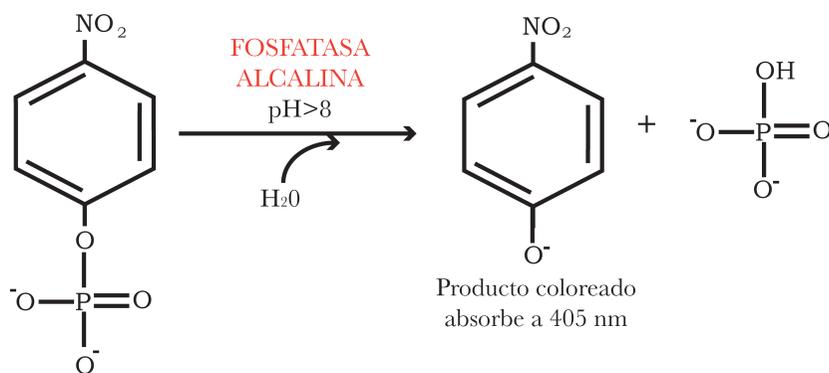


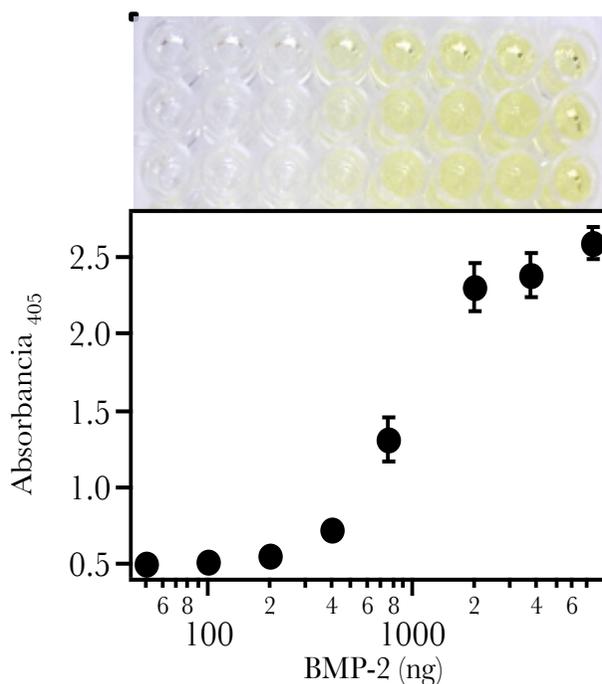
fig.2.24. Reacción de defosforilación del p-nitrofenilfosfato catalizada por la enzima fosfatasa alcalina. Debido a la eliminación del fosfato, se produce p-nitrofenol que es un producto coloreado que absorbe a 405 nm.

Las medidas de absorbancia a 405 nm se han llevado a cabo en un lector de placas modelo FL-600 de Biotek. Para la realización del ensayo, se necesitan los siguientes reactivos:

- Células C2C12 conservadas en medio DMEM tal y como se ha descrito previamente.
- Medio de cultivo DMEM enriquecido. El medio DMEM se conserva congelado en alícuotas de 45 mL en tubos falcon y previo a su utilización se le adiciona 5 mL de suero fetal bovino y 500  $\mu$ L de disolución de penicilina y de sulfato de estreptomicina a una concentración de 10 U/ $\mu$ L y 10  $\mu$ g/ $\mu$ L respectivamente.
- Disolución de proteína silvestre o variante de BMP-2 en 50 mM ácido acético a una concentración aproximada de 1 mg/mL.
- Tampón PBS, 20 mM fosfato, 150 mM NaCl, pH 7.4.
- Tampón de lisis, Tris 50mM, Tritón X-100 al 10%, MgCl<sub>2</sub> 2 mM, mP 6.8
- Tampón sustrato, se debe preparar fresco antes de cada ensayo. Se disuelven 100 mg de p-nitrofenilfosfato en 10 mL de agua desionizada y se mezcla igual volumen de disolución de p-nitrofenilfosfato con tampón comercial 2-amino-2-metil-1-propanol de SIGMA.
- Disolución de NaOH 0.5M.

Para la realización del ensayo, tras el pase celular se sembraron 10.000 células C2C12 por pocillo (10.000 cel/cm<sup>2</sup> de superficie aproximadamente) en una placa de 24 pocillos de Corning® Costar®; se añadió DMEM completo hasta un volumen de 400  $\mu$ L; finalmente se añadieron diferentes cantidades de proteína (5ng - 8000 $\mu$ g) por pocillo y se dejó incubando a 37°C y atmósfera de 5% de CO<sub>2</sub> durante 4 días. La disolución de proteína debe ser concentrada (1mg/mL), ya que al estar disuelta a pH ácido, si se añade mucho volumen de disolución se dañan las células C2C12. Por ello, en ningún ensayo el volumen añadido superó los 10  $\mu$ L de disolución. Tras la incubación, se eliminó el medio de cultivo de los pocillos con

cuidado de no desprender la capa de células adherida y se lavaron los pocillos con 400  $\mu\text{L}$  de PBS, el proceso de lavado se repitió tres veces para eliminar por completo los restos de proteínas del suero fetal bovino. Para la lisis celular se añadió 100  $\mu\text{L}$  de tampón de lisis a cada pocillo y se realizaron tres ciclos de congelación y descongelación dejando la placa 30 minutos a  $-20^{\circ}\text{C}$  y 30 minutos a  $37^{\circ}\text{C}$ . Se cogieron 10  $\mu\text{L}$  de lisado celular y se pasaron a una placa de 48 pocillos, se añadió 100  $\mu\text{L}$  de tampón sustrato y se dejó incubando 15 minutos a  $37^{\circ}\text{C}$ , para parar la reacción se añadió 100  $\mu\text{L}$  de NaOH 0.5M e inmediatamente se midió la absorbancia a 405 nm.



*fig.2.25. Ensayo de diferenciación celular de C2C12 a diferentes dosis de BMP-2 silvestre mediante medidas de actividad fosfatasa alcalina (FA). El ensayo de FA sobre en el lisado celular tratado con BMP-2, se realizó por triplicado para cada dosis de BMP-2. Así, en círculos negros se representa el valor medio y las barras de error representan la desviación estándar resultante de las tres medidas realizadas. Ilustración de la coloración amarilla del p-nitrofenol que se produce por la presencia de fosfatasa alcalina frente a la coloración transparente en aquellos pocillos donde, al no producirse la diferenciación, no se ha expresado la enzima fosfatasa alcalina.*

### 2.11.3. Cuantificación de la proliferación celular mediante la determinación del contenido total de proteínas por Bradford

Con objeto de cuantificar la proliferación de las células C2C12 al ser tratadas con BMP-2, se determinó el contenido total de proteínas en el lisado celular por el método MicroBradford de Biorad. Este ensayo se basa en la afinidad del colorante Coomassie Brilliant Blue G250 por las proteínas. Coomassie G250 es un colorante que al encontrarse en el entorno hidrofóbico del interior de una proteína, origina un color azul intenso que se puede cuantificar mediante medidas de absorbancia a 590 nm [73]. Las medidas de absorbancia a 590 nm se han llevado a cabo en un lector de placas modelo FL-600 de Biotek. Para la realización del ensayo, el lisado celular se obtuvo de la misma manera que para el ensayo de fosfatasa alcalina. Se cogieron 10  $\mu\text{L}$  de lisado celular y se pasaron a una placa de 48 pocillos, se añadió 60  $\mu\text{L}$  de reactivo Bradford de Biorad y 230  $\mu\text{L}$  de  $\text{H}_2\text{O}$  desionizada. Se incubó a 37°C durante 15 minutos y se midió la absorbancia a 590 nm (fig.2.27). Para cuantificar la cantidad total de proteínas se realizó una recta patrón empleando la proteína IgG (fig.2.26).

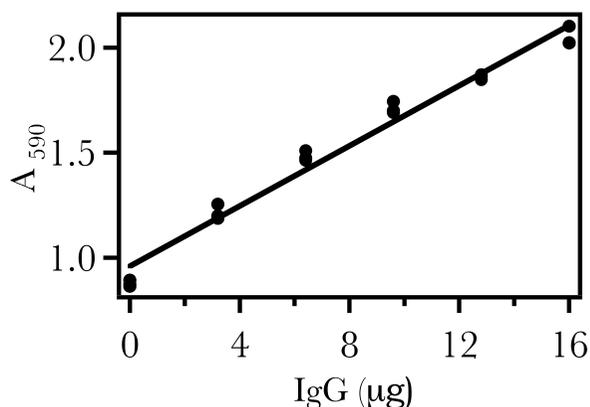


fig.2.26. Recta patrón del ensayo Bradford que relaciona cantidades conocidas de IgG con absorbancia a 590 nm.

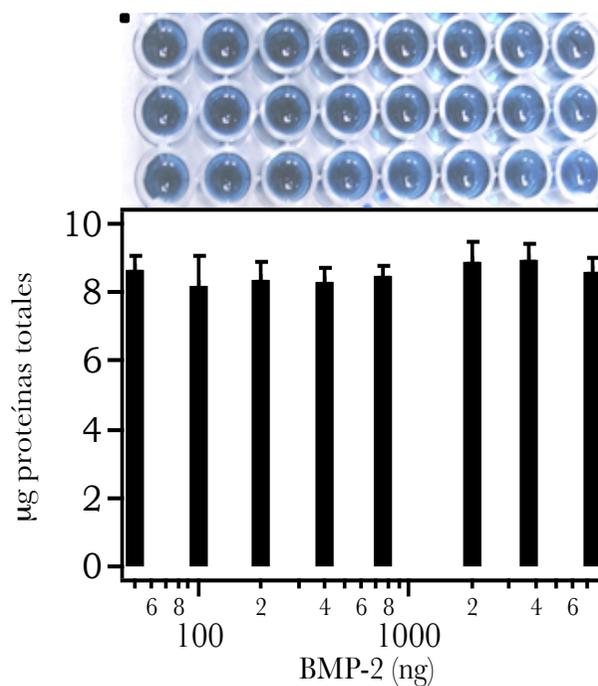


fig.2.27. Determinación del contenido total de proteínas tras incubar las células C2C12 durante 4 días con diferentes cantidades de BMP-2 silvestre. Para cada dosis de BMP-2, el ensayo Bradford se realizó por triplicado, así, las barras de error representan la desviación estándar resultante de las tres medidas realizadas. Ilustración de la coloración azul resultante de la unión inespecífica del colorante Coomassie G250 con las proteínas presentes.



## 2.12. Bibliografía

1. Bessette, P.H., et al., *Construction of designed protein libraries using gene assembly mutagenesis*. Methods in molecular biology, 2003. **231**: p. 29-37.
2. Langsetmo, K., J. Fuchs, and C. Woodward, *Escherichia coli thioredoxin folds into two compact forms of different stability to urea denaturation*. Biochemistry, 1989. **28**(8): p. 3211-20.
3. Pace, C.N., Shirley, B. A. and Thomson, J. A., *Protein structure: A practical approach*. 1989: p. 311.
4. Yasuhiko Nozaki, C.H.W.H., Serge N. Timasheff, [3] *The preparation of guanidine hydrochloride*. Methods in Enzymology, 1972. **26**.
5. Privalov, P.L., *Stability of proteins: small globular proteins*. Advances in protein chemistry, 1979. **33**: p. 167-241.
6. Privalov, P.L., *Stability of proteins. Proteins which do not present a single cooperative system*. Advances in protein chemistry, 1982. **35**: p. 1-104.
7. Privalov, P.L., *Thermodynamic problems of protein structure*. Annual review of biophysics and biophysical chemistry, 1989. **18**: p. 47-69.
8. Sanchez-Ruiz, J.M., *Differential scanning calorimetry of proteins*. Sub-cellular biochemistry, 1995. **24**: p. 133-76.
9. Sanchez-Ruiz, J.M., et al., *Differential scanning calorimetry of the irreversible thermal denaturation of thermolysin*. Biochemistry, 1988. **27**(5): p. 1648-52.
10. Freire, E., et al., *Calorimetrically determined dynamics of complex unfolding transitions in proteins*. Annual review of biophysics and biophysical chemistry, 1990. **19**: p. 159-88.
11. Cooper, A. and C.M. Johnson, *Differential scanning calorimetry*. Methods in molecular biology, 1994. **22**: p. 125-36.

12. Ibarra-Molero, B.S.-R., J. M., *Differential Scanning Calorimetry of Proteins: an Overview and Some Recent Developments*. Advanced Techniques in Biophysics, 2006. **10**.
13. Sanchez-Ruiz, B.I.-M.a.J.M., *Protein Folding and Misfolding*. Royal Society of Chemistry, London, 2006.
14. Sanchez-Ruiz, J.M., *Probing free-energy surfaces with differential scanning calorimetry*. Annual review of physical chemistry, 2011. **62**: p. 231-55.
15. Chowdhry, B.Z. and S.C. Cole, *Differential scanning calorimetry: applications in biotechnology*. Trends in biotechnology, 1989. **7**(1): p. 11-18.
16. Plotnikov, V., et al., *An autosampling differential scanning calorimeter instrument for studying molecular interactions*. Assay and drug development technologies, 2002. **1**(1 Pt 1): p. 83-90.
17. Lopez Mayorga, O. and E. Freire, *Dynamic analysis of differential scanning calorimetry data*. Biophysical chemistry, 1987. **27**(1): p. 87-96.
18. Plaza del Pino, I.M., B. Ibarra-Molero, and J.M. Sanchez-Ruiz, *Lower kinetic limit to protein thermal stability: a proposal regarding protein stability in vivo and its relation with misfolding diseases*. Proteins, 2000. **40**(1): p. 58-70.
19. Biltonen, R.L. and E. Freire, *Thermodynamic characterization of conformational states of biological macromolecules using differential scanning calorimetry*. CRC critical reviews in biochemistry, 1978. **5**(2): p. 85-124.
20. Takahashi, K. and J.M. Sturtevant, *Thermal denaturation of streptomyces subtilisin inhibitor, subtilisin BPN', and the inhibitor-subtilisin complex*. Biochemistry, 1981. **20**(21): p. 6185-90.
21. Bechtel, W.J. and J.A. Schellman, *Protein stability curves*. Biopolymers, 1987. **26**(11): p. 1859-77.

22. Pace, C.N. and D.V. Laurents, *A new method for determining the heat capacity change for protein folding*. Biochemistry, 1989. **28**(6): p. 2520-5.
23. Griko Iu, V., et al., [*Thermodynamic study of the structure of apomyoglobin*]. Biofizika, 1988. **33**(1): p. 18-26.
24. Griko Yu, V., S. Venyaminov, and P.L. Privalov, *Heat and cold denaturation of phosphoglycerate kinase (interaction of domains)*. FEBS letters, 1989. **244**(2): p. 276-8.
25. Makhatadze, G.I., S.J. Gill, and P.L. Privalov, *Partial molar heat capacities of the side chains of some amino acid residues in aqueous solution. The influence of the neighboring charges*. Biophysical chemistry, 1990. **38**(1-2): p. 33-7.
26. Makhatadze, G.I. and P.L. Privalov, *Protein interactions with urea and guanidinium chloride. A calorimetric study*. Journal of molecular biology, 1992. **226**(2): p. 491-505.
27. Schellman, J.A., *The thermodynamic stability of proteins*. Annual review of biophysics and biophysical chemistry, 1987. **16**: p. 115-37.
28. Brandts, J.F., *The Thermodynamics of Protein Denaturation. I. The Denaturation of Chymotrypsinogen*. Journal of the American Chemical Society, 1964. **86**(20).
29. Griko Iu, V. and P.L. Privalov, [*Cold denaturation of myoglobin in alkaline solutions*]. Doklady Akademii nauk SSSR, 1986. **291**(3): p. 709-11.
30. Privalov, P.L., et al., *Cold denaturation of myoglobin*. Journal of molecular biology, 1986. **190**(3): p. 487-98.
31. Griko, Y.V., et al., *Cold denaturation of staphylococcal nuclease*. Proceedings of the National Academy of Sciences of the United States of America, 1988. **85**(10): p. 3343-7.

32. Tamura, A., K. Kimura, and K. Akasaka, *Cold denaturation and heat denaturation of Streptomyces subtilisin inhibitor. 2. <sup>1</sup>H NMR studies*. Biochemistry, 1991. **30**(47): p. 11313-20.
33. Azuaga, A.I., et al., *Heat and cold denaturation of beta-lactoglobulin B*. FEBS letters, 1992. **309**(3): p. 258-60.
34. Ibarra-Molero, B., G.I. Makhatadze, and J.M. Sanchez-Ruiz, *Cold denaturation of ubiquitin*. Biochimica et biophysica acta, 1999. **1429**(2): p. 384-90.
35. Luisa Romero-Romero, M., et al., *Highly Anomalous Energetics of Protein Cold Denaturation Linked to Folding-Unfolding Kinetics*. PLoS One, 2011. **6**(7).
36. Plaza del Pino, I.M., C.N. Pace, and E. Freire, *Temperature and guanidine hydrochloride dependence of the structural stability of ribonuclease T1*. Biochemistry, 1992. **31**(45): p. 11196-202.
37. Romero-Romero, M.L., et al., *Highly anomalous energetics of protein cold denaturation linked to folding-unfolding kinetics*. PloS one, 2011. **6**(7): p. e23050.
38. Sanchez-Ruiz, J.M., et al., *Analysis of the thermal unfolding of porcine procarboxypeptidase A and its functional pieces by differential scanning calorimetry*. European journal of biochemistry / FEBS, 1988. **176**(1): p. 225-30.
39. Guzman-Casado, M., et al., *Differential scanning calorimetry of lobster haemocyanin*. European journal of biochemistry / FEBS, 1990. **188**(1): p. 181-5.
40. Morin, P.E., D. Diggs, and E. Freire, *Thermal stability of membrane-reconstituted yeast cytochrome c oxidase*. Biochemistry, 1990. **29**(3): p. 781-8.
41. Lepock, J.R., et al., *Thermal denaturation of the Ca<sup>2+</sup>(+)-ATPase of sarcoplasmic reticulum reveals two thermodynamically independent domains*. Biochemistry, 1990. **29**(3): p. 681-9.

42. Conejero-Lara, F., et al., *Effect of  $Zn^{2+}$  on the thermal denaturation of carboxypeptidase B*. Biochemistry, 1991. **30**(8): p. 2067-72.
43. Conejero-Lara, F., et al., *Differential scanning calorimetric study of carboxypeptidase B, procarboxypeptidase B and its globular activation domain*. European journal of biochemistry / FEBS, 1991. **200**(3): p. 663-70.
44. Galisteo, M.L., P.L. Mateo, and J.M. Sanchez-Ruiz, *Kinetic study on the irreversible thermal denaturation of yeast phosphoglycerate kinase*. Biochemistry, 1991. **30**(8): p. 2061-6.
45. Lumry, R. and H. Eyring, *Conformation Changes of Proteins*. The Journal of Physical Chemistry, 1954. **58**(2): p. 110-120.
46. Kholodenko, V. and E. Freire, *A simple method to measure the absolute heat capacity of proteins*. Analytical biochemistry, 1999. **270**(2): p. 336-8.
47. Georgescu, R.E., et al., *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*. European journal of biochemistry / FEBS, 2001. **268**(5): p. 1477-85.
48. Makhatadze, G.I. and P.L. Privalov, *Heat capacity of proteins. I. Partial molar heat capacity of individual amino acid residues in aqueous solution: hydration effect*. Journal of molecular biology, 1990. **213**(2): p. 375-84.
49. Josef Michl, E.W.T., *Spectroscopy with polarized light: solute alignment by photoselection in liquid crystals, polymers and membranes*. 1986.
50. Klinger D.S., L.J.W., Randall C.E., *Polarized light in optics and spectroscopy*. Academic Press, 1990.
51. Woody, R.W., H. Sugeta, and T.S. Kodama, *[Circular dichroism of proteins: recent developments in analysis and prediction]*. Tanpakushitsu kakusan koso. Protein, nucleic acid, enzyme, 1996. **41**(1): p. 56-69.

52. Lowry, T.M., *The origin of optical rotatory power*. Nature, 1935. **136**: p. 191-192.
53. Wu, J., J.T. Yang, and C.S. Wu, *Beta-II conformation of all-beta proteins can be distinguished from unordered form by circular dichroism*. Analytical biochemistry, 1992. **200**(2): p. 359-64.
54. Manavalan, P. and W.C. Johnson, *Sensitivity of Circular-Dichroism to Protein Tertiary Structure Class*. Nature, 1983. **305**(5937): p. 831-832.
55. Fandrich, M. and C.M. Dobson, *The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation*. The EMBO journal, 2002. **21**(21): p. 5682-90.
56. Lakowicz, J.R., *Principles of Fluorescence Spectroscopy*. Springer, 2006. **2006**.
57. Reutimann, H., et al., *A conformational study of thioredoxin and its tryptic fragments*. The Journal of biological chemistry, 1981. **256**(13): p. 6796-803.
58. Kelley, R.F. and E. Stellwagen, *Conformational transitions of thioredoxin in guanidine hydrochloride*. Biochemistry, 1984. **23**(22): p. 5095-102.
59. Tyndall, J.D., T. Nall, and D.P. Fairlie, *Proteases universally recognize beta strands in their active sites*. Chemical reviews, 2005. **105**(3): p. 973-99.
60. Park, C., et al., *Energetics-based protein profiling on a proteomic scale: identification of proteins resistant to proteolysis*. Journal of molecular biology, 2007. **368**(5): p. 1426-37.
61. Anfinsen, C.B. and H.A. Scheraga, *Experimental and theoretical aspects of protein folding*. Advances in protein chemistry, 1975. **29**: p. 205-300.
62. Park, C. and S. Marqusee, *Pulse proteolysis: a simple method for quantitative determination of protein stability and ligand binding*. Nature methods, 2005. **2**(3): p. 207-12.

63. Park, C. and S. Marqusee, *Probing the high energy states in proteins by proteolysis*. Journal of molecular biology, 2004. **343**(5): p. 1467-76.
64. Tur-Arlandis, G., et al., *Proteolytic scanning calorimetry: a novel methodology that probes the fundamental features of protein kinetic stability*. Biophysical journal, 2010. **98**(6): p. L12-4.
65. Holmgren, A., et al., *Three-dimensional structure of Escherichia coli thioredoxin-S2 to 2.8 Å resolution*. Proceedings of the National Academy of Sciences of the United States of America, 1975. **72**(6): p. 2305-9.
66. Engstrom, N.E., et al., *Isolation and characterization of calf liver thioredoxin*. The Journal of biological chemistry, 1974. **249**(1): p. 205-10.
67. Holmgren, A., *Bovine thioredoxin system. Purification of thioredoxin reductase from calf liver and thymus and studies of its function in disulfide reduction*. The Journal of biological chemistry, 1977. **252**(13): p. 4600-6.
68. Holmgren, A., *Reduction of disulfides by thioredoxin. Exceptional reactivity of insulin and suggested functions of thioredoxin in mechanism of hormone action*. The Journal of biological chemistry, 1979. **254**(18): p. 9113-9.
69. Holmgren, A., *Thioredoxin catalyzes the reduction of insulin disulfides by dithiothreitol and dihydrolipoamide*. The Journal of biological chemistry, 1979. **254**(19): p. 9627-32.
70. Blau, H.M., C.P. Chiu, and C. Webster, *Cytoplasmic activation of human nuclear genes in stable heterocaryons*. Cell, 1983. **32**(4): p. 1171-80.
71. Yaffe, D. and O. Saxel, *Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle*. Nature, 1977. **270**(5639): p. 725-7.
72. Katagiri, T., et al., *Bone morphogenetic protein-2 converts the differentiation pathway of C2C12 myoblasts into the osteoblast lineage*. The Journal of cell biology, 1994. **127**(6 Pt 1): p. 1755-66.

73. Bradford, M.M., *A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding*. Analytical biochemistry, 1976. **72**: p. 248-54.

### **3. Sección experimental: objetivos y resultados**



### 3.1. Proteína Morfogenética de Hueso 2

Las Proteínas Morfogenéticas de Hueso (BMPs, *Bone Morphogenetic Proteins*) son un grupo de factores de crecimiento conocido inicialmente por su capacidad de inducir formación de hueso y cartílago [1, 2]. Estas proteínas pertenecen a la superfamilia de los Factores de Crecimiento Transformante tipo  $\beta$  (TGF- $\beta$ , *Transforming Growth Factor beta*), que son citoquinas reguladoras de procesos celulares. Está descrita su participación en el control del crecimiento, la proliferación y la apoptosis celular [3-5]. Las BMPs se pueden clasificar en al menos cuatro grupos, en base a criterios estructurales y funcionales: el grupo BMP2/4, GDF-5 (*Growth and Differentiation factor 5*), OP-1 (*Osteogenic Protein 1*) y BMP-9 [6] (fig 3.1.1). La mayoría de las proteínas que se engloban en los grupos BMP2/4, OP-1 y BMP-9 han demostrado inducir la formación de cartílago y hueso *in vivo*, mientras que las del grupo GDF-5 se caracterizan por la formación de cartílago y tendón. De todas las BMPs, la BMP-2 es la proteína que presenta mayor capacidad de inducir regeneración ósea y de formar hueso ectópico en adultos. Por lo que se considera de gran interés en el campo de la biomedicina, con aplicaciones relacionadas con la fusión espinal, el tratamiento de fracturas, la cirugía ortopédica y la cirugía oral [7-10].

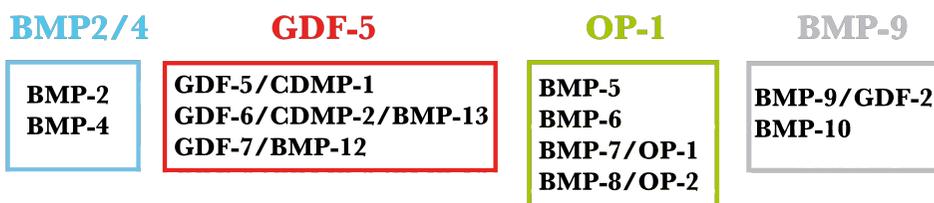


fig 3.1.1. Las BMPs se pueden clasificar en, al menos, cuatro grupos de acuerdo con criterios estructurales y funcionales [11].

La BMP-2 se sintetiza, como un precursor de 396 aminoácidos que consta de un péptido señal, un dominio y la secuencia correspondiente a la proteína madura, que se encuentra en el extremo C-terminal del precursor. En el proceso de

maduración, el precursor sufre diferentes modificaciones post-traduccionales (glicosilaciones, proteólisis y dimerización) hasta alcanzar finalmente la estructura homodimérica activa que es secretada fuera de la célula (fig 3.1.2). De los cuatro sitios N-glicosilados presentes en el precursor, solo se mantiene uno en la cadena polipeptídica de la BMP-2 madura, no obstante, para que la proteína presente actividad no precisa de esta glicosilación [12, 13].

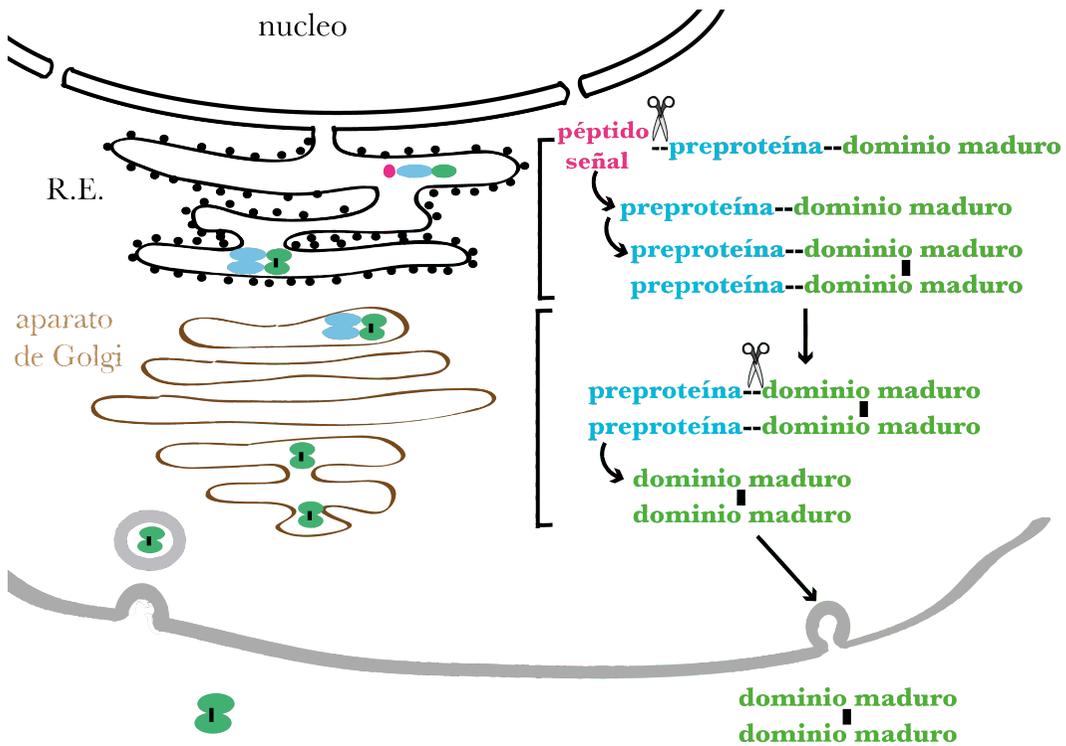


fig 3.1.2. Esquema de la producción y maduración de BMP-2 en un sistema de expresión eucariota. En primer lugar, el precursor sintetizado (pre-pro-BMP-2) es digerido perdiendo el péptido señal y en una etapa posterior se produce la dimerización de los monómeros de pro-BMP-2, ambas modificaciones se producen en el retículo endoplasmático. El dímero de pro-BMP-2 es transportado al aparato de Golgi donde, de nuevo, es digerido obteniéndose la proteína dimérica BMP2 madura. Finalmente el dímero maduro se secreta al medio extracelular.

La BMP-2 madura es una proteína homodimérica donde los monómeros se unen covalentemente por un puente disulfuro entre las Cys78 de cada subunidad [14]. Cada monómero, de 114 aminoácidos y 13 kDa de peso molecular, presenta una

estructura de nudo cistina (*cysteine knot*) constituida por tres puentes disulfuro intracatenarios. Como consecuencia de las uniones entre Cys43/Cys111 y Cys47/Cys113 se forma un anillo a través del cual pasa el tercer puente disulfuro formado por la unión Cys14/Cys79 (fig 3.1.3).

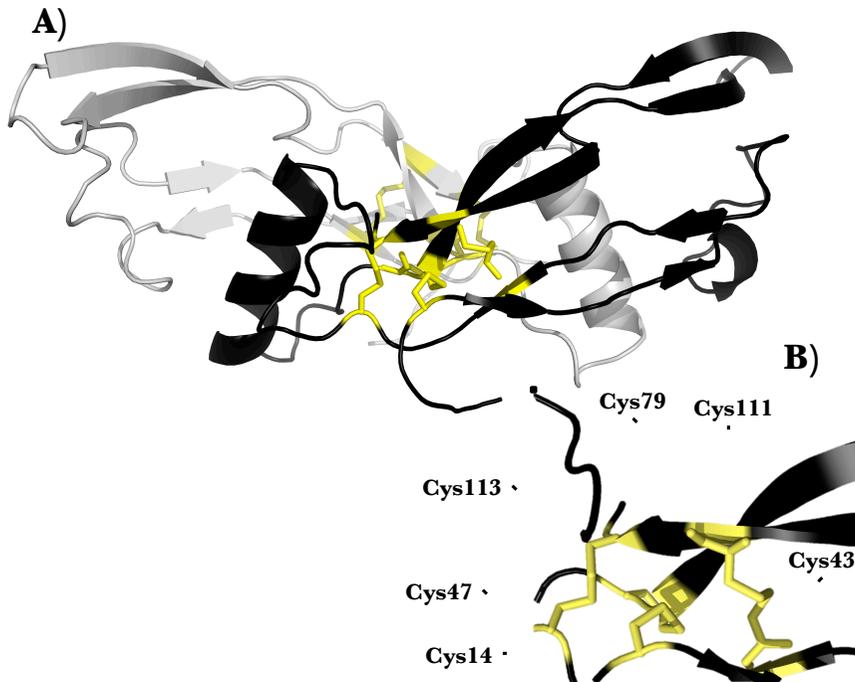


fig 3.1.3. A) Estructura cristalográfica de la proteína BMP-2 madura resuelta por Scheufler y colaboradores [14]. En gris y negro se muestran cada uno de los monómeros unidos covalentemente por un puente disulfuro entre las cisteínas 78 de cada subunidad. Cada monómero contiene una estructura de nudo cistina. En la ilustración, los enlaces disulfuros se muestran como barras amarillas. B) Ampliación del nudo cistina de uno de los monómeros donde se muestran las uniones entre Cys43/Cys111 y Cys47/Cys113, a través del anillo que forman estos enlaces, pasa el tercer puente disulfuro entre Cys14/Cys79.

Para producir BMP-2 activa, se puede utilizar un sistema de expresión eucariota, como las células de ovario de hámster chino (CHO) o un sistema de expresión bacteriano, como *E. coli*. En las células CHO, la BMP-2 madura se secreta al medio extracelular tras haber sufrido los procesos de maduración que se

representan en la fig 3.1.2. En *E. coli*, al no disponer la célula de la maquinaria necesaria para que se den las glicosilaciones y los procesos de proteólisis necesarios, se expresa directamente la cadena polipeptídica correspondiente al monómero maduro de BMP-2. Además, el entramado tan complejo de puentes disulfuro (los tres puentes intracatenarios y el intercatenario responsable de la dimerización) no se forma correctamente, por lo que se precisa de una etapa de replegamiento *in vitro*. A pesar de ello, la producción en *E. coli*, permite obtener mayores cantidades de proteína activa que otros sistemas de expresión, por esta razón, este ha sido el sistema utilizado para la obtención de BMP-2 silvestre y variantes, en la presente Tesis Doctoral.

La BMP-2 desencadena una respuesta celular al unirse a receptores de membrana, incidiendo en multitud de procesos celulares y alterando la expresión génica. Estos cambios están relacionados con procesos de desarrollo embrionario y con la diferenciación de tejidos mesenquimales, mayoritariamente a tejido óseo. Por su capacidad para diferenciar tejidos hacia un linaje osteoblástico, la BMP-2 es usada con fines terapéuticos en diversas aplicaciones como, por ejemplo, en el tratamiento de fracturas o en la cirugía dental [7-10].

En el trabajo que se expone a continuación nos planteamos explorar la posibilidad de que el enfoque de consenso, nos pudiera conducir a variantes de BMP-2 optimizadas desde el punto de vista de sus aplicaciones, es decir, conforme su capacidad para diferenciar células de músculo a células de hueso. Para ello, se realizó un alineamiento de la secuencia monomérica de BMP-2 con secuencias homólogas contenidas en las bases de datos de proteínas. Bajo la hipótesis de que las mutaciones que aparecen con más frecuencia en cada posición del alineamiento, pudiesen estar relacionadas con su actividad biológica, realizamos un análisis de consenso. Obtuvimos variantes de BMP-2 con las mutaciones propuestas por el análisis y para comprobar la hipótesis planteada, realizamos un estudio comparativo de la actividad biológica de las variantes de BMP-2 obtenidas.

Con objeto de caracterizar la función biológica se realizaron ensayos celulares *in vitro* basados en la detección de la actividad fosfatasa alcalina asociada a la inducción del fenotipo osteoblástico (descrito en la sección 2.11.2 de metodología).

### **3.1.1. Análisis bioinformático del alineamiento de secuencias:**

#### **Método de consenso**

Un alineamiento de secuencias consiste en una disposición ordenada de secuencias, de manera que se maximicen las coincidencias de aminoácidos para cada posición con respecto a una secuencia diana o secuencia “*query*”. El alineamiento de secuencias encierra información acerca de la historia evolutiva de la proteína *query*, que puede resultar útil para la mejora de alguna propiedad de dicha proteína.

El método de consenso consiste en realizar primero un recuento de la frecuencia de aparición de cada aminoácido para cada posición en el alineamiento. En segundo lugar, se elabora la secuencia consenso como aquella que contiene los aminoácidos más habituales en el alineamiento para cada posición. A lo largo de la evolución, las proteínas han ido sufriendo mutaciones que han podido ser toleradas o rechazadas en función del efecto que producen en las propiedades de las proteínas. De manera que existen determinados límites selectivos, asociado a las características fenotípicas de las proteínas, que determinan las mutaciones que son aceptadas. Si una mutación produce un efecto positivo en una propiedad, será aceptada. Sin embargo, si una mutación produce un efecto negativo en una propiedad, solo será aceptada si no produce un descenso en la propiedad por debajo del límite selectivo, esta situación implica que previamente se hayan dado mutaciones que hayan aumentado significativamente el valor de la propiedad. Por estadística, las mutaciones que conllevan una mejora en una propiedad están más representadas en las secuencias de proteínas actuales que las mutaciones que conllevan un empeoramiento en una propiedad. Así, se supone que cuando

realizamos un alineamiento de una proteína con secuencias del mismo origen evolutivo (secuencias homólogas) podemos pensar que mutaciones al aminoácido que aparece con más frecuencia en cada posición del alineamiento (aminoácido consenso) presumiblemente tienen un efecto beneficioso, en el sentido de aumentar la estabilidad o la función por ejemplo. La eficacia del método de consenso depende de la bondad del alineamiento de secuencias, por lo que es muy importante que represente correctamente el equilibrio mutacional/selectivo sobre cada posición. Esta metodología ha sido utilizada con éxito, para obtener proteínas mejoradas en su estabilidad [15-17] y en su actividad [16]. Además, recientemente se ha conseguido en nuestro laboratorio, combinado los residuos más frecuentes en cada posición de un alineamiento, una mejora simultánea de ambas propiedades [16].

En este trabajo, nos planteamos explorar la posibilidad de que el enfoque de consenso, nos pudiera conducir a variantes de BMP-2 optimizadas conforme su capacidad para diferenciar células de músculo a células de hueso. Para ello, obtuvimos mutantes simples de BMP-2 con las mutaciones sugeridas por el análisis de consenso. *A priori*, no podemos relacionar la alta frecuencia de aparición de un residuo en el alineamiento de secuencias, con que la mutación a dicho residuo en la proteína *query* produzca un aumento de su actividad. Con lo cual, todas las variantes serán estudiadas bajo su actividad biológica, mediante la realización de ensayos celulares *in vitro*, basados en la detección de la actividad fosfatasa alcalina asociada a la inducción del fenotipo osteoblástico. Tras analizar los resultados, realizaremos mutantes múltiples con las mutaciones de las variantes que hayan resultado mejoradas en cuanto a la actividad biológica a la BMP-2 silvestre.

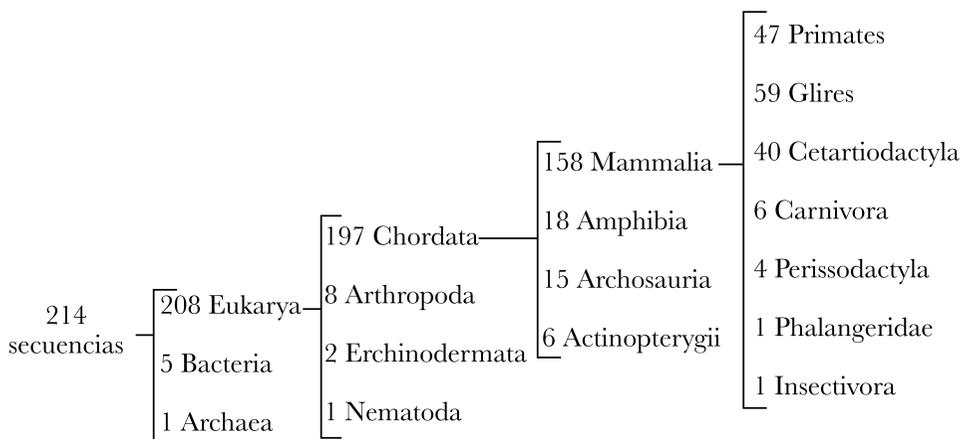
En cuanto al procedimiento experimental llevado a cabo, los alineamientos de secuencias se obtuvieron realizando una búsqueda con BLAST 2 [18] (<http://blast.wustl.edu>) en las bases de datos UniProtKB/TrEMBL y

UniProtKB/Swiss-Prot obtenidas de la página web <http://www.uniprot.org/> con fecha de 28 de julio de 2008. Los alineamientos de secuencias se realizaron tomando como *query* la secuencia monomérica de la BMP-2 madura cuyos códigos de PDB (*Protein Data Bank*) y de Uniprot son 3BMP y P12643 respectivamente. Los alineamientos se realizaron usando el algoritmo de Smith-Waterman [19, 20] y se filtraron por un límite de identidad mayor de 0.3, calculándose la identidad como el número de coincidencias entre una secuencia dada y la secuencia *query* dividido por el número de residuos en la secuencia *query*. Se eligió un corte de 0.3 porque se asume que proteínas con una identidad de al menos 0.25-0.30 tienen el mismo plegamiento [21], salvo contadas excepciones, este corte de identidad nos garantizará que la inmensa mayoría de las secuencias que queden en el alineamiento pertenezcan a proteínas homologas a la proteína *query*. Para la selección de secuencias con identidad mayor 0.30 se utilizó el programa *sec.exe*, desarrollado en nuestro grupo de investigación, que utiliza como fichero de entrada el alineamiento de secuencias obtenido con BLAST 2 y que nos proporciona el alineamiento con aquellas secuencias con un índice de identidad mayor a 0.3. Para el cálculo de frecuencias de aparición de un residuo en cada posición se utilizó el programa *FCR-SISI.exe* que utiliza como fichero de entrada el alineamiento de secuencias cribado por un límite de identidad de 0.3 y nos proporciona un fichero que contiene la siguiente información: la frecuencia de aparición del residuo *query* en el alineamiento, el residuo más frecuente diferente al *query* y su frecuencia de aparición en el alineamiento. Este programa también ha sido desarrollado en nuestro grupo de investigación. Para el alineamiento realizado con las secuencias contenidas en la base de datos Swiss-Prot, de las 213 secuencias iniciales se seleccionaron 168 (por tener un índice de identidad mayor a 0.3 con respecto a la *query*) y para el alineamiento realizado con las secuencias contenidas en la base de datos TrEMBL, de las 844 secuencias del alineamiento se seleccionaron 603 (fig 3.1.4).



Para clasificar las secuencias según la **función biológica** de la proteína que codifican, se buscó en la página web <http://www.uniprot.org/uniprot/> los procesos biológicos donde está implicada la proteína que codifica cada una de las secuencias que conforman el alineamiento. Por ejemplo, para la BMP-2, la proteína que codifica la secuencia *query*, está descrita su implicación en procesos biológicos de condrogénesis, diferenciación y osteogénesis. Se seleccionaron aquellas secuencias que codifican proteínas implicadas en procesos de osteogénesis, tras la criba quedaron 36 secuencias. Finalmente, se descartó realizar el análisis de consenso de las 36 secuencias resultantes ya que debido al número tan pequeño de secuencias, el resultado del análisis estaría muy condicionado por cada una de las secuencias y sería poco fiable.

Para organizar las secuencias según la **clasificación taxonómica** de los organismos a los que pertenecen las proteínas que codifican, se buscó en la página web <http://www.uniprot.org/uniprot/> la clasificación taxonómica de los organismos a los que pertenece la proteína que codifica cada una de las secuencias que conforman el alineamiento. En fig.3.1.5 se muestra la clasificación taxonómica de las 214 secuencias del alineamiento.



*fig.3.1.5. Esquema de la organización de las secuencias del alineamiento Swiss-Prot/BMP-2 según la clasificación taxonómica de los organismos a los que pertenecen las proteínas que codifican.*

La mayoría de las secuencias del alineamiento codifican proteínas pertenecientes a organismos con tejido óseo. Además no se encontraron diferencias significativas en los residuos más frecuentes del alineamiento para cada posición cuando realizamos el recuento en los diferentes alineamientos que resultan al agrupar las secuencias según su clasificación taxonómica, siempre y cuando dichos alineamientos contengan un número significativo de secuencias (el alineamiento con menos secuencias del que se calcularon los residuos de consenso fue aquel que contiene las secuencias que codifican proteínas pertenecientes a *Cetartiodactyla*, con 40 secuencias).

Por otra parte, puede ocurrir, que en el alineamiento exista un grupo numeroso de secuencias muy parecidas entre sí y diferentes a su vez de la secuencia *query*. La presencia de este grupo puede deberse a polimorfismos, a variantes diseñadas en un laboratorio sobre una proteína natural, al estudio exhaustivo de una misma proteína en diferentes organismos, etcétera... En cualquier caso, la sobre-representación de secuencias parecidas podría alterar el resultado del análisis de consenso. Con objeto de eliminar un posible grupo de secuencias muy parecidas entre sí, se filtró el alineamiento Swiss-Prot/BMP-2 según la **similaridad** entre sus secuencias. Con este propósito, se creó para las 214 secuencias del alineamiento, una matriz de distancias, donde las distancias se definen como 1 menos el índice de identidad para cada pareja de secuencias (distancias de Dayhoff) [22]. Usando el programa *K-means* [23, 24] incluido en MATLAB, se agruparon las secuencias en base a la matriz de distancias. Para ello, la función *K-means* origina los grupos de secuencias o *clusters* utilizando un algoritmo iterativo que minimiza la suma de las distancias existentes entre cada una de las secuencias. Como resultado, se obtienen diferentes *clusters* cumpliéndose que las secuencias de un mismo *cluster* son semejantes entre sí y diferentes a las de otro *cluster*. Para determinar el número más apropiado de *clusters* en los que se puede dividir el alineamiento se realizó el análisis *K-means* considerando diferentes números de

*clusters* (desde 2 hasta 5) y se calculó, para todas las posibilidades de agrupación, el valor de *Silhouette* de cada secuencia.

Para obtener el valor de *Silhouette* [25] se necesitan las divisiones obtenidas (los *clusters*) y las distancias entre cada secuencia (fig 3.1.6). Si consideramos una secuencia  $i$  asignada al *cluster* A y que el *cluster* A contiene más secuencias además de  $i$ :

$a(i)$  = media de las distancias entre la secuencia  $i$  y todas las demás secuencias del *cluster* A (en la fig 3.1.6 es la media de los segmentos rojos)

$d(i,C)$  = la media de las distancias entre la secuencia  $i$  y todas las secuencias del *cluster* C (en la fig 3.1.6 es la media de los segmentos azules)

$b(i)$  = valor mínimo de las distancias entre la secuencia  $i$  y las secuencias que pertenezcan al *cluster* C (en la fig 3.1.6 es el menor de los segmentos azules)

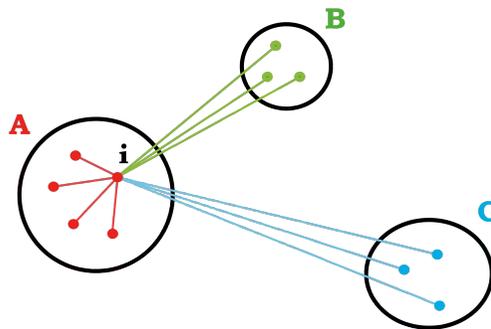


fig 3.1.6. Ilustración de los elementos involucrados en el cálculo de  $s(i)$  (valor de *Silhouette* para la secuencia  $i$ ), donde la secuencia  $i$  pertenece al *cluster* A. El *cluster* B sería la segunda mejor opción donde incluir el elemento  $i$ . Las distancias entre  $i$  y los demás secuencias del *cluster* A se representan en rojo,  $a(i)$  será la media de la longitud de los segmentos rojos. Las distancias entre  $i$  y las demás secuencias del *cluster* C se representan como segmentos azules,  $d(i, C)$  será la media de dichas distancias. Las distancias entre  $i$  y las demás secuencias del *cluster* B se representan como segmentos verdes, la media de dichas distancias debe ser igual al menor de los segmentos azules.

El valor de *Silhouette*,  $s(i)$ , se obtiene de la combinación de  $a(i)$  y  $b(i)$  de la siguiente manera:

$$s(i) = \begin{cases} 1 - a(i)/b(i) & \text{si } a(i) < b(i) \\ 0 & \text{si } a(i) = b(i) \\ b(i)/a(i) - 1 & \text{si } a(i) > b(i) \end{cases} \quad \text{ec. 3.1.1}$$

Reescribiendo la ecuación ec. 3.1.1 en una sola expresión:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad \text{ec. 3.1.2}$$

Para decidir si el número de *clusters* de secuencias obtenidos es el apropiado hemos de fijarnos en la gráfica del valor de *Silhouette* (fig 3.1.7):

- si  $a(i) < b(i)$  entonces  $s(i)$  tendrá un valor alto (cercano a 1), esto implica que las distancias dentro de un mismo *cluster* son más pequeñas que la menor de las distancias entre dos *cluster* diferentes, con lo cual la secuencia  $i$  está bien situada dentro del *cluster* A.
- si  $a(i) = b(i)$  entonces  $s(i) = 0$ , esto implica que las distancias entre un mismo *cluster* son iguales a la menor de las distancias entre dos *cluster* diferentes, con lo que no está claro si la secuencia  $i$  debe incluirse en el *cluster* A o en el B.
- si  $a(i) > b(i)$  entonces  $s(i)$  tendrá un valor bajo (cercano a -1), esto implica que las diferencias dentro del *cluster*  $a(i)$  son mayores que la menor de las diferencias entre dos *cluster* diferentes, con lo cual la secuencia  $i$  está mal situada dentro del *cluster* A.

Resumiendo, el número de *clusters* de secuencias es adecuado si estos se encuentran bien separados y las secuencias que lo forman poseen un valor alto de *Silhouette*.

En la fig 3.1.7, se representa el valor de *Silhouette* de cada secuencia, para las diferentes agrupaciones consideradas (de 2 a 5 *clusters*). Se decidió dividir el

alineamiento en dos *clusters* y eliminar el grupo de 39 secuencias que aparece como *cluster 1* en A, *cluster 2* en B y *cluster 4* en C y D, (fig 3.1.7) ya que sus secuencias tienen un valor de *Silhouette* muy alto, lo que significa que son muy parecidas entre sí y diferentes al resto de secuencias. Este grupo de secuencias, muy parecidas entre sí y diferentes a la secuencia *query*, representa el 20% del alineamiento, por lo que condiciona enormemente el análisis de consenso.

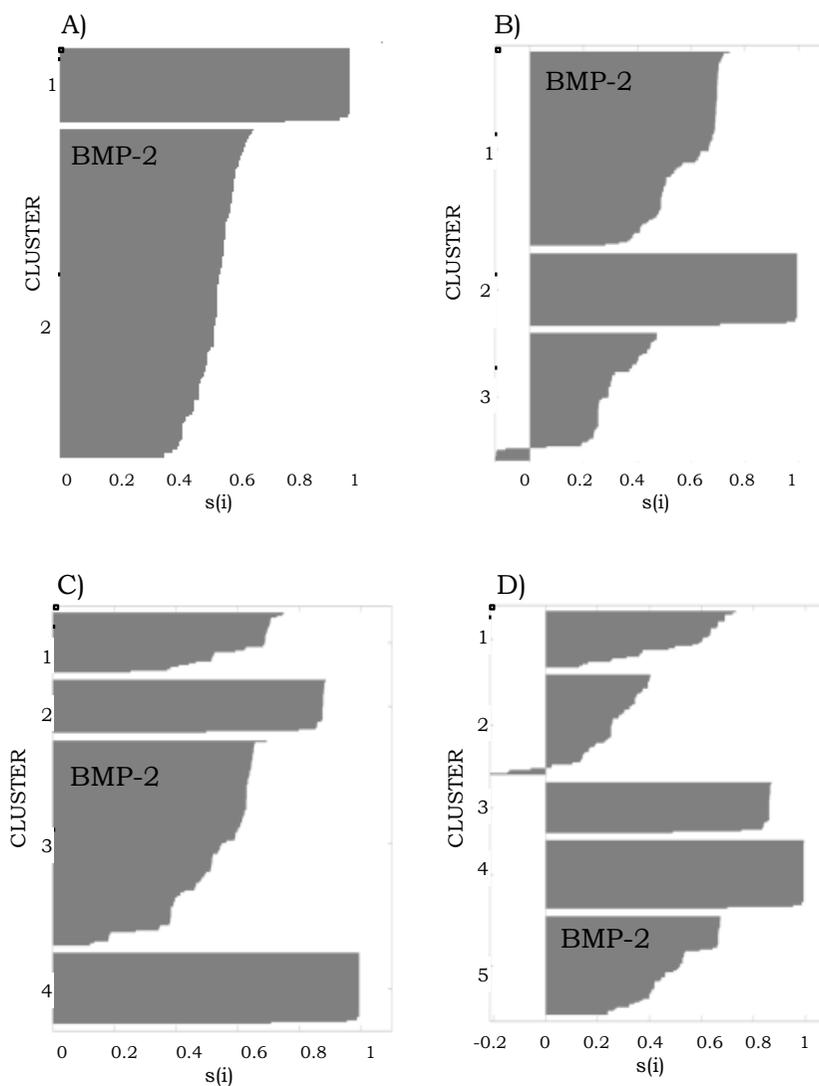


fig 3.1.7. Representación del valor de *Silhouette* para cada secuencia al dividir las 214 secuencias del alineamiento en A) 2 clusters, B) 3 clusters, C) 4 clusters y D) 5 clusters.

Si nos fijamos en las secuencias que contiene el *cluster* eliminado, vemos que de las 39 secuencias, 36 codifican la miostatina, también conocida como factor de crecimiento y diferenciación tipo 8 (GDF8) y tres codifican el factor de crecimiento y diferenciación tipo 11 (GDF11). La sobre-representación de la miostatina en el alineamiento podría deberse a que haya sido estudiada exhaustivamente y, como parte de su estudio, se haya secuenciado en numerosos organismos. Además, se sabe que la miostatina no induce la formación de hueso ni cartílago *in vivo* [6]. En fig 3.1.8 se muestra la secuencia representativa del *cluster* eliminado, el tamaño de los residuos es proporcional a la frecuencia de aparición en el alineamiento, como se puede observar, las secuencias que conforman este *cluster* son altamente parecidas entre sí.



fig 3.1.8. Representación del alineamiento de secuencias del *cluster* eliminado. El tamaño del código del aminoácido es proporcional a su frecuencia de aparición en el alineamiento. Las secuencias que conforman este *cluster* son muy parecidas entre sí (como se deduce del tamaño de los códigos que representan los aminoácidos en prácticamente todas las posiciones) y representan el 20% del alineamiento total, por lo que condicionan enormemente el análisis de consenso. Por esta razón, el análisis de consenso se realizó sobre las demás secuencias del alineamiento.

Del resto de secuencias del alineamiento, se seleccionaron aquellas que codifican proteínas de organismos pertenecientes a la subclase Mammalia y al orden Primates, y se realizó el análisis de consenso a los dos alineamientos resultantes.

Las mutaciones propuestas por el consenso en ambos casos fueron las mismas, así que decidimos trabajar sobre el alineamiento de secuencias que codifican proteínas de organismos del orden Primates.

Tras realizar el análisis de consenso se seleccionaron:

- Aquellas posiciones donde el aminoácido de consenso no coincidía con el aminoácido de la secuencia *query*.
- Dichas posiciones se filtraron a su vez por la frecuencia de aparición del aminoácido de consenso en el alineamiento, quedando seleccionadas aquellas posiciones donde el aminoácido de consenso aparecía en al menos el 40% de las secuencias.
- Por último, de las posiciones seleccionadas se eligieron aquellas donde el cociente  $\frac{\text{frecuencia}_{\text{consenso}}}{\text{frecuencia}_{\text{query}}} \geq 2$ .

En la tabla 3.1.1 se muestran las posiciones finalmente seleccionadas, los residuos *query* y de consenso para dichas posiciones y el cociente entre la frecuencia de aparición del aminoácido de consenso con respecto a la frecuencia de aparición del aminoácido *query*.

<b>Posición</b>	<b>aa<sub>query</sub></b>	<b>aa<sub>consenso</sub></b>	$\frac{\text{Fr}_{\text{consenso}}}{\text{Fr}_{\text{query}}}$
<b>26</b>	V	L	6.7
<b>33</b>	V	I	4.0
<b>41</b>	F	N	3.5
<b>83</b>	E	K	8.0
<b>86</b>	A	P	3.0
<b>97</b>	K	N	2.4

*tabla 3.1.1. En la columna 1 se muestran las posiciones seleccionadas en base al criterio anteriormente explicado, en la columna 2 el aminoácido query, en la columna 3 el aminoácido más frecuente en el alineamiento y finalmente en la columna 4 se muestra cuan más frecuente es el aminoácido de consenso con respecto al aminoácido query.*

En base a los criterios anteriormente explicados, las mutaciones que finalmente se decidieron realizar son: V26L, V33I, F41N, E83K, A86P y K97N y se muestran etiquetadas en la estructura del dímero maduro de la BMP-2 (fig 3.1.9).

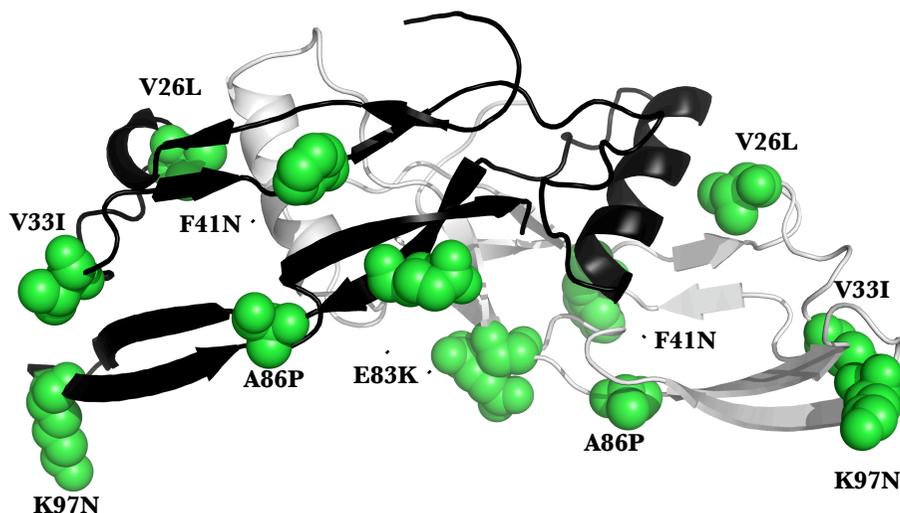


fig 3.1.9. Estructura de la proteína morfogenética de hueso 2 (3bmp.pdb [14]). Un monómero se muestra en negro y el otro monómero en gris. Las posiciones seleccionadas por el análisis bioinformático realizado se muestran como esferas verdes y están etiquetadas con la mutación propuesta.

### 3.1.2. Estudio de la capacidad de diferenciación ósea de BMP-2 silvestre y variantes de BMP-2

Con las mutaciones seleccionadas (V26L, V33I, F41N, E83K, A86P y K97N) se han construido variantes simples sobre BMP-2. Las variantes se han obtenido por mutagénesis dirigida (sección 2.1) con el gen de la BMP-2 insertado en el vector pT7.7. La transformación, expresión y purificación de la proteína se ha explicado en las secciones 2.4.1, 2.4.3 y 2.4.4.3.

Los resultados que se muestran a continuación corresponden a ensayos realizados sobre la línea celular C2C12 [26], inicialmente aislada de tejido muscular de ratón adulto (sección 2.11). La implantación de BMP-2 en tejido muscular induce la formación ósea, por lo que este modelo celular es uno de los más utilizados en

estudios con esta proteína. Además, las células C2C12 son de fácil manejo y proliferan rápidamente, lo que facilita el trabajo de investigación. Al producirse la diferenciación de la línea celular mioblástica C2C12 hacia un linaje osteoblástico, se inhibe la expresión de proteínas propias del fenotipo muscular y se induce la expresión de proteínas propias del fenotipo óseo tales como la fosfatasa alcalina. Esta enzima es un marcador temprano de la diferenciación ósea y su expresión basal en las células C2C12 es nula, por ello se utilizó este marcador para la evaluación de la respuesta celular a la BMP-2.

Inicialmente, se estudiaron los mutantes simples V26L, V33I, F41N, E83K, A86P y K97N. Se testó la capacidad de dichos mutantes para diferenciar la línea celular mioblástica C2C12 a un linaje osteoblástico mediante el ensayo de fosfatasa alcalina (sección 2.11.2 de metodología). En fig 3.1.10 y fig 3.1.11 se muestran los resultados de los ensayos de diferenciación celular para las formas mutantes simples. Las células C2C12 mostraron diferentes resultados de diferenciación dependiendo del número de pases celulares al que se hayan sometido, por ello, todos los ensayos se realizaron con células de menos de 10 pases celulares (explicado en la sección 2.11.1 de metodología). En cada tanda de ensayos siempre se testó la BMP-2 silvestre con objeto de analizar los datos de los mutantes comparativamente.

Para descartar que las diferencias de actividad fosfatasa alcalina se debieran a un desigual crecimiento celular, se cuantificó mediante Bradford la cantidad de proteínas totales, y se tomó el valor obtenido como una estimación de la proliferación celular (sección 2.11.3 de metodología). En fig 3.1.12 se muestran, a modo de ejemplo, los resultados del ensayo Bradford para la BMP-2 silvestre y para los mutantes V26L y V33I, en todos los casos las proteínas totales resultaron  $\approx 10 \mu\text{g}$  lo que indica que la modulación en la capacidad diferenciadora de BMP-2 silvestre, V26L y V33I (fig 3.1.11) no se debe a diferencias en el crecimiento

celular. Al tratar las células C2C12 con el resto de variantes de BMP-2, el ensayo de Bradford mostró una proliferación celular semejante.

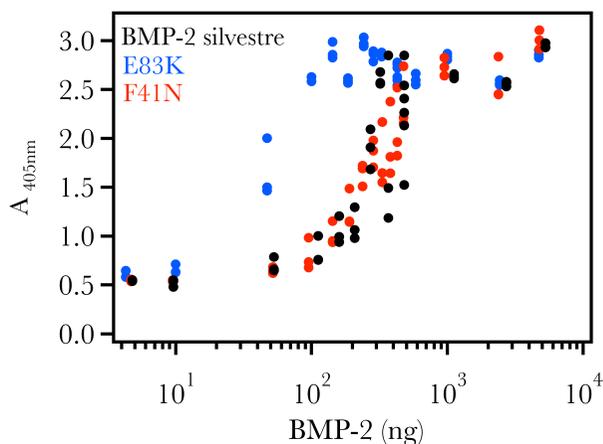


fig 3.1.10. Ensayo de diferenciación celular a diferentes dosis de BMP-2 silvestre (círculos rellenos negros) y de las proteínas mutantes E83K (círculos rellenos azules) y F41N (círculos rellenos rojos). Los ensayos de fosfatasa alcalina se realizaron por triplicado tras cuatro días de incubación de las C2C12 con BMP-2 silvestre o con las proteínas mutantes.

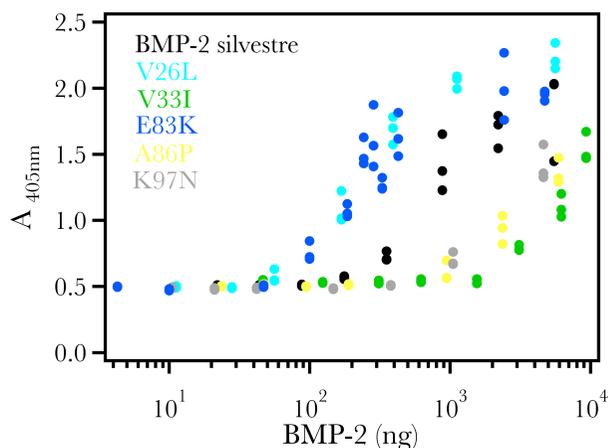


fig 3.1.11. Ensayo de diferenciación celular a diferentes dosis de BMP-2 silvestre (círculos rellenos negros) y de las proteínas mutantes V26L (círculos rellenos celeste), V33I (círculos rellenos verdes), E83K (círculos rellenos azules), A86P (círculos rellenos amarillos) y K97N (círculos rellenos grises). Los ensayos de fosfatasa alcalina se realizaron por triplicado tras cuatro días de incubación de las C2C12 con BMP-2 silvestre o con las proteínas mutantes.

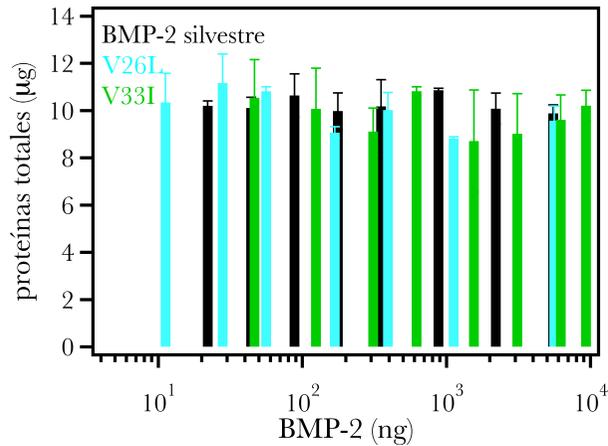


fig 3.1.12. Ensayo de respuesta celular a diferentes dosis de BMP-2 silvestre (barras negras) y de las proteínas mutantes V26L (barras celeste) y V33I (barras verdes). La masa de proteínas totales se cuantificó mediante el ensayo de Bradford para estimar la proliferación celular. Los ensayos se realizaron por triplicado tras cuatro días de incubación de las C2C12 con BMP-2 silvestre o con las formas mutantes V26L y V33I. Las barras de error representan la desviación estándar asociada a la realización por triplicado de los ensayos.

Tras observar la modulación de la capacidad diferenciadora de las formas mutantes simples ensayadas, se decidieron estudiar otras formas mutantes múltiples. Al ser el mutante E83K el más activo de los estudiados, se eligió como punto de partida para la construcción de nuevos mutantes dobles. Además también se decidió estudiar el mutante V33I/A86P ya que combina las formas mutantes simples que resultaron ser menos activas. Suponiendo que el efecto de las mutaciones fuese aditivo podríamos obtener un amplio rango de modulación de la actividad diferenciadora. Se obtuvieron las formas mutantes V26L/E83K, V33I/E83K, F41N/E83K, E83K/A86P, E83K/K97N y V33I/A86P por mutagénesis dirigida como se ha detallado en la sección 2.1. La transformación, expresión y purificación de las nuevas formas se realizó de manera semejante a la proteína BMP-2 silvestre tal y como se ha indicado en las secciones 2.4.1, 2.4.3 y 2.4.4.3. En fig 3.1.13 y fig 3.1.14 se muestran los resultados de los ensayos de diferenciación celular para las formas mutantes dobles.

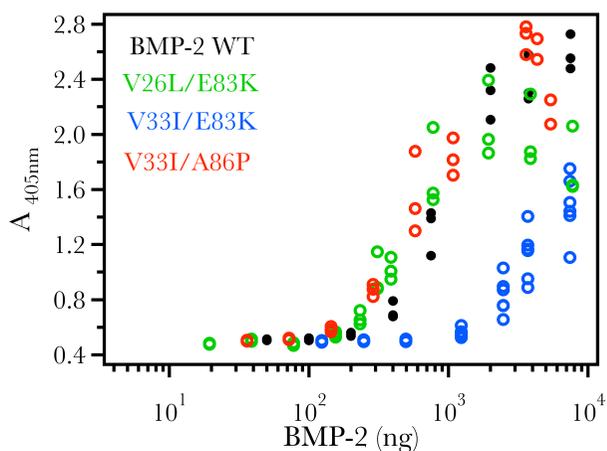


fig 3.1.13. Ensayo de diferenciación celular a diferentes dosis de BMP-2 silvestre (círculos rellenos negros) y de las proteínas mutantes V26L/E83K (círculos vacíos verdes), V33I/A86P (círculos vacíos rojos) y V33I/E83K (círculos vacíos azules). Los ensayos de fosfatasa alcalina se realizaron por triplicado tras cuatro días de incubación de las C2C12 con BMP-2 silvestre o con las proteínas mutantes.

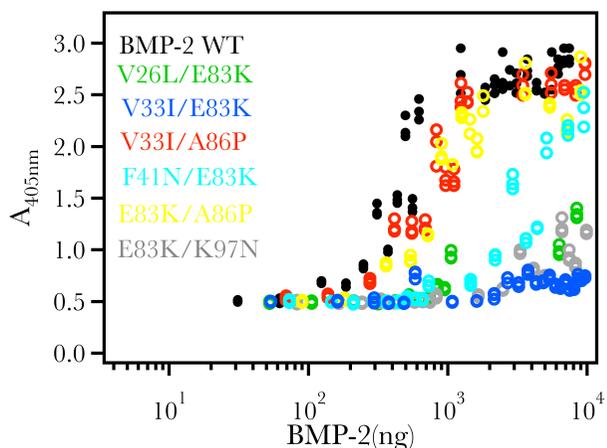


fig 3.1.14. Ensayo de diferenciación celular a diferentes dosis de BMP-2 silvestre (círculos rellenos negros) y de las proteínas mutantes V26L/E83K (círculos vacíos verdes), V33I/E83K (círculos vacíos azules), V33I/A86P (círculos vacíos rojos), F41N/E83K (círculos vacíos celestes), E83K/A86P (círculos vacíos amarillos) y E83K/K97N (círculos vacíos grises). Los ensayos de fosfatasa alcalina se realizaron por triplicado tras cuatro días de incubación de las C2C12 con BMP-2 silvestre o con las proteínas mutantes.

Como ya se ha indicado, en cada tanda de ensayos, se testó como control la actividad fosfatasa alcalina del lisado celular de las C2C12 tras haber sido incubadas con BMP-2 silvestre. Los resultados de las variantes se analizaron en términos comparativos con respecto la actividad obtenida para el control. El análisis conjunto de los resultados reveló cierta irreproducibilidad. A modo de ejemplo se muestra en fig 3.1.15 los resultados de la variante E83K y del control realizados en dos tandas diferentes de experimentos. Como se puede observar, la variante E83K resultó más activa que el control en ambos casos, pero la mejora de la actividad resultó mayor en la tanda de experimentos realizados en primer lugar (fig 3.1.15 A) que en la repetición del experimento (fig 3.1.15 B).

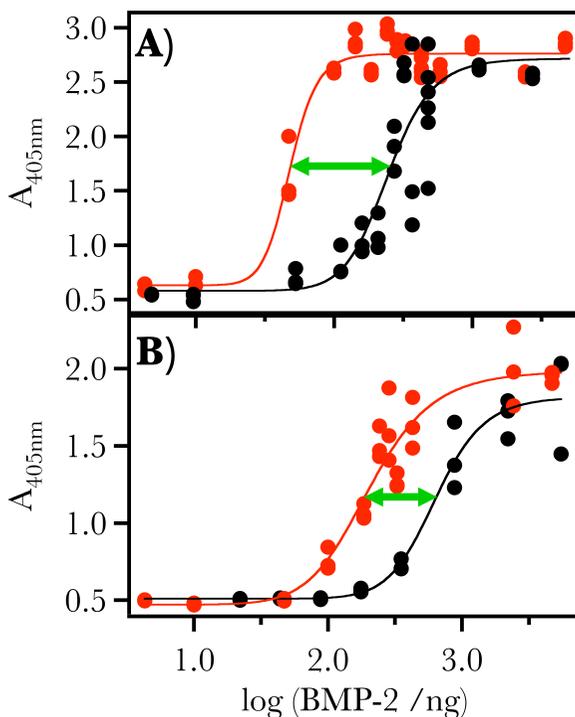


fig 3.1.15. Ensayo de diferenciación celular a diferentes dosis de BMP-2 silvestre (círculos rellenos negros) y a diferentes dosis de la variante E83K/BMP-2 (círculos rellenos rojos). Las líneas continuas no tienen significado teórico, sólo sirven para guiar la vista. En A) y B) se muestran los resultados obtenidos para las mismas proteínas en diferentes tandas de experimentos. La longitud de la flecha verde indica la diferencia de actividad del mutante E83K con respecto a la proteína control, BMP-2 silvestre.

Como hemos dicho anteriormente, la BMP-2 en su forma madura es una proteína homodimérica donde los monómeros se unen covalentemente por un puente disulfuro entre la Cys78 de cada subunidad [14], cada monómero a su vez

presenta una estructura de nudo cistina constituida por otros tres puentes disulfuro. Este entramado tan complejo de puentes disulfuro no se pliega correctamente en un sistema de expresión bacteriano y necesita de una etapa de replegamiento *in vitro* (sección 2.4.4.3 de metodología). La irreproducibilidad de los resultados realizados hizo cuestionar dicha etapa de replegamiento *in vitro* para las formas mutantes, ya que la expresión y purificación fue puesta a punto por el grupo de Ingeniería Tisular de la Universidad Complutense de Madrid a cargo de José Luis López Lacomba para la obtención de BMP-2 silvestre y no para las formas mutantes. Por ello, en colaboración con el grupo de J.L López Lacomba se está trabajando en el desarrollo de un nuevo proceso de purificación en un sistema de expresión eucariota que permita obtener la proteína correctamente plegada. Este trabajo, actualmente en progreso, consiste en la puesta a punto de la expresión y purificación de la BMP-2 en células de ovario de hámster chino (CHO) donde la BMP-2 se plegará de acuerdo con el esquema que se muestra en fig 3.1.2 [27]. Obtener las variantes de BMP-2 plegadas correctamente desde su sistema de expresión, reduciría el tiempo y el coste de producción de las proteínas. Esto nos permitiría estudiar un número mayor de variantes y además abre la posibilidad al estudio de bibliotecas de variantes.

### **3.1.3. Caracterización biofísica de BMP-2 silvestre y variantes de BMP-2**

#### **3.1.3.1. Medidas de calorimetría diferencial de barrido**

Al ser la BMP-2 insoluble a valores neutros y básicos de pH, la caracterización termodinámica de su desplegamiento queda limitada a valores ácidos, o al uso de agentes que eviten la agregación de la proteína en un rango de pH cercano a las condiciones fisiológicas donde realiza su función *in vivo*.

Para el estudio a pH ácido, se dializó la BMP-2 silvestre frente a 50 mM ácido acético en un rango de pH de 2.7 a 3.4. Los experimentos de DSC se realizaron a

una velocidad de barrido de 150 K/h, de 10 a 120°C a diferentes concentraciones de proteína (0.6-0.8 mg/mL). Como se muestra en fig 3.1.16, en ningún caso se observó transición calorimétrica con un área que razonablemente se correspondiera con lo esperable para una proteína de 26 kDa [28].

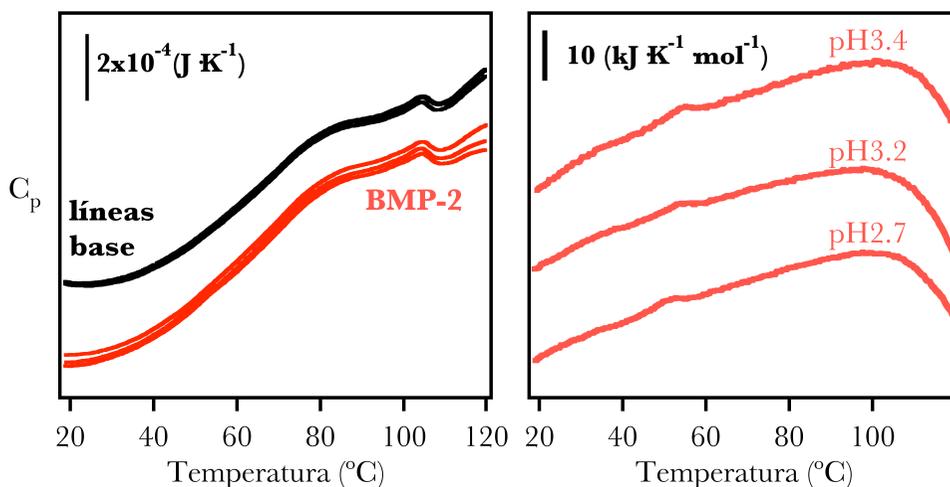


fig 3.1.16. A) Perfiles de capacidad calorífica frente a temperatura, en negro se muestran las líneas base y en rojo los barridos correspondientes a disoluciones de BMP-2 en 50 mM ácido acético a diferentes pHs en un rango de 2.7-3.4. B) Perfiles normalizados de capacidad calorífica frente a temperatura de disoluciones de BMP-2 en 50 mM ácido acético a una concentración de 0.6 mg/mL a pH=3.4, de 0.8 mg/mL a pH=3.2 y de 0.8 mg/mL a pH=2.7.

Por limitaciones de solubilidad, los experimentos realizados a un pH cercano al fisiológico requerían o el uso de disoluciones con alta concentración de agente desnaturante o con agentes que evitasen la agregación de la proteína.

Se prepararon disoluciones de BMP-2 a una concentración de 0.4 mg/mL, en 50 mM HEPES pH 7.5 y diferentes concentraciones de cloruro de guanidina en un rango de concentración de 2.8 a 6.9 M. Los barridos de DSC se realizaron a 90K/h, de 10 a 120°C y de nuevo no se observó transición calorimétrica. Ante la posibilidad de que la BMP-2 fuera tan estable que requiriese un agente

desnaturalizante más fuerte que el cloruro de guanidina, se prepararon disoluciones de BMP-2 a una concentración de 0.4 mg/mL, en 50 mM HEPES pH 7.5 y una concentración de 5.5 M de tiocianato de guanidina [29]. Los barridos de DSC tampoco mostraron transición alguna.

Está descrito que la arginina afecta a las interacciones proteína-proteína evitando su agregación. La arginina aumenta la tensión superficial del agua lo que favorece su interacción con la cadena lateral de los aminoácidos y con el enlace polipeptídico [30]. El uso de arginina, nos permitió disolver la BMP-2 en un tampón a pH neutro. Se preparó una disolución de BMP-2 a 0.4 mg/mL en HEPES 100 mM, arginina 1 M a pH 7.0. Los barridos realizados a 150 K/h, de 10 a 120°C no mostraron transición alguna.

Al no observarse en ninguna de las condiciones estudiadas indicios del desplegamiento de la BMP-2, no se pudo caracterizar su estabilidad por calorimetría diferencial de barrido y, por ello, nos propusimos su caracterización mediante estudios de proteólisis.

### **3.1.3.2. Estudios de proteólisis de variantes de BMP-2**

Varios trabajos muestran cómo sometiendo una proteína a incubación con proteasas y analizando el lisado final, o bien por SDS-PAGE [31, 32] o bien por DSC [33], se puede obtener información de la estabilidad termodinámica o cinética de la proteína en cuestión. Al no haber obtenido por calorimetría medida alguna de estabilidad, como alternativa para la caracterización de la estabilidad de BMP-2 silvestre y de las variantes de BMP-2, nos propusimos un estudio proteolítico con termolisina.

En el trabajo que se describe a continuación se han sometido a pulsos de termolisina muestras de variantes de BMP-2 con objeto de caracterizar la estabilidad cinética de dichas proteínas. Esta técnica consume muy poca cantidad de muestra por lo que resulta de gran utilidad para el estudio de proteínas de

elevado coste de producción, como es el caso que nos atañe. La termolisina cataliza la hidrólisis del enlace peptídico formado preferentemente por aminoácidos con cadenas laterales grandes e hidrofóbicas tales como la fenilalanina o la leucina [34, 35], en fig 3.1.17 se muestran los posibles puntos de corte de la proteína BMP-2 con termolisina. La proteína será digerida con la exposición al medio de dichas zonas de corte que frecuentemente son secuencias enterradas en la estructura de la proteína [36].

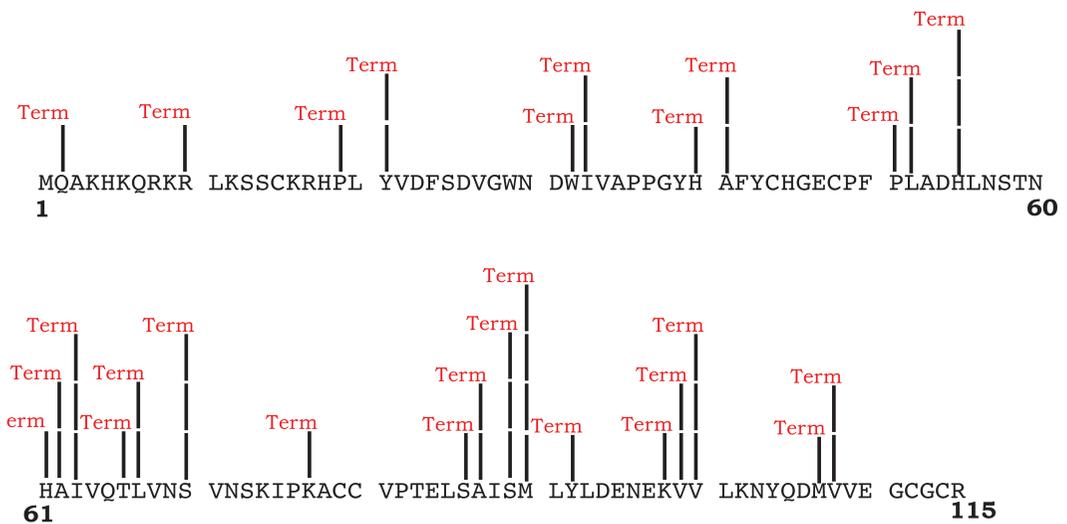


fig 3.1.17. Secuencia monomérica de BMP-2 donde se muestran los sitios de corte de termolisina.

Los estudios de proteólisis se realizaron como se ha descrito en el apartado 2.9 del capítulo de metodología. Para la BMP-2 silvestre los experimentos se realizaron en un rango de concentraciones de termolisina de 0.01-0.1 mg/mL y para las variantes de BMP-2 a 0.01 mg/mL de termolisina. La concentración de proteína fue de ~0.25 mg/mL en 100 mM Acético, 10 mM CaCl<sub>2</sub> a pH=5.5 y la digestión se realizó a una temperatura de 25°C.

En fig 3.1.18 y fig 3.1.19 se muestran, a modo de ejemplo, los gels del análisis proteolítico y el perfil de intensidad de banda frente a tiempo para las variantes

V26L y E83K/K97N, ambas formas mutantes son las que resultaron tener mayor y menor capacidad para diferenciar a linaje osteoblástico las células C2C12. El ajuste a una ecuación biexponencial del perfil de digestión de la variante V26L (fig 3.1.18) mostró como el 45% de la población era digerida en un tiempo de vida media de 70 min y el 55% de población restante era digerido en un tiempo de vida media de 60 segundos. El perfil de digestión de la variante E83K/K97N (fig 3.1.19) también se ajustó a una ecuación biexponencial, dicho ajuste mostró cómo el 22% de la población era digerido en un tiempo de vida media de 5h, mientras que el 73% restante era digerido en un tiempo de vida media de 28 min.

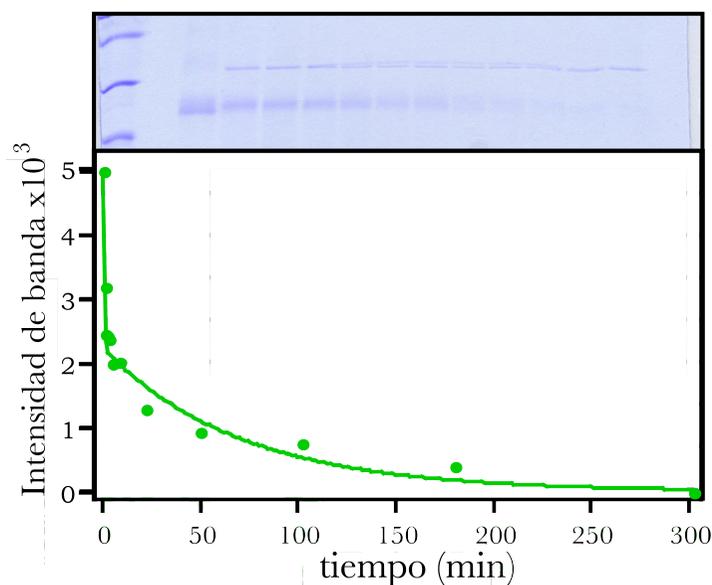


fig 3.1.18. Fragmento del gel de acrilamida que muestra la disminución en la intensidad de la banda correspondiente al mutante V26L al ser digerida por la termolisina. La gráfica muestra la intensidad de la banda (analizada con el programa ImageJ) frente al tiempo de incubación. En puntos verdes se muestran los datos experimentales y en línea continua el ajuste a una ecuación biexponencial. La concentración de termolisina fue de 0.01 mg/mL y la concentración de V26L fue de 0.27 mg/mL en 100 mM Acético, 10 mM CaCl<sub>2</sub> a pH=5.5, la digestión se realizó a una temperatura de 25°C.

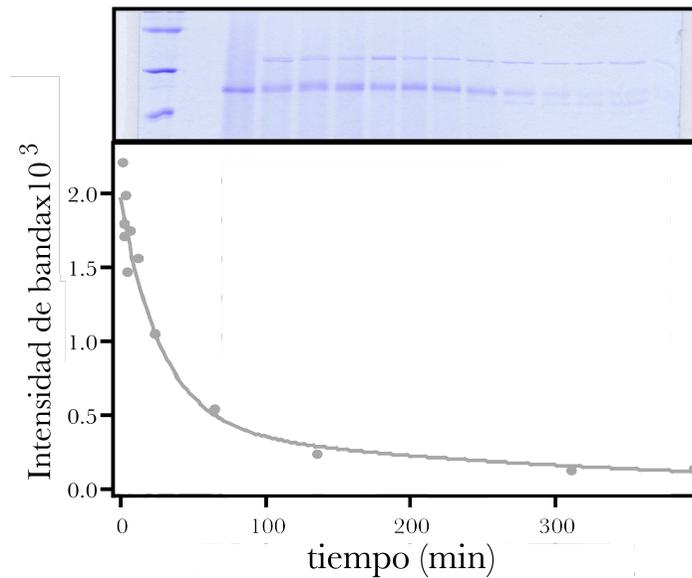


fig 3.1.19. Fragmento del gel de acrilamida que muestra la disminución en la intensidad de la banda correspondiente al mutante E83K/K97N de BMP-2 a una concentración de 0.27 mg/mL debido a su digestión por la termolisina. La gráfica muestra la intensidad de la banda (analizada con el programa ImageJ) frente al tiempo de incubación. En puntos grises se muestran los datos experimentales y en línea continua el ajuste a una ecuación biexponencial. La concentración de termolisina fue de 0.01 mg/mL y la concentración de E83K/K97N fue de 0.27 mg/mL en 100 mM Acético, 10 mM CaCl<sub>2</sub> a pH=5.5, la digestión se realizó a una temperatura de 25°C.

Se analizaron todas las variantes de BMP-2 y el perfil de intensidad de banda *versus* tiempo se ajustó a las ecuaciones ec.2.70 y ec.2.71. Analizando los resultados se concluyó que las muestras estudiadas contenían varias poblaciones de proteína plegada de diferente forma, ya que los perfiles obtenidos indicaban una población de proteína que se digería en un tiempo medio muy corto, y otra población, que tardaba mucho más tiempo en ser digerida.

La heterogeneidad en las muestras de proteína se resolverá con la puesta a punto del proceso de purificación en el sistema de expresión eucariota en el que se está trabajando en colaboración con el grupo de Ingeniería Tisular de la Universidad Complutense de Madrid a cargo de J.L López Lacomba.



### 3.1.4. Bibliografía

1. Urist, M.R., *Bone: formation by autoinduction*. Science, 1965. **150**(3698): p. 893-9.
2. Urist, M.R. and B.S. Strates, *Bone morphogenetic protein*. Journal of dental research, 1971. **50**(6): p. 1392-406.
3. Burt, D.W., *Evolutionary grouping of the transforming growth factor-beta superfamily*. Biochemical and biophysical research communications, 1992. **184**(2): p. 590-5.
4. Herpin, A., C. Lelong, and P. Favrel, *Transforming growth factor-beta-related proteins: an ancestral and widespread superfamily of cytokines in metazoans*. Developmental and comparative immunology, 2004. **28**(5): p. 461-85.
5. Hogan, B.L., *Bone morphogenetic proteins: multifunctional regulators of vertebrate development*. Genes & development, 1996. **10**(13): p. 1580-94.
6. Miyazono, K. and T. Shimanuki, *Chapter 55 - Bone Morphogenetic Protein Receptors and Actions*, in *Principles of Bone Biology (Third Edition)*, P.B. John, et al., Editors. 2008, Academic Press: San Diego. p. 1177-1196.
7. Cochran, D.L., et al., *Evaluation of recombinant human bone morphogenetic protein-2 in oral applications including the use of endosseous implants: 3-year results of a pilot study in humans*. Journal of periodontology, 2000. **71**(8): p. 1241-57.
8. Govender, S., et al., *Recombinant human bone morphogenetic protein-2 for treatment of open tibial fractures: a prospective, controlled, randomized study of four hundred and fifty patients*. The Journal of bone and joint surgery. American volume, 2002. **84-A**(12): p. 2123-34.
9. Einhorn, T.A., et al., *A single percutaneous injection of recombinant human bone morphogenetic protein-2 accelerates fracture repair*. The Journal of bone and joint surgery. American volume, 2003. **85-A**(8): p. 1425-35.

10. Granjeiro, J.M., et al., *Bone morphogenetic proteins: from structure to clinical use*. Brazilian journal of medical and biological research = Revista brasileira de pesquisas medicas e biologicas / Sociedade Brasileira de Biofisica ... [et al.], 2005. **38**(10): p. 1463-73.
11. M. Kawabata, K.M., *Bone morphogenetic proteins*. E. Canalis (Ed.), *Skeletal Growth Factors*, Lippincott Williams & Wilkins, 2000: p. 269-290.
12. Yano, K., et al., *Osteoinductive capacity and heat stability of recombinant human bone morphogenetic protein-2 produced by Escherichia coli and dimerized by biochemical processing*. Journal of bone and mineral metabolism, 2009. **27**(3): p. 355-63.
13. Bessa, P.C., et al., *Osteoinduction in human fat-derived stem cells by recombinant human bone morphogenetic protein-2 produced in Escherichia coli*. Biotechnology letters, 2008. **30**(1): p. 15-21.
14. Scheufler, C., W. Sebald, and M. Hulsmeyer, *Crystal structure of human bone morphogenetic protein-2 at 2.7 Å resolution*. Journal of molecular biology, 1999. **287**(1): p. 103-15.
15. Steipe, B., et al., *Sequence statistics reliably predict stabilizing mutations in a protein domain*. Journal of molecular biology, 1994. **240**(3): p. 188-92.
16. Rodriguez-Larrea, D., et al., *Role of conservative mutations in protein multi-property adaptation*. The Biochemical journal, 2010. **429**(2): p. 243-9.
17. Lehmann, M., et al., *The consensus concept for thermostability engineering of proteins: further proof of concept*. Protein engineering, 2002. **15**(5): p. 403-11.
18. Altschul, S.F., et al., *Basic local alignment search tool*. Journal of molecular biology, 1990. **215**(3): p. 403-10.
19. Smith, T.F., M.S. Waterman, and W.M. Fitch, *Comparative biosequence metrics*. Journal of molecular evolution, 1981. **18**(1): p. 38-46.

20. Smith, T.F. and M.S. Waterman, *Identification of common molecular subsequences*. Journal of molecular biology, 1981. **147**(1): p. 195-7.
21. Dokholyan, N.V. and E.I. Shakhnovich, *Understanding hierarchical protein evolution from first principles*. Journal of molecular biology, 2001. **312**(1): p. 289-307.
22. Dayhoff, M.O., R.M. Schwartz, and B.C. Orcutt, *Atlas of Protein Sequence and Structure*. 1978. **5**: p. 354-352.
23. Forgy, E.W., *Cluster Analysis of Multivariate Data - Efficiency Vs Interpretability of Classifications*. Biometrics, 1965. **21**(3): p. 768-&.
24. Jancey, R.C., *Multidimensional Group Analysis*. Australian Journal of Botany, 1966. **14**(1): p. 127-&.
25. Rousseeuw, P.J., *Silhouettes: a graphical aid to the interpretation and validation of cluster analysis*. Journal of Computational and Applied Mathematics, 1986. **20**: p. 53-65.
26. Katagiri, T., et al., *Bone morphogenetic protein-2 converts the differentiation pathway of C2C12 myoblasts into the osteoblast lineage*. The Journal of cell biology, 1994. **127**(6 Pt 1): p. 1755-66.
27. Israel, D.I., et al., *Expression and characterization of bone morphogenetic protein-2 in Chinese hamster ovary cells*. Growth factors, 1992. **7**(2): p. 139-50.
28. Robertson, A.D. and K.P. Murphy, *Protein Structure and the Energetics of Protein Stability*. Chemical reviews, 1997. **97**(5): p. 1251-1268.
29. Perez-Jimenez, R., et al., *The efficiency of different salts to screen charge interactions in proteins: a Hofmeister effect?* Biophysical journal, 2004. **86**(4): p. 2414-29.
30. Arakawa, T., et al., *Suppression of protein interactions by arginine: a proposed mechanism of the arginine effects*. Biophysical chemistry, 2007. **127**(1-2): p. 1-8.

31. Park, C. and S. Marqusee, *Probing the high energy states in proteins by proteolysis*. Journal of molecular biology, 2004. **343**(5): p. 1467-76.
32. Park, C. and S. Marqusee, *Pulse proteolysis: a simple method for quantitative determination of protein stability and ligand binding*. Nature methods, 2005. **2**(3): p. 207-12.
33. Tur-Arlandis, G., et al., *Proteolytic scanning calorimetry: a novel methodology that probes the fundamental features of protein kinetic stability*. Biophysical journal, 2010. **98**(6): p. L12-4.
34. Heinrikson, R.L., *Applications of thermolysin in protein structural analysis*. Methods in enzymology, 1977. **47**: p. 175-89.
35. Pank, M., et al., *Hydrophobic interaction in thermolysin specificity*. FEBS letters, 1982. **142**(2): p. 297-300.
36. Tyndall, J.D., T. Nall, and D.P. Fairlie, *Proteases universally recognize beta strands in their active sites*. Chemical reviews, 2005. **105**(3): p. 973-99.

### **3.2. Estudio de tiorredoxinas precámbricas resucitadas**

La posibilidad de viajar en el tiempo, a nivel molecular, para reconstruir secuencias de proteínas de organismos ya extinguidos fue propuesta por Linus Pauling y Emile Zuckerkandl en los años 50. Estos autores, sugirieron que mediante el análisis comparativo de secuencias de proteínas actuales, se podrían inferir las secuencias de proteínas relacionadas de organismos ancestrales [1]. Sin embargo, dicha idea no se materializó hasta la era de la genómica, a mediados de los años noventa, gracias al avance de la bioinformática y la posibilidad de disponer de bases de datos con una gran cantidad de secuencias de proteínas [2-5].

Desde entonces, existen algunos trabajos que acreditan la obtención de proteínas ancestrales y que han revelado información sobre el entorno físico que rodeaba a los organismos extintos, así como información sobre su adaptación a los cambios en ese entorno [6-9].

Además, el estudio de las proteínas ancestrales ha mostrado que tienen características diferentes a las de las proteínas actuales, como muy alta estabilidad y, en algunos casos, niveles incrementados de su función, que son muy interesantes desde el punto de vista de la ingeniería de proteínas [6, 10, 11]. La resurrección ancestral, por tanto, se revela como una de las principales metodologías para el desarrollo de proteínas modificadas con aplicaciones biotecnológicas y biomédicas.

Como resultado de un trabajo colaborativo entre los grupos de investigación de Eric A. Gaucher (Georgia Institute of Technology, Atlanta), de Julio M. Fernández (Universidad de Columbia, Nueva York) y de José M. Sánchez Ruiz (Universidad de Granada), se han resucitado y caracterizado siete tiorredoxinas.

Eric A. Gaucher (Georgia Institute of Technology, Atlanta) llevó a cabo la reconstrucción de las secuencias de las tiorredoxinas ancestrales pertenecientes a siete nodos de un árbol filogenético, que fue construido utilizando 203 secuencias de tiorredoxinas actuales representativas de los tres dominios de la vida. El árbol

filogenético reconstruido, que incluye los nodos extintos resucitados, se muestra en fig 3.2.1.

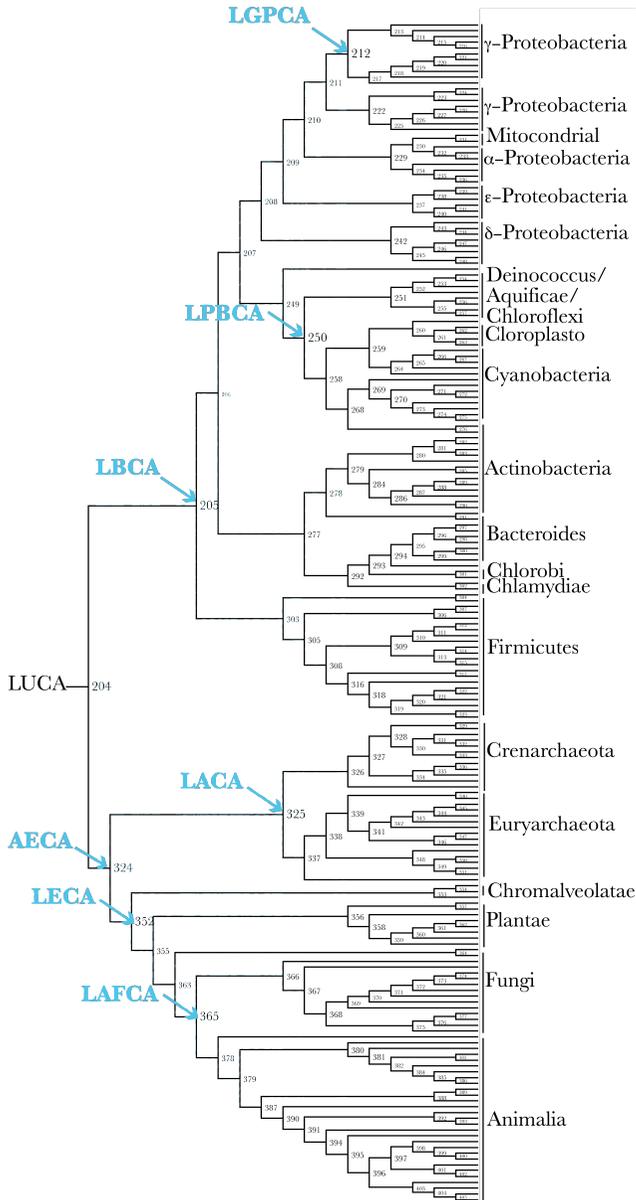


fig 3.2.1. *Árbol filogenético construido para la obtención de las secuencias asociadas a los nodos extintos que se muestran en azul. Se han utilizado un total de 203 secuencias de tioredoxinas actuales representativas de los tres dominios de la vida (Bacteria, Archaea y Eukarya). Los nodos de interés se indican con flechas: el último ancestro común de bacterias (LBCA), el último ancestro común de arquea (LACA), el ancestro común de arquea y eucariota (AECA), el último ancestro común de eucariotas (LECA), el último ancestro común de cianobacterias y los grupos de deinococcus y thermus (LPBCA), el último ancestro común de  $\gamma$ -proteobacteria (LGPCA) y el último ancestro común de animales y hongos (LAFCA). Los números que aparecen se corresponden a los nodos [6].*

En fig 3.2.2, se muestra el alineamiento múltiple con las secuencias que, con mayor probabilidad, se corresponden con las tioredoxinas de los organismos extintos asociados a ciertos nodos del árbol filogenético. También se ha incluido

en el alineamiento, las secuencias de la tiorredoxina de *E. coli* y la tiorredoxina humana.

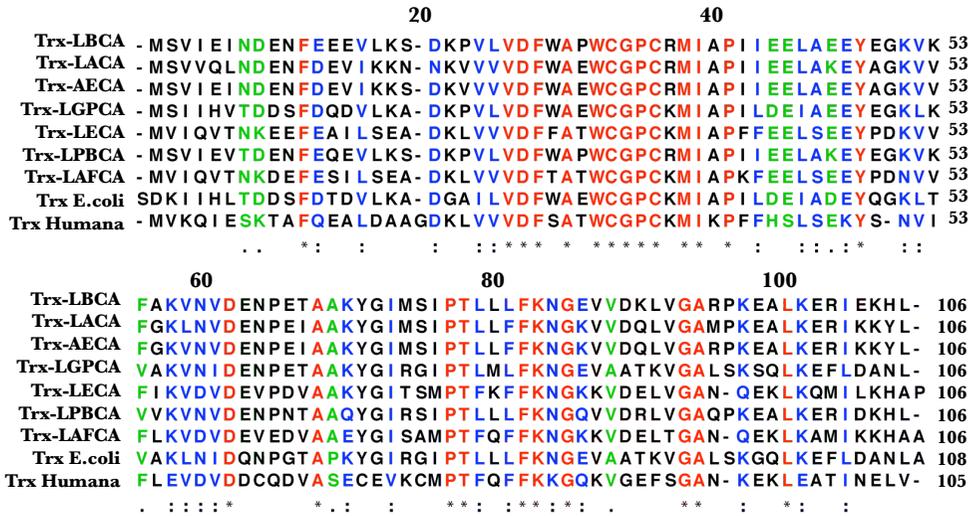


fig 3.2.2. Alineamiento múltiple de las secuencias de tiorredoxinas ancestrales perteneciente a organismos de nodos ya extintos y las tiorredoxinas actuales de *E. coli* y humana. Se indican en rojo los residuos conservados, como se puede observar el sitio activo, CGPC, está altamente conservado. Los residuos en azul y verde, representan sustituciones por residuos similares.

En nuestro grupo de investigación, en la Universidad de Granada, se resucitaron experimentalmente las siete tiorredoxinas precámbricas y se llevó a cabo una caracterización biofísica exhaustiva de las mismas. Cabe resaltar que estas enzimas ancestrales son las más antiguas resucitadas hasta la fecha, datando la tiorredoxina del nodo asociado al último ancestro común de las bacterias (LBCA de sus siglas en inglés) de hace aproximadamente 4200 millones de años. Así, hemos resucitado proteínas que operaban en un tiempo muy cercano al origen de la vida. En cuanto a la caracterización biofísica, el estudio de la termoestabilidad de las tiorredoxinas ancestrales puso de manifiesto la extrema estabilidad que presentaban estas proteínas [6]; siendo, hasta lo que sabemos, de las más estables conocidas hasta la fecha (fig 3.2.3).

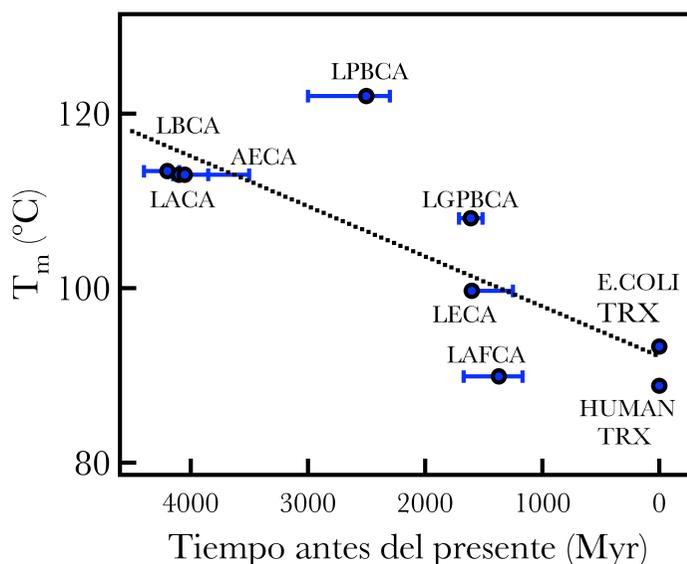
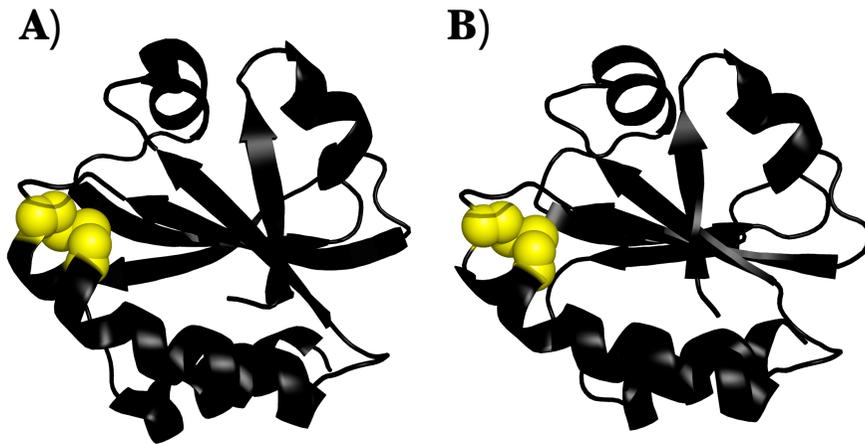


fig 3.2.3. Temperatura de desnaturalización ( $T_m$ ) versus tiempo geológico para las tiorredoxinas precámbricas y para las tiorredoxinas actuales humana y de *E. coli*. La línea discontinua representa el ajuste lineal de los datos mostrando una tendencia de paleotemperatura con un decrecimiento en  $T_m$  de  $5.8 \pm 1.8$  K/Ga [6].

En el grupo de Julio M. Fernández (Universidad de Columbia, Nueva York) se estudió la actividad enzimática de estas tiorredoxinas ancestrales por Microscopía de Fuerza Atómica (AFM) de molécula individual. Este trabajo reveló que los patrones de catálisis correlacionan con los dominios de la vida y que dicho mecanismo se estableció hace aproximadamente 4000 millones de años. En cuanto a la eficiencia catalizadora, las tiorredoxinas más antiguas presentan una mayor actividad en condiciones ácidas que sus análogas actuales [6, 12].

Muy recientemente, en colaboración con José A. Gavira perteneciente al Laboratorio de Estudios Cristalográficos de Granada (CSIC), se han resuelto las estructuras tridimensionales de las siete tiorredoxinas precámbricas por cristalografía de rayos X. Los estudios estructurales muestran que todas las tiorredoxinas resucitadas tienen el mismo *fold* que sus análogas actuales, véase fig 3.2.4.



*fig 3.2.4. A) Estructura tridimensional de la tiorredoxina de E. coli (2trx). B) Estructura tridimensional de la tiorredoxina de LPBCA (2yj7). Ambas estructuras han sido resueltas mediante cristalografía de rayos X.*

Las propiedades que presentan estas tiorredoxinas ancestrales, su hiper-estabilidad y su capacidad para realizar la catálisis enzimática a pHs ácidos, resultan muy interesantes en el campo de ingeniería de proteínas. Entender los mecanismos energéticos, moleculares y evolutivos responsables de dichas propiedades, mejoradas respecto de las tiorredoxinas actuales, pueden revelar información muy valiosa para el diseño de proteínas.

- 1) Por ello, nos planteamos realizar un estudio exhaustivo, mutacional y de los factores energéticos de la estabilidad termodinámica y cinética, sobre la tiorredoxina ancestral más estable, la tiorredoxina perteneciente al último ancestro común de las cianobacterias y del grupo de deinococcus y thermus (LPBCA de sus siglas en inglés). Reconstruimos su curva de estabilidad y estudiamos la cinética de su desplegamiento inducido por guanidina. También realizamos un estudio mutacional entre la tiorredoxina de LPBCA y la tiorredoxina perteneciente al último ancestro común de bacterias (LBCA de sus siglas en inglés). Los resultados de este trabajo se muestran en la sección 3.2.1.

- 2) En el transcurso de los estudios sobre la tiorredoxina de LPBCA, los resultados experimentales parecían indicar que la desnaturalización fría ocurría de manera no cooperativa. Esto nos sugirió la hipótesis de que la cooperatividad pudiera haber sido adquirida a lo largo de la evolución. Con objeto evaluar dicha hipótesis nos planteamos el estudio comparativo de los patrones energéticos de la desnaturalización de la tiorredoxina de LPBCA y de su análoga actual, la tiorredoxina de *E. coli*. Ambas proteínas mostraron el mismo patrón en la desnaturalización fría, con lo cual, se descartó la hipótesis inicialmente planteada. Sin embargo, pudimos explicar el patrón de la desnaturalización fría de la tiorredoxina de *E. coli* en base a la lentitud de la cinética de desplegamiento a baja temperatura, siendo este estudio de gran utilidad para la comprensión de la desnaturalización fría de proteínas. Los resultados se muestran en la sección 3.2.2.

La estabilidad de las proteínas es un factor limitante para muchos objetivos en ingeniería de proteínas. Por ejemplo, para el diseño de nuevos sitios activos, con frecuencia, se hacen necesarias mutaciones que conllevan una pérdida significativa de la estabilidad. Esto hace que, previamente, haya que estabilizar la proteína para obtener un armazón molecular lo suficientemente estable capaz de alojar las mutaciones necesarias para conformar el nuevo sitio activo. En este sentido, las proteínas ancestrales, por su hiper-estabilidad, pueden resultar muy apropiadas como punto de partida para proyectos de ingeniería de proteínas, ya que permiten acumular mutaciones desestabilizantes, sin que se comprometa su viabilidad.

- 3) En el último trabajo que se presenta en esta Tesis Doctoral nos planteamos explorar la posibilidad de utilizar una enzima ancestral resucitada como punto de partida en un objetivo de ingeniería de proteínas. En concreto, investigamos la posibilidad de rediseñar el sitio activo de la tiorredoxina de LPBCA. El estudio mutacional realizado entre la tiorredoxina de LPBCA

y de LBCA (descrito en la sección 3.2.1.1.1), reveló tres mutaciones estabilizantes que realizamos sobre la tiorredoxina de LPBCA para obtener una variante más estable aun. Sobre dicha variante eliminamos un grupo ionizable enterrado, esencial para la catálisis enzimática, e intentamos reproducir su efecto en la actividad con un grupo expuesto al solvente. Para ello, seleccionamos con criterios estructurales varios residuos de la superficie proteica y construimos una biblioteca combinatorial con todas las posibles combinaciones de los residuos seleccionados. Los resultados de este trabajo se muestran en la sección 3.2.3.



### **3.2.1. Estudio mutacional dirigido a analizar los factores energéticos que determinan la estabilidad termodinámica y cinética de la tiorredoxina de LPBCA**

A lo largo de la evolución, las proteínas deben adaptarse a las condiciones que las rodean y por ello sufren cambios continuamente. A pesar de que estos cambios en la modulación de las proteínas deben involucrar varias propiedades, recientes trabajos han destacado el papel de la estabilidad de las proteínas en la evolución molecular [13-16]. En este sentido, un trabajo colaborativo entre los grupos de investigación de Eric A. Gaucher (Georgia Institute of Technology, Atlanta), de Julio M. Fernández (Universidad de Columbia, Nueva York) y de José M. Sánchez Ruiz (Universidad de Granada), en el que se han resucitado y caracterizado siete tiorredoxinas precámbricas, ha mostrado resultados muy reveladores respecto a la estabilidad de la tiorredoxina a lo largo de su evolución:

- Por una parte, cabe destacar que las siete tiorredoxinas ancestrales resucitadas son más estables que sus análogas actuales, siendo incluso algunas de ellas, de las proteínas más estables conocidas hasta la fecha (fig 3.2.5).
- Por otra parte, si se representa la temperatura de desnaturalización ( $T_m$ ) de las tiorredoxinas resucitadas *versus* el tiempo geológico en el que se estima que operaban, se observa una disminución de  $T_m$  que correlaciona con la velocidad de enfriamiento de los antiguos océanos, estimada a partir de la deposición de isótopos de oxígeno en rocas [6]. Sin embargo, los valores de  $T_m$  obtenidos son aproximadamente 50°C más altos que la temperatura inferida para los océanos [8] (fig 3.2.5).

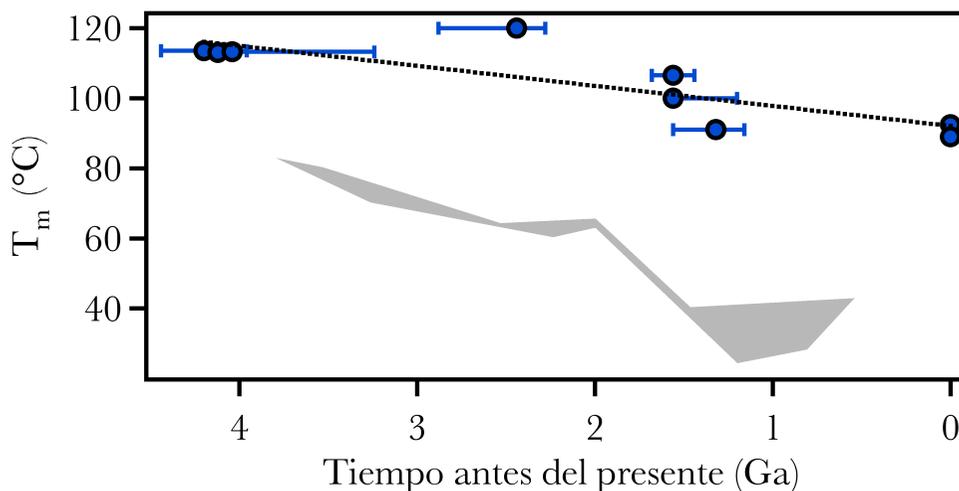


fig 3.2.5. Temperatura de desnaturalización ( $T_m$ ) versus tiempo geológico para las tiorredoxinas precámbricas y para las tiorredoxinas actuales, humana y de *E. coli*, en círculos azules. La línea discontinua representa el ajuste lineal de los datos mostrando una tendencia de la paleotemperatura con un decrecimiento en  $T_m$  de  $5.8 \pm 1.8$  K/Ga [6]. El área gris representa la temperatura de los antiguos océanos inferida a partir de estudios de isótopos de oxígeno en rocas [8].

Para investigar los mecanismos moleculares y los factores energéticos que afectan a la hiper-estabilidad de la tiorredoxina, nos planteamos un análisis comparativo entre la tiorredoxina ancestral de LPBCA (la más estable de las siete tiorredoxinas resucitadas) y la tiorredoxina de *E. coli*, su análoga actual. En fig 3.2.6 se puede observar la diferencia de estabilidad entre ambas proteínas, medida como la diferencia entre sus temperaturas de desnaturalización. En este trabajo, hemos reconstruido la curva de estabilidad de la tiorredoxina de LPBCA, que hasta lo que sabemos, es la primera curva de estabilidad reconstruida para una proteína ancestral y hemos estudiado las cinéticas del desplegamiento inducido por guanidina a diferentes temperaturas. Ambos estudios han ayudado a la comprensión de los factores energéticos responsables de la estabilidad tan inusual que presenta la tiorredoxina de LPBCA, proporcionando información útil para el diseño de proteínas estabilizadas.

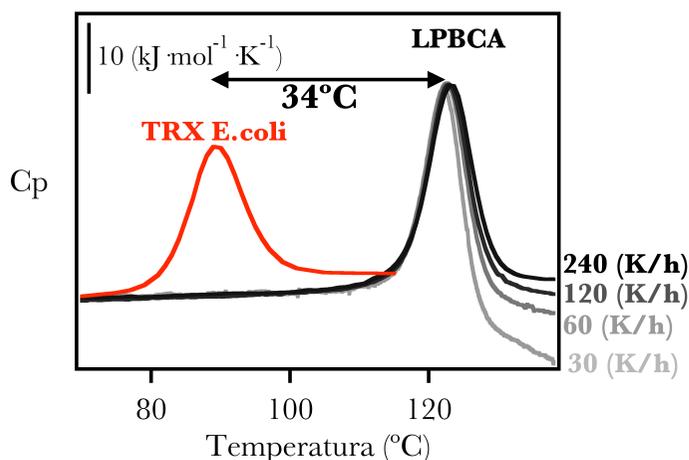


fig 3.2.6. Perfil calorimétrico de la tiorredoxina de *E. coli* y de LPBCA. Se puede observar la diferencia de estabilidad medida como el incremento entre sus temperaturas de desnaturalización ( $T_m$ ). Los experimentos de DSC se llevaron a cabo una concentración de proteína de 0.5 mg/mL, en 50 mM HEPES a pH 7. Para la tiorredoxina de *E. coli* la velocidad de barrido fue de 200 K/h y para la tiorredoxina de LPBCA, se muestran varias trazas realizadas a diferentes velocidades de barrido (30 K/h, 60 K/h, 120 K/h y 240 K/h).

La diferencia de aproximadamente 50°C entre las temperaturas de desnaturalización de las tiorredoxinas y la temperatura del entorno en el que operaban dichas proteínas, parece sugerir que la selección no haya actuado directamente sobre la estabilidad termodinámica, pero sí sobre otro factor relacionado con la estabilidad termodinámica [17]. Estudios teóricos y un gran número de estudios experimentales sobre desnaturalización térmica irreversible de proteínas sugieren que la estabilidad termodinámica (valor positivo para la energía libre de desnaturalización) por sí sola no garantiza que la proteína permanezca en el estado nativo durante una cierta escala de tiempo, ya que las alteraciones irreversibles (incluso si ocurren en estados desnaturalizados poco poblados o estados parcialmente desplegados) pueden perjudicar el estado nativo [18]. De acuerdo con esto, y teniendo en cuenta la diferencia de 50°C entre las  $T_m$  de las tiorredoxinas estudiadas y la temperatura del ambiente en el que operaban, proponemos que la tiorredoxina ha sido seleccionada de forma natural para tener

una estabilidad cinética significativa y esta se refleja en la estabilidad termodinámica [17]. Para comprobar la validez de esta hipótesis nos planteamos el estudio comparativo de las cinéticas del desplegamiento inducido por guanidina a diferentes temperaturas para la tiorredoxina de LPBCA y para la tiorredoxina de *E. coli*. Este estudio ha proporcionado información relevante que ayuda a la comprender la relación entre la presión evolutiva a la que ha sido sometida la tiorredoxina y su estabilidad.

### **3.2.1.1. Estudio de la estabilidad termodinámica de la tiorredoxina de LPBCA por calorimetría diferencial de barrido**

La estabilidad termodinámica de una proteína se define con frecuencia como la diferencia entre la energía de Gibbs del estado nativo y desnaturalizado, la dependencia de  $\Delta G_N^D$  con la temperatura se conoce como curva de estabilidad (sección 2.6.4.1.2). Así, para caracterizar la estabilidad termodinámica de la tiorredoxina de LPBCA nos propusimos reconstruir su curva de estabilidad. Para ello, se necesita un valor fiable de  $\Delta_N^D C_p$ , que se puede calcular directamente de la traza de DSC o bien a partir de la pendiente resultante al representar  $\Delta_N^D H(T_m)$  versus  $T_m$  para varias transiciones de DSC con diferentes valores de  $T_m$ .

En fig 3.2.7 se muestran los perfiles calorimétricos de la tiorredoxina de LPBCA realizados con una disolución de proteína de 0.5 mg/mL, en 50 mM HEPES a pH 7 a diferentes velocidades de barrido (30K/h, 60K/h, 120K/h y 240K/h). Como se puede observar, las trazas están distorsionadas cinéticamente, de manera que a menor velocidad de barrido, aumenta su asimetría y disminuye la reversibilidad del proceso de desplegamiento (obtenido como la proporción de señal calorimétrica que se recupera en el segundo barrido con respecto a la señal calorimétrica del primer barrido). La distorsión cinética presente en las trazas de DSC (fig 3.2.7) impedía obtener el valor de la capacidad calorífica del estado desnaturalizado ( $C_p^D$ ), y por tanto no es fue posible calcular el valor de  $\Delta_N^D C_p$  del

análisis directo de las trazas. Por ello, decidimos obtener diferentes variantes sobre la tiorredoxina de LPBCA con diferentes estabilidades térmicas. La pendiente del ajuste lineal de la representación de  $\Delta_N^D H(T_m)$  versus  $T_m$  para las variantes obtenidas, nos dará el valor de  $\Delta_N^D C_p$  de la tiorredoxina de LPBCA.

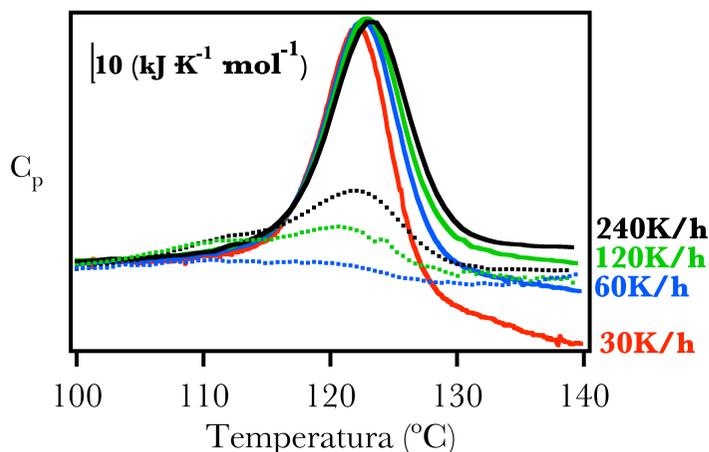


fig 3.2.7. Perfil de capacidad calorífica frente a temperatura de la tiorredoxina de LPBCA. La proteína se dializó frente a 50 mM HEPES a pH 7 y los experimentos se realizaron a una concentración de proteína de 0.5 mg/mL a diferentes velocidades de barrido. En línea continua se muestra el primer barrido al que se somete la muestra y en línea discontinua el segundo.

### 3.2.1.1.1. Estudio mutacional entre la tiorredoxina de LBCA y de LPBCA

Con objeto de obtener variantes de la tiorredoxina de LPBCA con diferentes estabilidades, que nos permitan inferir el  $\Delta_N^D C_p$ , realizamos un análisis mutacional entre la tiorredoxina de LPBCA y su antecesora, la tiorredoxina de LBCA. Ambas proteínas se diferencian en 13 mutaciones y su estabilidad difiere en 9°C (fig 3.2.8). Para reducir el tiempo de producción y purificación de las proteínas, las variantes propuestas por el estudio mutacional se purificaron con una cola de 6 histidinas en la zona del amino terminal. Así se obtuvieron, mediante mutagénesis dirigida, 13 variantes de la tiorredoxina de LPBCA con las

mutaciones simples V7I, T8N, Q14E, K47E, V55F, V56A, N65E, Q69K, R73M, Q85E, R89K, Q94R y D104E.

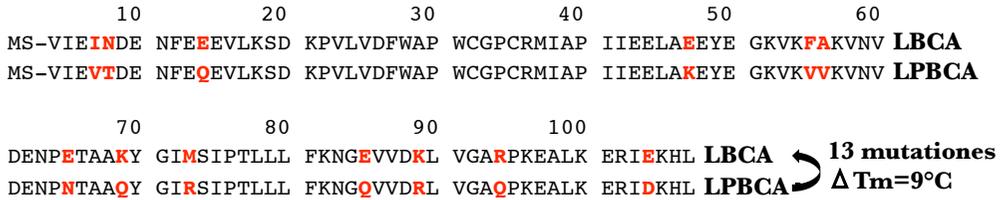


fig 3.2.8. Alineamiento de las secuencias de las tioredoxinas de LPBCA y de LBCA. La numeración utilizada es aquella que hace coincidir las posiciones equivalentes de las tioredoxinas de LPBCA, de LBCA y la de *E. coli*.

Los experimentos de calorimetría de las formas mutantes simples se realizaron de 20°C a 140°C a una velocidad de barrido de 240 K/h. Las muestras de proteína se dializaron frente a 50 mM HEPES a pH 7 a una concentración de 0.5 mg/mL. Las mutaciones V55F, Q69K y D104E resultaron estabilizantes, mientras que V7I, T8N, Q14E, K47E, V56A, N65E, R73M, Q85E, R89K y Q94R resultaron desestabilizantes (fig 3.2.9).

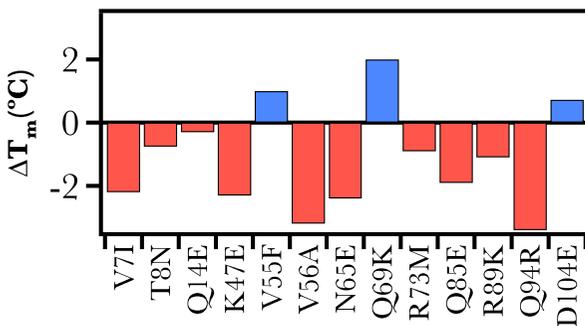


fig 3.2.9. Diferencia de estabilidad entre las variantes de la tioredoxina de LPBCA y la tioredoxina de LPBCA silvestre. Se ha tomado como  $T_m$  la temperatura a la cual, la señal de capacidad calorífica es máxima. En azul se muestran las variantes que resultaron más estables y en rojo las menos estables.

Para obtener proteínas cuyas  $T_m$  cubran un amplio rango de temperaturas, se seleccionó la variante menos estable (Q94R) y se construyeron las variantes doble (V55F/Q69K) y triple (V55F/Q69K/D104E) con las mutaciones estabilizadoras.

Esta nueva tanda de mutantes se clonaron en el vector pET30a (+) (sección 2.2 de metodología) para obtener las proteínas sin cola de histidina y así, estudiarlas sin el efecto de la cola de 6 histidinas en el desplegamiento (desestabiliza la proteína).

### **3.2.1.1.2. Análisis de las trazas de DSC**

En fig 3.2.10 se muestran los perfiles calorimétricos de las variantes Q94R/LPBCA, V55F/Q69K/LPBCA (doble), V55F/Q69K/D104E/LPBCA (triple) y la tiorredoxina de LPBCA silvestre realizados a diferente velocidad de barrido (30K/h, 60K/h, 120K/h y 240K/h). Los experimentos se llevaron a cabo desde 20°C hasta 140°C, a una concentración de proteína de 0.5 mg/mL en 50 mM HEPES a pH 7. No se muestran las trazas correspondientes al segundo barrido al que se ha sometido la muestra para facilitar la visualización de la figura, sin embargo, en todos los casos se cumple que a 30 K/h no se recupera la señal calorimétrica de la traza, a 60 K/h se recupera el 5%, a 120 K/h el 10% y a 240 K/h el 20%.

Como se puede observar en fig 3.2.10, las trazas de DSC muestran efecto de velocidad de barrido, acentuándose dicho efecto, a velocidades de barrido más lentas. Así, a medida que aumenta la velocidad de barrido, la transición es más simétrica, más reversible y se desplaza a mayores temperaturas. Esta dependencia con la velocidad de barrido se atribuye a procesos (como autólisis, agregación o alteraciones químicas de los residuos) que ocurren a alta temperatura, haciendo que los estados significativamente poblados durante el desplegamiento se alejen de sus valores de equilibrio. A bajas temperaturas, sin embargo, los estados significativamente poblados durante el desplegamiento si se encuentran en sus valores de equilibrio. El hecho de que el efecto de velocidad de barrido aumente a menores velocidades, se debe a que estas alteraciones irreversibles ocurren lentamente, de manera que se aprecian más, cuanto mayor sea el periodo de tiempo en el que transcurre la transición.

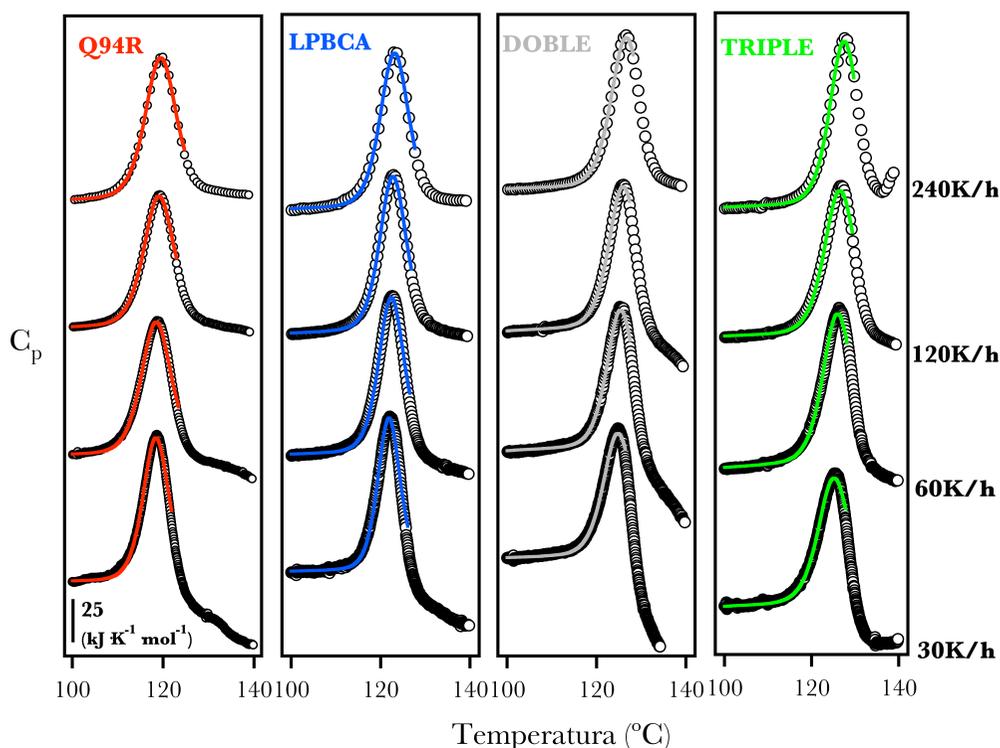


fig 3.2.10. Perfiles de capacidad calorífica frente a temperatura de las variantes Q94R, doble y triple y de la tiorredoxina de LPBCA. Las trazas se encuentran desplazadas a lo largo del eje y para facilitar su visualización. Las disoluciones se proteína dializaron frente a 50 mM HEPES a pH7, los experimentos se han llevado a cabo diferentes velocidades de barrido (30K/h, 60K/h, 120K/h y 240K/h) a una concentración de proteína de 0.5 mg/mL. En círculos se muestran los datos experimentales y en línea continua el ajuste al modelo de equilibrio de dos estados.

En este trabajo, para el análisis de las trazas, se descartaron los datos experimentales de capacidad calorífica obtenidos a alta temperatura, donde ocurren los procesos irreversibles que hacen que los estados significativamente poblados se alejen de sus valores de equilibrio. Los datos experimentales obtenidos a bajas temperaturas, donde los estados significativamente poblados sí se encontraban en sus valores de equilibrio, se analizaron según el modelo de equilibrio de dos estados (sección 2.6.4.1.1 de metodología). Para elegir el rango de temperaturas en el que ajustar dicho modelo, se eliminaron los datos

experimentales a partir de la temperatura más alta, tras la cual, el modelo proporciona un ajuste que se aleja de los datos experimentales. Como se puede observar en fig 3.2.10, para el mutante Q94R y la tiorredoxina de LPBCA, se descartaron para el ajuste los datos experimentales obtenidos para una temperatura cuatro grados mayor que la temperatura a la cual la capacidad calorífica es máxima y para las variantes doble y triple, se descartaron los datos experimentales obtenidos para una temperatura dos grados mayor que la temperatura a la cual la capacidad calorífica es máxima. Los parámetros termodinámicos obtenidos del ajuste se muestran en la tabla 3.2.1.

<b>Proteína</b>	<b>T<sub>m</sub> (°C)</b>	<b><math>\Delta H^{vH}</math> (kJ mol<sup>-1</sup>)</b>	<b><math>\Delta_N^D H_m</math> (kJ mol<sup>-1</sup>)</b>	<b>v (K/h)</b>
<b>Q94R</b>	118.45	681.8	489.0	30
<b>Q94R</b>	118.52	591.0	519.0	60
<b>Q94R</b>	119.12	580.9	520.2	120
<b>Q94R</b>	119.51	568.6	568.6	240
<b>LPBCA</b>	121.89	785.2	462.1	30
<b>LPBCA</b>	122.43	693.8	531.8	60
<b>LPBCA</b>	122.77	670.9	548.3	120
<b>LPBCA</b>	123.14	625.3	586.2	240
<b>doble</b>	124.82	606.7	485.4	30
<b>doble</b>	125.97	628.3	541.9	60
<b>doble</b>	125.41	650.4	508.0	120
<b>doble</b>	126.67	595.0	611.4	240
<b>triple</b>	125.21	552.3	529.9	30
<b>triple</b>	126.01	594.8	583.9	60
<b>triple</b>	126.53	568.4	592.2	120
<b>triple</b>	127.49	612.3	628.9	240

tabla 3.2.1. Parámetros termodinámicos obtenidos del ajuste al modelo de equilibrio de dos estados en el rango de temperaturas en el que se cumple la condición de equilibrio.

Una buena concordancia con el modelo reversible de dos estados viene dado por un valor de  $\Delta_N^D H_m / \Delta H^{vH}$  cercano a la unidad [19]. En el caso que nos atañe, se calculó dicho cociente para cada traza calorimetría utilizando los datos derivados del ajuste, y se representó  $\Delta_N^D H_m / \Delta H^{vH}$  en función de la inversa de la velocidad de barrido. El cociente extrapolado a velocidad de barrido infinita, utilizando un polinomio de segundo grado, resultó dentro del rango establecido por Privalov y colaboradores [19] como buen indicativo de la concordancia con el modelo de dos estados de equilibrio (fig 3.2.11)

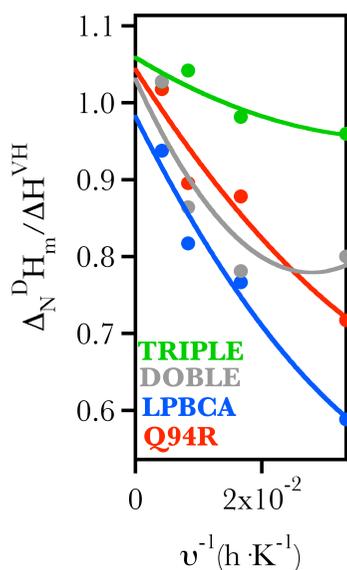


fig 3.2.11. Representación del cociente  $\Delta_N^D H_m / \Delta H^{vH}$  frente a la inversa de la velocidad de barrido. En círculos se muestran los datos experimentales y en línea continua el ajuste a un polinomio de segundo grado. En todos los casos, a medida que aumenta la velocidad de barrido, el valor del cociente  $\Delta_N^D H_m / \Delta H^{vH}$  tiende a la unidad, resultando su extrapolación a velocidad de barrido infinita (teniendo en cuenta el error experimental asociado al valor), dentro del rango establecido por Privalov y colaboradores como buen indicativo de la concordancia con el modelo de equilibrio de dos estados.

Los datos derivados del ajuste, nos proporcionan valores de  $T_m$  y de  $\Delta_N^D H_m$  en función de la velocidad de barrido, teniendo en cuenta la tendencia a la unidad del cociente  $\Delta_N^D H_m / \Delta H^{vH}$  a velocidad de barrido infinita, podemos considerar los datos de equilibrio como los valores que obtenemos de su extrapolación a velocidad de barrido infinita. Así, de la representación de  $T_m$  frente a la inversa de la velocidad de barrido, se obtuvo  $T_m^0$  como la temperatura de desnaturalización extrapolada mediante el ajuste a un polinomio de segundo grado a velocidad de barrido infinita (fig 3.2.12 A). Para obtener el valor de la entalpía calorimétrica de

equilibrio, se representó  $\Delta_N^D H_m$  frente a la  $T_m$  y se realizó el ajuste lineal de los datos. La proyección de  $T_m^0$  sobre dicho ajuste lineal nos proporciona el valor de entalpía calorimétrica a velocidad de barrido infinita ( $\Delta_N^D H_m^0$ ).

Finalmente, el  $\Delta_N^D C_p$  lo obtenemos del ajuste lineal de  $\Delta_N^D H_m^0$  a diferentes temperaturas, obteniéndose un valor de  $6.36 \text{ kJ} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$  (fig 3.2.12 B).

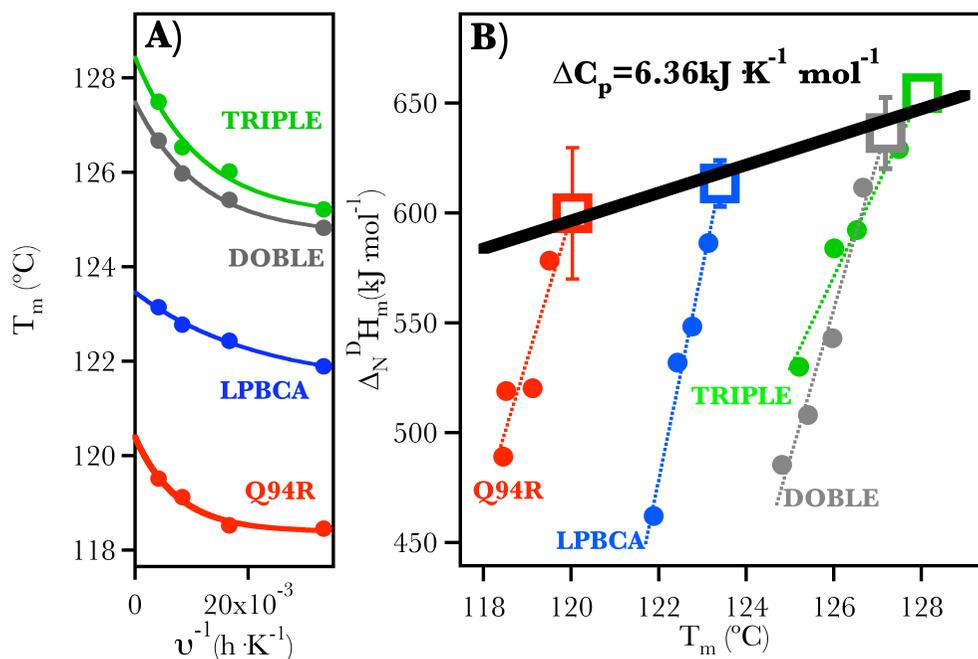


fig 3.2.12. A) Representación de la temperatura de desnaturalización frente a la inversa de la velocidad de barrido. Los datos experimentales se muestran en círculos cerrados y en línea continua se muestra el ajuste a un polinomio de segundo grado. De la extrapolación a velocidad de barrido infinita se obtiene  $T_m^0$  para cada una de las variantes estudiadas. B) Representación de la entalpía calorimétrica frente a la temperatura de desnaturalización. En círculos cerrados se muestran los datos experimentales obtenidos del ajuste al modelo de equilibrio de dos estados, en línea discontinua se muestra el ajuste lineal de la entalpía versus  $T_m$  para cada una de las variantes estudiadas. La proyección de  $T_m^0$  sobre dicho ajuste lineal nos proporciona el valor de entalpía calorimétrica a velocidad de barrido infinita ( $\Delta_N^D H_m^0$ ), que se muestra en la gráfica como un cuadrado abierto. El ajuste lineal de  $\Delta_N^D H_m^0$  a diferentes temperaturas nos proporciona un valor de  $\Delta_N^D C_p$  de  $6.36 \text{ kJ} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$ .

El valor obtenido de  $\Delta_N^D C_p = 6.36 \text{ kJ} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$ , es similar al de otras variantes de tiorredoxina estudiadas en nuestro grupo de investigación (fig 3.2.13) en contra de lo esperado, según la predicción que indica que  $\Delta_N^D C_p$  disminuye a altas temperaturas, anulándose a temperaturas superiores a  $100^\circ\text{C}$  (sección 2.6.4.1.2. de metodología). Se realizó la media aritmética de los valores de  $\Delta_N^D C_p$  para todas las variantes recopiladas obteniéndose un valor de  $\Delta_N^D C_p = 5.76 \pm 0.12 \text{ kJ} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$  y se consideró que su valor era independiente de la temperatura en el rango de estudiado ( $50^\circ\text{C}$ - $130^\circ\text{C}$ ).

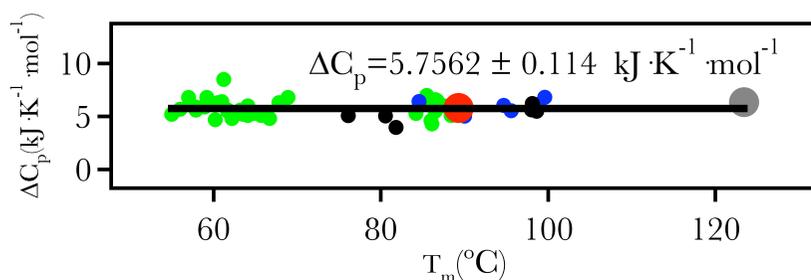


fig 3.2.13. A) Datos de  $\Delta_N^D C_p$  obtenidos del ajuste directo de perfiles de capacidades caloríficas absolutas al modelo de equilibrio de dos estados de variantes de tiorredoxina. En verde se muestran los datos de tiorredoxina silvestre de *E. coli* y fragmentos de tiorredoxina (1-73, 74-108) a diferentes pHs [20], en negro los datos de las variantes estabilizadas de tiorredoxina 3x y 6x a diferentes pHs [21], en azul se muestran los datos experimentales a pH 7 obtenidos con variantes de tiorredoxina con mutaciones disruptivas del núcleo hidrofóbico estudiada en nuestro grupo de investigación, en rojo se muestra el valor de  $\Delta_N^D C_p$  para la tiorredoxina de *E. coli* a pH 7 y en gris el dato obtenido para la tiorredoxina de LPBCA por el procedimiento anteriormente descrito.

### 3.2.1.2. Reconstrucción de la curva de estabilidad de la tiorredoxina ancestral de LPBCA y su estudio comparativo con la curva de estabilidad de la tiorredoxina de *E. coli*

Una vez obtenido el  $\Delta_N^D C_p$ , las curvas de estabilidad de la tiorredoxina de LPBCA y de *E. coli* (fig 3.2.15A) fueron construidas a partir de las ecuaciones 2.15, 2.16 y 2.17 descritas en el capítulo 2.6.4.1.1. de metodología. Dado el estudio

descrito en el apartado anterior, el  $\Delta_N^D C_p$  se consideró el mismo para ambos casos con un valor de  $5.8 \text{ kJ} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$ . Se consideraron los valores de  $\Delta H=613 \text{ kJ} \cdot \text{mol}^{-1}$  y  $T_m=123.4^\circ\text{C}$  para la tiorredoxina de LPBCA obtenidos de la extrapolación a velocidad de barrido infinita (fig 3.2.12). Para la tiorredoxina de *E. coli* se tomaron los valores de  $\Delta H=425 \text{ kJ} \cdot \text{mol}^{-1}$  y  $T_m=89.3^\circ\text{C}$  obtenidos al ajustar el modelo reversible de dos estados a la traza calorimétrica que se muestra en fig 3.2.14 (el experimento de DSC se realizó desde  $20^\circ\text{C}$  hasta  $110^\circ\text{C}$ , a una concentración de tiorredoxina de *E. coli* de  $0.5 \text{ mg/mL}$ , en HEPES  $50 \text{ mM}$  pH 7 a una velocidad de barrido de  $200 \text{ K/h}$ ).

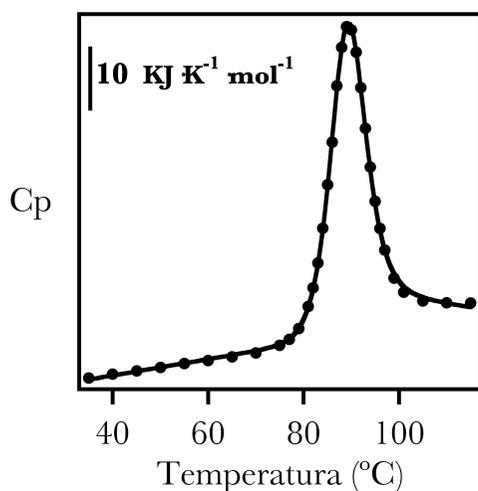


fig 3.2.14. Perfil calorimétrico de la tiorredoxina de *E. coli*. El experimento de DSC se realizó desde  $20^\circ\text{C}$  hasta  $110^\circ\text{C}$ , a una concentración de tiorredoxina de *E. coli* de  $0.5 \text{ mg/mL}$ , en HEPES  $50 \text{ mM}$  pH 7 a una velocidad de barrido de  $200 \text{ K/h}$ . En círculos se muestran los datos experimentales y en línea continua el ajuste al modelo de equilibrio de los estados.

La diferencia de estabilidad entre ambas proteínas calculadas como  $\Delta\Delta G=\Delta G(\text{TrxLPBCA})-\Delta G(\text{Trx.}E. coli)$  a  $60^\circ\text{C}$  es de  $40 \text{ kJ} \cdot \text{mol}^{-1}$ (fig 3.2.15 B). El análisis exhaustivo de las curvas de estabilidad muestra que la contribución mayoritaria a la estabilización es de carácter entrópico y su valor, calculado como  $T \cdot \Delta\Delta S=T \cdot \Delta S(\text{TrxLPBCA})-T \cdot \Delta S(\text{Trx.}E. coli)$  a  $60^\circ\text{C}$ , es  $47 \text{ kJ} \cdot \text{mol}^{-1}$ (fig 3.2.15 B).

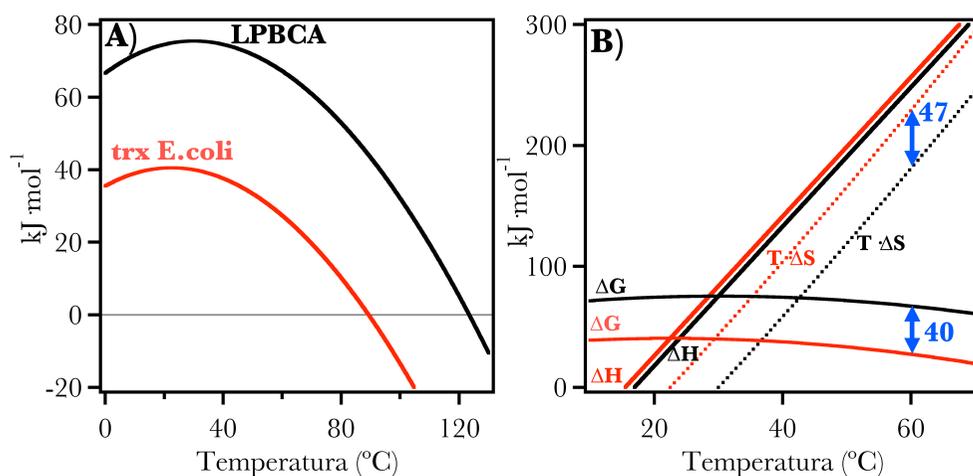


fig 3.2.15. A) Curva de estabilidad para la tiorredoxina de *E. coli* (línea continua roja) y para la de LPBCA (línea continua negra). B) Representación de la contribución entálpica a la estabilidad de la tiorredoxina de *E. coli* con trazo continuo rojo y de la contribución entrópica con trazo discontinuo rojo. Para la tiorredoxina de LPBCA la contribución entálpica se representa con trazo continuo negro y la entrópica con trazo discontinuo del mismo color. Como se puede observar, la diferencia de estabilidad entre ambas proteínas, medida como la diferencia de variación de energía libre ( $\Delta\Delta G$ ) a  $60^{\circ}\text{C}$ , es de  $40 \text{ kJ mol}^{-1}$ . La contribución entrópica a la estabilidad de LPBCA, medida como  $T \cdot \Delta\Delta S$  a  $60^{\circ}\text{C}$ , es de  $47 \text{ kJ mol}^{-1}$ , siendo la contribución mayoritaria a la estabilidad.

Para justificar la estabilización entrópica de la tiorredoxina de LPBCA, se realizó un recuento de las prolina y glicinas de su secuencia, ya que, es bien sabido que las mutaciones que implican eliminar una glicina o insertar una prolina contribuyen entrópicamente a la estabilización de la proteína [22]. Atendiendo a las secuencias proteicas, se ha observado una relación de  $\text{Pro}/_{\text{gly}} = 5/9$  para la tiorredoxina de *E. coli* y de  $\text{Pro}/_{\text{gly}} = 7/5$  para la tiorredoxina de LPBCA. Se ha estimado que el coste entrópico de eliminar una glicina es de  $\Delta S = -10 \text{ J} \cdot \text{K}^{-1} \text{ mol}^{-1}$  y de  $\Delta S = -17 \text{ J} \cdot \text{K}^{-1} \text{ mol}^{-1}$  el de introducir una prolina [23]. Considerando la eliminación de 4 glicinas y la inserción de 2 prolina, la contribución entrópica total a la estabilidad, calculada como  $T \cdot \Delta\Delta S$  a  $60^{\circ}\text{C}$ , sería

de  $-25 \text{ kJ mol}^{-1}$ . En nuestro grupo de investigación se ha comprobado experimentalmente como parte de la tesis doctoral de Álvaro Inglés Prieto que las mutaciones de glicina a prolina sobre la tiorredoxina de *E. coli* explican  $\sim 20^\circ\text{C}$  de incremento en la temperatura de desnaturalización. Además, este estudio también señaló la mutación K90L como responsable de  $\sim 5^\circ\text{C}$  de aumento en la temperatura de desnaturalización. Para determinar si la mutación K90L contribuye entrópicamente a la estabilización, se calculó, a partir de los valores de ASA (área superficial accesible) [23], el  $\Delta S$  de la Lys90 de la tiorredoxina de *E. coli* y de la Leu90 de la tiorredoxina de LPBCA. Para ello, se realizó el cálculo de área superficial accesible con una modificación del algoritmo de Shrake-Rupley [24]. Este algoritmo sitúa aleatoriamente 2000 puntos alrededor de la esfera de van der Waals que representa cada átomo. Se utilizó un radio de  $1.4 \text{ \AA}$  para la sonda del disolvente y los radios de Chothia [25] para los átomos de la proteína. Los cálculos se realizaron con el programa ASA-ENG.exe, desarrollado en nuestro grupo de investigación, y revelaron como la contribución entrópica a la estabilidad de la mutación K90L a  $60^\circ\text{C}$ , calculada como  $T \cdot \Delta\Delta S = T \cdot \Delta S(\text{Leu90}) - T \cdot \Delta S(\text{Lys90})$ , es de  $-40 \text{ kJ mol}^{-1}$ . Hay que tener en cuenta que para el cálculo de la contribución entrópica de K90L, sólo se han considerado los residuos implicados directamente en la mutación, sin embargo, los residuos de la periferia de la Leu90 y de la Lys90 también pueden influir en la contribución, en cualquier caso, el resultado obtenido nos proporciona información cualitativa que ayuda a comprender el papel, en la energética de la estabilidad de la tiorredoxina de LPBCA, de la mutación K90L.

Así, a partir del estudio de la curva de estabilidad de la tiorredoxina de LPBCA y tras su comparativa con la curva de estabilidad de la tiorredoxina de *E. coli*, hemos determinado que la estabilización termodinámica de la tiorredoxina de LPBCA es principalmente de carácter entrópico, y además, hemos podido justificar dicha

estabilización entrópica (al menos en parte) en base al aumento del cociente Pro/gly y a la mutación K90L.

### 3.2.1.3. Estudio experimental de las velocidades de plegamiento de la tiorredoxina de LPBCA y *E. coli* seguida por fluorescencia

Estudios previos realizados en nuestro laboratorio sugieren que, en el curso de la evolución de la tiorredoxina, la selección natural no rechaza las mutaciones que causan un descenso en la estabilidad termodinámica de algunos kJ/mol, pero que sí rechaza aquellas que causan un descenso en la estabilidad cinética de algunos kJ/mol [17]. Este argumento se basa en el efecto que tendría un descenso de unos pocos kJ/mol en la  $T_m$  (si la disminución se produce en la estabilidad termodinámica) o en el tiempo de vida media de la proteína (si la disminución se produce en la estabilidad cinética).

En términos de temperatura de desnaturalización, con la ecuación Gibbs-Helmholtz (ec.3.2.1) se puede ver como la  $T_m$  cambia casi linealmente con  $\Delta\Delta G$  y un descenso de unos pocos kJ/mol en  $\Delta G$ , causa un descenso moderado en el valor de la  $T_m$ . Como la  $T_m$  de la tiorredoxina de *E. coli* es de 89.3°C, un descenso de unos pocos kJ/mol en  $\Delta G$ , deja la  $T_m$  bastantes grados por encima de la temperatura fisiológica y por tanto seguirán, existiendo valores de  $\Delta G > 0$  a la temperatura fisiológica (fig 3.2.16 A).

$$\Delta G = \Delta H_m \cdot \left(1 - \frac{T}{T_m}\right) + \Delta C_p \cdot \left[T - T_m - T \ln\left(\frac{T}{T_m}\right)\right] \quad \text{ec.3.2.1}$$

Considerando la estabilidad cinética, determinada por la velocidad de desplegamiento y suponiendo que la teoría del estado de transición es aplicable, entonces:

$$\tau = \frac{1}{k_{UNF}} \propto \exp\left(\frac{\Delta G_{UNF}^\ddagger}{RT}\right) \quad \text{ec.3.2.2}$$

donde  $k_{\text{UNF}}$  es la constante de velocidad de desplegamiento de primer orden,  $\tau$  es el correspondiente tiempo de vida media y  $\Delta G_{\text{UNF}}^\ddagger$  es la energía libre de activación de desplegamiento. Esta ecuación indica que  $\tau$  decrece con la energía libre de activación de una manera exponencial. Como resultado, un descenso de algunos kJ/mol en  $\Delta G_{\text{UNF}}^\ddagger$  puede producir un gran cambio en  $\tau$ , provocando un descenso de algunas semanas a horas e incluso a minutos (fig 3.2.16 B).

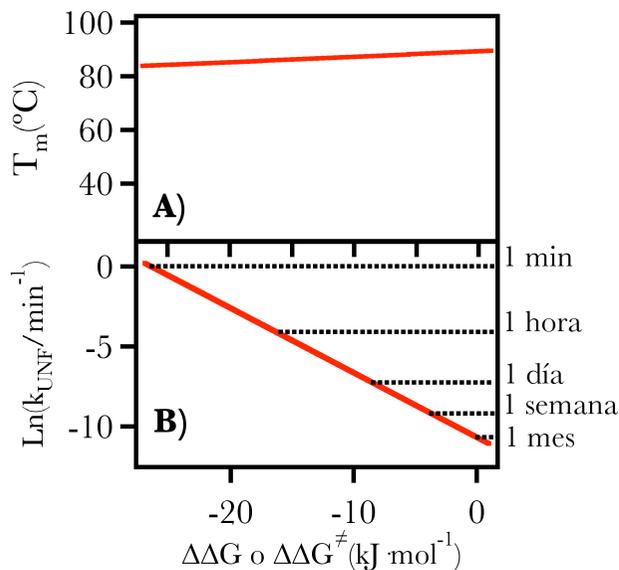


fig 3.2.16. Efecto de un descenso de unos pocos  $\text{kJ/mol}$  en la variación del cambio de energía libre, sobre la  $T_m$  (si la disminución se produce en la estabilidad termodinámica) o, sobre el tiempo de vida media de la proteína (si la disminución se produce en la estabilidad cinética). A) Dependencia de la temperatura de desnaturalización en equilibrio ( $T_m$ ) con la variación del cambio de la energía libre ( $\Delta\Delta G$ ). Los valores de  $T_m$  se calculan resolviendo numéricamente la ecuación de Gibbs-Helmholtz (ec.3.2.1) usando los parámetros termodinámicos para tiorredoxina de *E. coli* ( $T_m=89.3^\circ\text{C}$ ,  $\Delta H=425 \text{ kJ mol}^{-1}$  y  $\Delta_N^D C_P=5.8 \text{ kJ } \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$ ). B) Dependencia de la constante de velocidad de desplegamiento con la variación de la energía libre de activación ( $\Delta\Delta G_{\text{UNF}}^\ddagger$ ). Los valores de la constante de velocidad se calculan usando la ecuación ec.3.2.2. A modo ilustrativo, en este cálculo, se ha considerado la constante de velocidad para la tiorredoxina de *E. coli* ( $\Delta\Delta G_{\text{UNF}}^\ddagger=0$ ) igual a  $3.31 \cdot 10^{-5} \text{ min}^{-1}$ , que corresponde a un tiempo de vida media de 1 mes. El eje de la derecha se indican valores representativos de tiempos de vida media, que se corresponden con las constantes de velocidad correspondientes del eje de la izquierda.

Este trabajo sugiere que la selección natural actúa, en el caso de la tiorredoxina, sobre la estabilidad cinética [17]. Siguiendo la misma línea argumental, la diferencia de estabilidad termodinámica entre la tiorredoxina de LPBCA y la de *E. coli* debe ser un reflejo de la diferencia de estabilidad cinética entre ambas proteínas. El estudio cinético descrito a continuación muestra como de hecho esto es lo que ocurre.

Las cinéticas de desplegamiento inducidas por guanidina se realizaron a diferentes temperaturas para la tiorredoxina de *E. coli* y para la tiorredoxina de LPBCA (sección 2.8.3. del capítulo de metodología). En todos los casos, los valores de  $k_{\text{UNF}}$  se obtuvieron mediante el ajuste monoexponencial de los perfiles de emisión de fluorescencia a 350 nm frente a tiempo (ver figura 2.18 a modo de ejemplo). A cada temperatura, la representación del  $\ln(k_{\text{UNF}})$  frente a la concentración de guanidina mostraban una relación lineal, ya que el rango de concentraciones de guanidina estudiado para cada temperatura, se corresponde con la rama de desplegamiento de una gráfica *chevron* (logaritmo de la constante de velocidad *versus* concentración de guanidina).

Esta relación lineal se puede describir como:

$$\ln(k_{\text{UNF}}) = k_{\text{UNF}}^0 + \frac{m_{\text{U}}}{RT} \cdot c \quad \text{ec.3.2.3}$$

donde  $m_{\text{U}}$  describe el efecto de la concentración de guanidina sobre las energías libre de Gibbs de activación para el proceso de desplegamiento.  $m_{\text{U}}$  está relacionado con el cambio del área superficial accesible expuesto al disolvente desde el estado nativo al estado de transición. Experimentalmente los valores de  $m_{\text{U}}$  obtenidos para la tiorredoxina de LPBCA y para la de *E. coli* no mostraron dependencia con la temperatura en el rango estudiado, por ello se realizó un ajuste global de los datos a la ecuación ec.3.2.3 manteniendo el valor de  $m_{\text{U}}$  constante para cada proteína. Se obtuvo un valor de  $m_{\text{U}} = 8.94 \pm 0.14 \text{ kJ} \cdot \text{mol}^{-1} \cdot \text{M}^{-1}$

para la tiorredoxina de *E. coli* y de  $m_U = 10.52 \pm 0.07 \text{ kJ mol}^{-1} \cdot \text{M}^{-1}$  para la tiorredoxina de LPBCA. Los errores describen la incertidumbre asociada al proceso de ajuste.

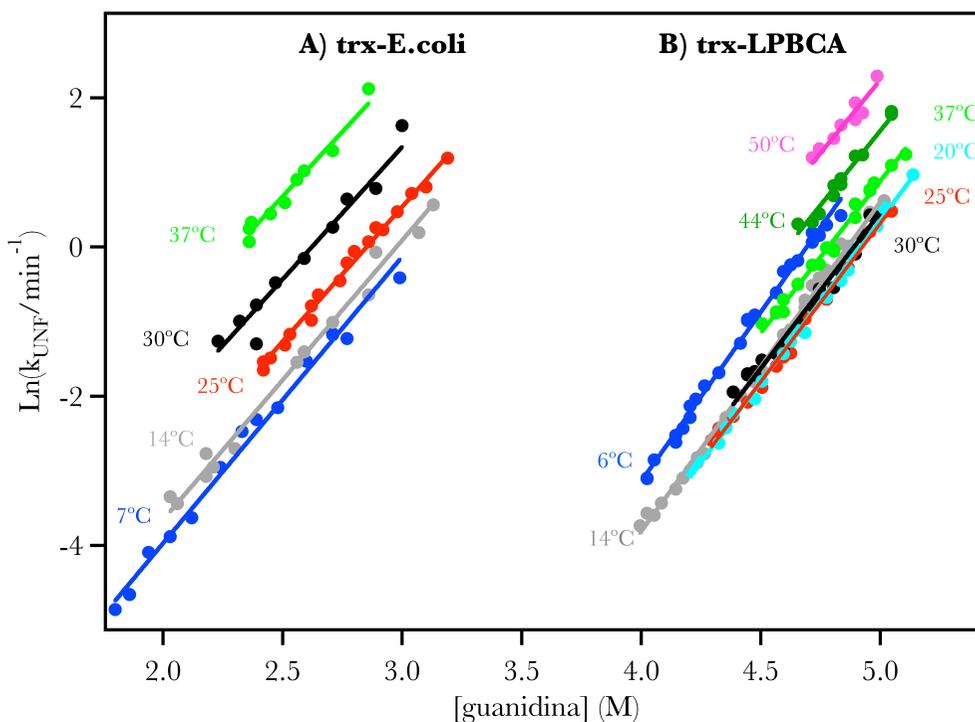


fig 3.2.17. Representación de la constante de desplegamiento  $k_{UNF}$  frente a concentración de guanidina (en un rango de concentración que se corresponde con la rama de desplegamiento de la gráfica chevron) a diferentes temperaturas para la tiorredoxina de *E. coli* (A) y para la de LPBCA (B). En círculos se muestran los datos experimentales y en línea continua el ajuste a la ecuación ec.3.2.3. Se realizó un ajuste global manteniendo el valor de  $m_U$  constante con la temperatura para cada proteína.

Los valores de  $k_{UNF}^0$  extrapolados a concentración nula de guanidina (fig 3.2.18) deben de tomarse con cautela debido a las desviaciones de la linealidad que existen a baja concentración de guanidina, estas desviaciones están vinculadas al apantallamiento de las interacciones carga-carga [26]. Aún así, los valores extrapolados son útiles ya que aportan información cualitativa de la escala de

tiempo en el que se desarrolla el proceso, en ausencia de desnaturalizantes químicos. Además estas desviaciones no afectan a la extrapolación de las  $k_{UNF}$  a concentración de guanidina 3.5 M (fig 3.2.18).

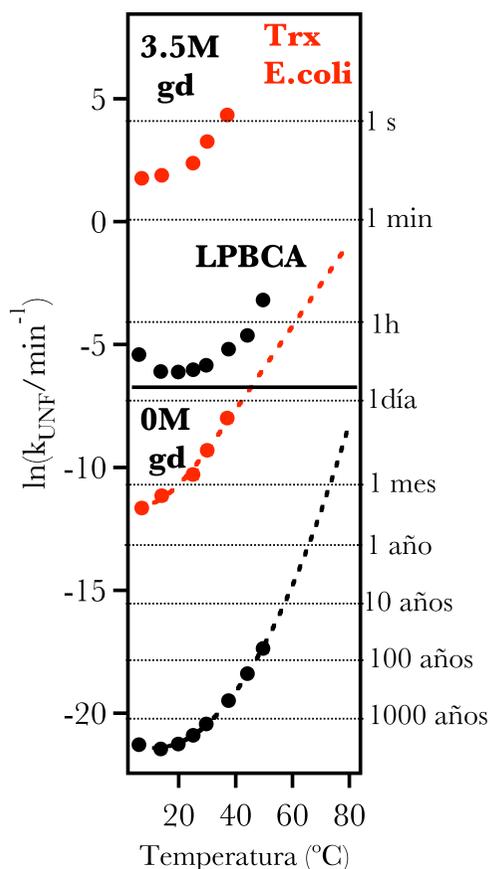


fig 3.2.18. Dependencia con la temperatura de las constante de desplegamiento de la tiorredoxina de *E. coli* (en círculos rojos) y de la tiorredoxina de LPBCA (en círculos negros) a 3.5 M guanidina y a 0M guanidina. Datos obtenidos mediante la extrapolación lineal mostrada en fig 3.2.17. La escala de tiempo se indica a la derecha de la gráfica.

Tras el estudio realizado, cabe resaltar varios puntos:

- La velocidad de desplegamiento para la tiorredoxina de *E. coli* a  $25^{\circ}\text{C}$  en agua es de  $k_{UNF}^0 = 3.40 \cdot 10^{-5} \text{ min}^{-1}$ , que resulta en un tiempo de vida de  $\sim 20$  días. La tiorredoxina de LPBCA tiene una edad geológica de unos 3 giga-

años, la temperatura de los océanos estimada [8] para dicha edad geológica es de 70°C aproximadamente. Para la tiorredoxina de LPBCA la cinética de desplegamiento a 70°C  $k_{\text{UNF}}^0 = 7.65 \cdot 10^{-6} \text{ min}^{-1}$ , que resulta en un tiempo de vida media de  $\sim 40$  días. Ambos tiempos de vida son comparables lo que corrobora la teoría de que la evolución selecciona sobre la estabilidad cinética.

- Existe una buena correlación entre la energía libre de Gibbs de activación de desplegamiento ( $\Delta\Delta G_{\text{UNF}}^\ddagger = 37 \text{ kJ mol}^{-1}$ ) y el valor de energía libre de Gibbs termodinámico ( $\Delta\Delta G = 40 \text{ kJ mol}^{-1}$ ) a una temperatura de referencia de 60°C. Los valores de energía libre de Gibbs de activación de desplegamiento a 60°C se calcularon mediante la extrapolación con la ecuación de Gibbs-Helmholtz (ec.3.2.1) de los datos de  $\Delta G_{\text{UNF}}^\ddagger$  frente a temperatura. Los valores de  $\Delta G_{\text{UNF}}^\ddagger$  se calcularon a partir de los datos experimentales de  $\ln(k_{\text{UNF}})$ , sabiendo que  $k_{\text{UNF}} = k_0 \cdot \exp(-\Delta G_{\text{UNF}}^\ddagger / R \cdot T)$  y utilizando como factor pre-exponencial  $\ln(k_0 / \text{min}^{-1}) = 24.9$  [27]. La correlación entre los valores de  $\Delta\Delta G_{\text{UNF}}^\ddagger$  y  $\Delta\Delta G$  apoya el argumento de la existencia de una selección natural para la estabilidad cinética en la tiorredoxina.



### 3.2.2. Highly Anomalous Energetics of Protein Cold Denaturation Linked to Folding-Unfolding Kinetics

En el transcurso de los experimentos que se realizaron para caracterizar la estabilidad termodinámica de la tiorredoxina de LPBCA, se observó un patrón energético en el desplegamiento extremadamente raro. Los experimentos de calorimetría diferencial de barrido que se realizaron en un rango de temperaturas de 2 a 100°C, a una velocidad de barrido de 90 K/h, con una concentración de tiorredoxina de LPBCA de 5 mg/mL en 50 mM HEPES, 4M guanidina a pH 7, mostraron que si bien en el desplegamiento a alta temperatura se observaba un pico endotérmico, a baja temperatura se observaba algo parecido a una transición de segundo orden, de manera que la desnaturalización fría parecía ocurrir sin efecto entálpico asociado (fig 3.2.19).

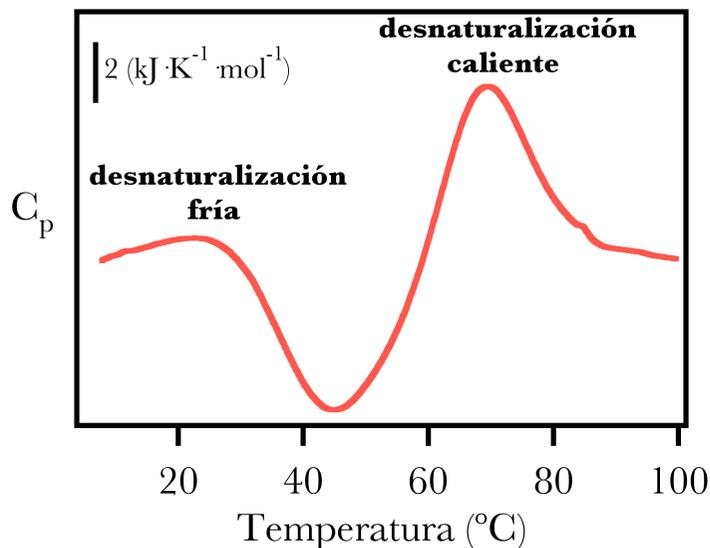


fig 3.2.19. Perfil calorimétrico de la tiorredoxina ancestral de LPBCA. El experimento se llevo a cabo a una velocidad de barrido de 90K/h con una concentración de proteína de 5 mg/mL en 50 mM HEPES, 4M guanidina a pH 7, desde 2 hasta 100°C.

Las evidencias experimentales obtenidas parecían indicar que la desnaturalización fría de la tiorredoxina de LPBCA (estudiada en el sentido del plegamiento) ocurría gradualmente de manera no cooperativa como ocurre en el desplegamiento de FSD-1ss y en la apomioglobina [28, 29]. Esto nos llevó a plantear la hipótesis de que la evolución podría haber seleccionado las tiorredoxinas para que desplegaran a baja temperatura de manera cooperativa. Para confirmar o descartar esta hipótesis realizamos un estudio comparativo del desplegamiento frío entre la tiorredoxina ancestral de LPBCA y su homóloga actual, la tiorredoxina de *E. coli*. Sin embargo, el perfil energético de ambas proteínas mostraron el mismo patrón, en ambos casos la desnaturalización fría parecía ocurrir sin variación entálpica, de manera que se descartó la hipótesis inicialmente planteada. En cualquier caso, los estudios realizados proporcionaron información relevante para comprender el proceso de la desnaturalización fría de proteínas.

A continuación se describe un resumen del estudio detallado del proceso de desnaturalización fría de la tiorredoxina de *E. coli*. publicado en la revista PLoS ONE con el título: “*Highly Anomalous Energetics of Protein Cold Denaturation Linked to Folding-Unfolding Kinetics*”

Aun no está claro si la desnaturalización fría es un mero espejo de la desnaturalización por calor, o si por el contrario tiene rasgos energéticos y estructurales propios. En este trabajo se muestra, para la tiorredoxina de *E. coli*, que los perfiles de desnaturalización por calor y por frío muestran características energéticas drásticamente diferentes. Mientras que la desnaturalización por calor muestra una transición endotérmica, la desnaturalización fría se produce sin apenas efecto de calor asociado, ocurre gradualmente como una transición de segundo orden (véase la figura nombrada como “figura 1C” en el artículo que se inserta más adelante). Dos posibles situaciones podrían explicar este resultado experimental:

- 1) Que la desnaturalización por frío ocurriera de manera no cooperativa
- 2) Que la barrera de energía libre asociada al plegamiento/desplegamiento de la proteína fuera muy alta a baja temperatura, lo que haría que el proceso de desplegamiento fuera extremadamente lento. A mayores temperaturas, aumentaría la velocidad del proceso que alcanzaría una velocidad significativa a temperaturas cercanas a la temperatura de inversión de la entalpía (temperatura a la cual  $\Delta H=0$ ).

Para distinguir entre las dos explicaciones propuestas, nos planteamos estudiar la energética del desplegamiento de la tiorredoxina de *E. coli* por DSC a diferentes velocidades de barrido y estudiar las cinéticas de desplegamiento a diferentes temperaturas. Los resultados obtenidos confirmaron la segunda explicación propuesta, con lo cual, el anómalo perfil encontrado para la desnaturalización fría de la tiorredoxina de *E. coli*, se pudo explicar en base a la lentitud del proceso de desplegamiento a baja temperatura. Sin embargo, a alta temperatura, la velocidad del proceso de desplegamiento aumenta, de manera que en el tiempo en el que transcurre el experimento de DSC, el desplegamiento ocurre en condiciones cercanas al equilibrio.

Nosotros proponemos, que el perfil encontrado para la tiorredoxina de *E. coli* (de desnaturalización caliente en condiciones cercanas al equilibrio y de desnaturalización fría cinéticamente controlada) es posiblemente un patrón muy común en las proteínas. Cabe preguntarse porqué no se ha observado en otros estudios anteriores la fuerte distorsión cinética de la desnaturalización fría. Pensamos que una posible explicación podría ser, que debido al uso de agentes desnaturalizantes para el estudio de la desnaturalización fría (algo que ocurre frecuentemente), como los agentes desnaturalizantes aumentan la constante de velocidad del proceso de desplegamiento, se enmascara la lentitud real del proceso.

De esta manera la cinética emerge como uno de los puntos clave a tener en cuenta a la hora de estudiar procesos de desnaturalización por frío, pudiendo la estabilización cinética a baja temperatura, jugar un papel primordial en la adaptación al frío de organismos psicrófilos.

# Highly Anomalous Energetics of Protein Cold Denaturation Linked to Folding-Unfolding Kinetics

M. Luisa Romero-Romero, Alvaro Inglés-Prieto, Beatriz Ibarra-Molero, Jose M. Sanchez-Ruiz\*

Departamento de Química Física, Facultad de Ciencias, Universidad de Granada, Granada, Spain

## Abstract

Despite several careful experimental analyses, it is not yet clear whether protein cold-denaturation is just a “mirror image” of heat denaturation or whether it shows unique structural and energetic features. Here we report that, for a well-characterized small protein, heat denaturation and cold denaturation show dramatically different experimental energetic patterns. Specifically, while heat denaturation is endothermic, the cold transition (studied in the folding direction) occurs with negligible heat effect, in a manner seemingly akin to a gradual, second-order-like transition. We show that this highly anomalous energetics is actually an apparent effect associated to a large folding/unfolding free energy barrier and that it ultimately reflects kinetic stability, a naturally-selected trait in many protein systems. Kinetics thus emerges as an important factor linked to differential features of cold denaturation. We speculate that kinetic stabilization against cold denaturation may play a role in cold adaptation of psychrophilic organisms. Furthermore, we suggest that folding-unfolding kinetics should be taken into account when analyzing *in vitro* cold-denaturation experiments, in particular those carried out in the absence of destabilizing conditions.

**Citation:** Romero-Romero ML, Inglés-Prieto A, Ibarra-Molero B, Sanchez-Ruiz JM (2011) Highly Anomalous Energetics of Protein Cold Denaturation Linked to Folding-Unfolding Kinetics. *PLoS ONE* 6(7): e23050. doi:10.1371/journal.pone.0023050

**Editor:** Vladimir N. Uversky, University of South Florida College of Medicine, United States of America

**Received:** May 18, 2011; **Accepted:** July 5, 2011; **Published:** July 29, 2011

**Copyright:** © 2011 Romero-Romero et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by FEDER (“Fondo Europeo de Desarrollo Regional”) funds and Grants BIO2009-09562 and CSD2009-00088 from the Spanish Ministry of Science and Innovation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: [sanchezr@ugr.es](mailto:sanchezr@ugr.es)

## Introduction

The existence of cold denaturation is a straightforward prediction of a widely-accepted phenomenological view of protein folding thermodynamics [1]. The simplest rendering of the prediction is as follows. The enthalpy change for protein unfolding ( $\Delta H$ ) is temperature-dependent (as given by the positive unfolding heat capacity change and the Kirchoff equation) and equals zero at the so-called enthalpy-inversion temperature ( $T_H$ ). At temperatures above  $T_H$ ,  $\Delta H$  is positive, the unfolding process is endothermic and, hence, it is driven by a temperature increase (heat denaturation). Below  $T_H$ ,  $\Delta H$  is negative, the unfolding process is exothermic and it is driven by a temperature decrease (cold denaturation).

Several facts make cold denaturation an important subject of study. A large number of organisms (in particular, microorganisms in the oceans) are psychrophiles [2] and cold-adaptation likely involves decreased cold-denaturation propensity in many of their proteins. [It is interesting in this context that recent findings of elevated ubiquitin-conjugated protein levels in Antarctic fish have been interpreted as evidence of cold denaturation *in vivo* [3]]. Also, unfolded and partially-unfolded states may be highly susceptible to irreversible alterations (misfolding and aggregation, for instance) and, hence, cold denaturation may compromise the shelf-life of protein pharmaceuticals upon low-temperature storage. [It is interesting in this context that evidence for cold-denaturation of monoclonal antibodies has been recently reported [4]].

Unfortunately, cold denaturation is difficult to study *in vitro* as the equilibrium cold-denaturation temperature is commonly below

the freezing point of water. With a few relevant exceptions [5,6], experimental cold denaturation *in vitro* is achieved with the help of destabilizing solvent conditions, high pressure, mutations or even core fluorination [7–13]. Furthermore, despite several careful experimental analyses, a number of fundamental issues remain unsolved and a consensus about the molecular mechanism of cold denaturation has yet to emerge [14]. In particular, it is not clear whether cold-denaturation is just a “mirror image” of heat denaturation or whether it shows unique structural and energetic features. In this context, the residual structure in the cold-denatured state and the cooperativity of the cold denaturation process have been frequently addressed in the literature [15–27].

Here we report that, for a well-characterized small protein, heat-denaturation and cold-denaturation show dramatically different experimental energetic patterns. Specifically, while heat denaturation is endothermic, the cold transition (studied in the folding direction) occurs with negligible heat effect ( $\Delta H \sim 0$ ), in (apparent) serious conflict with the widely-accepted view of protein thermodynamics summarized in the opening paragraph. We show that this highly anomalous behavior does not primarily arise from residual-structure or decreased-cooperativity effects; rather, it is an apparent effect that mainly reflects differential kinetic features of cold denaturation that have received little attention in previous studies.

## Results and Discussion

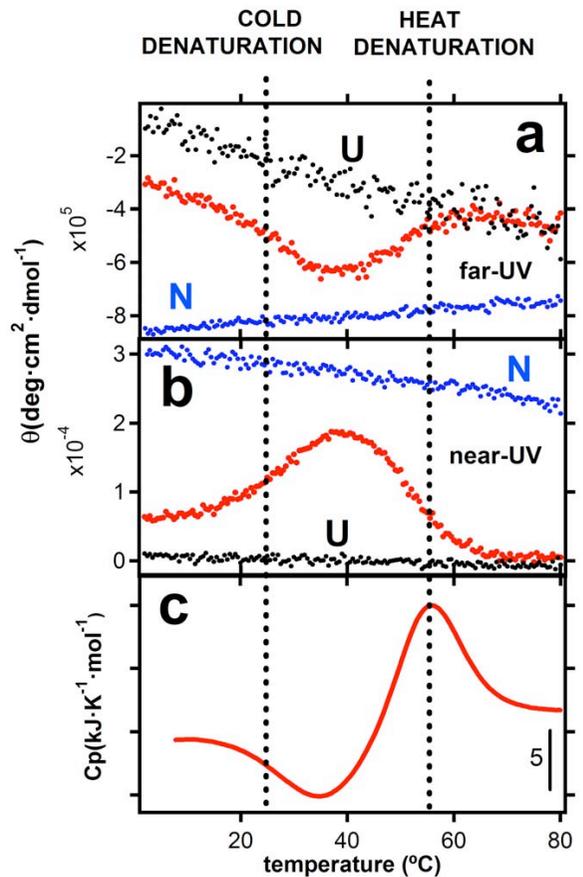
Figures 1a and 1b show the results of far-UV and near-UV circular dichroism experiments in which *E. coli* thioredoxin solutions were heated at a constant temperature scanning-rate of

1.5 degrees/min. [Thioredoxin solutions were kept at 2°C for 3 hours prior to the start of the temperature scanning experiment; we checked, nevertheless, that a several-days incubation at 2°C led essentially to the same results]. Protein ellipticity is an average over protein states and, consequently, folding-unfolding transitions are revealed by sigmoidal-like changes in profiles of ellipticity vs. temperature. The profiles for solutions including 2M guanidine reveal two well-defined sigmoidal-like changes obviously corresponding to the cold and heat denaturation processes. A differential scanning calorimetry (DSC) profile for the same thioredoxin solution and using the same temperature protocol is shown in Figure 1c. DSC measures the heat capacity of a protein solution, which is the temperature-derivative of the average enthalpy. Hence, a folding-unfolding transition is usually revealed by a heat capacity “peak” (the derivative of a sigmoidal-like profile), the area under which equals the total enthalpy change for the transition. Then, DSC thermograms for heat-and-cold denaturation typically show two peaks, often loosely resembling the double hump of a Bactrian camel. The experimental calorimetric profile in Figure 1c is, therefore, most unusual, as a peak is observed only for heat denaturation. Cold denaturation in this profile is actually signaled by a gradual increase in heat capacity upon temperature decrease (roughly from the heat capacity level expected for the native protein to the level expected for the unfolded protein). That is, contrary to heat denaturation, cold denaturation appears to occur with negligible heat effect (close to zero transition enthalpy change).

The two following scenarios may provide a plausible explanation for the astonishing experimental result shown in figure 1: Scenario A) Cold denaturation is a gradual, non-cooperative process, which could be loosely described as a “second-order-like” transition. Actually, this type of interpretation has been previously discussed in connection with the gradual heat-capacity profiles for the denaturation of a compact intermediate state of apomyoglobin and also for the denaturation of FSD-1ss, a designed mini protein studied as a model of primordial folding [28,29]. Scenario B) The free-energy barrier for folding-unfolding is large and, at low temperature, the system is kinetically trapped in the initial distribution of states. Upon temperature increase, the rate of folding/unfolding increases and becomes significant at temperatures close to the enthalpy inversion temperature. Consequently the cold transition occurs in an out-of-equilibrium, kinetically-controlled manner with little heat effect.

To decide between the two explanations proposed above, we have characterized the scanning-rate effect on the DSC transitions, a clear proxy of kinetic control [30]. The heat-denaturation transition is essentially independent of scanning-rate within the range 0.25–4.17 degrees/min, while a large scanning-rate effect is found for the cold-denaturation transition (Figure 2a), indicating a strong kinetic effect in the latter.

Scenario B is further supported by the rates of unfolding determined from the time-dependence of the protein fluorescence at given, constant temperatures (Figure 2b and 2c). The first-order rate constant for thioredoxin unfolding ( $k_U$ ) in 2M guanidine increases with temperature in a manner that is consistent with slow rates at cold-denaturation temperatures and fast rates at heat-denaturation temperatures (see Materials and Methods for a more detailed discussion on the temperature dependence of the unfolding rate constants). Furthermore, the availability of the temperature-dependence of  $k_U$  allows us to carry out a realistic simulation of a temperature-scanning experiment under scenario B. The simulation procedure we have used is essentially that described by Plaza del Pino et al [31]. Very briefly, a temperature-scanning experiment can be seen as a series of very small, equally-

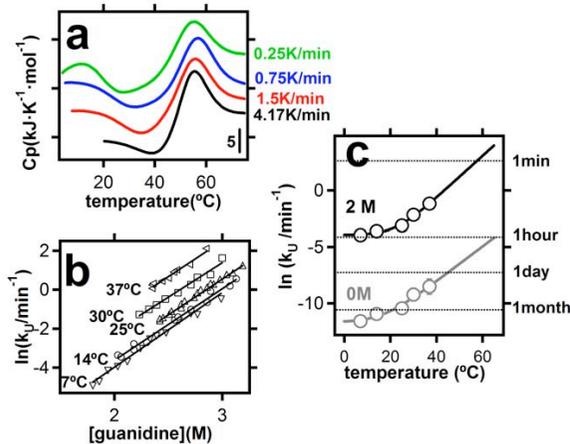


**Figure 1. Heat and cold denaturation of thioredoxin at pH 7 in the presence of 2M guanidine, as followed by, a) far-UV circular dichroism (222 nm), b) near-UV circular dichroism (280 nm) and, c) differential scanning calorimetry.** The three experiments (far-UV, near-UV and DSC) were carried out with the same solution and following exactly the same heating protocol (see text and “Materials and Methods” for details). The native-state (N) and unfolded-state (U) profiles in panels a and b were obtained in the absence of guanidine and in the presence of 6M guanidine, respectively. doi:10.1371/journal.pone.0023050.g001

spaced temperature jumps and the relaxation within each jump can be computed from the folding/unfolding kinetics. Then, the relaxation kinetics, together with the folding/unfolding energetics allows denaturation profiles to be calculated (see Materials and Methods for details).

Simulated profiles of fraction of native protein and heat capacity versus temperature thus obtained are shown in Figure 3. They do reproduce the relevant features of the experimental results summarized in figure 1, including in particular a) the strong scan-rate dependence of the cold-denaturation transition, b) the appearance of a cold-denaturation “peak” at the lower scan-rates (i.e., when equilibrium is approached) and c), the shift towards lower temperatures of the heat capacity minimum between the heat and cold denaturation transitions which is observed upon scan-rate decrease.

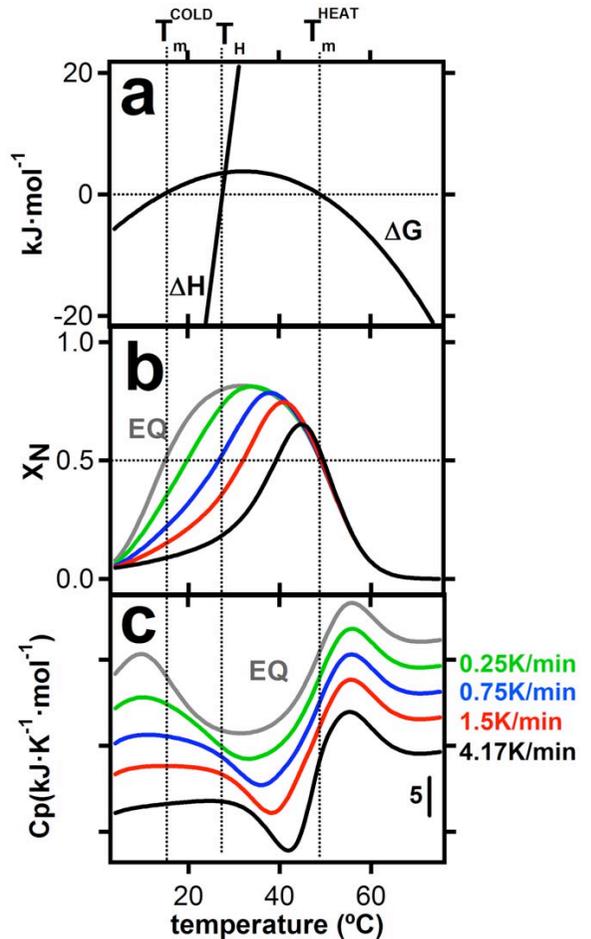
It is important to note that the simulations shown in figure 3 are based upon reasonable estimates of the energetic parameters of thioredoxin unfolding (see Figure 3a and Materials and Methods



**Figure 2. Cold denaturation and folding-unfolding kinetics.** a) Scan-rate effect on the DSC profiles for thioredoxin heat and cold denaturation at pH 7 and in the presence of 2M guanidine. All the scanning experiments started at 2°C. The differences observed in the lowest temperature are related to the instrument equilibration time (see Materials and Methods for details). b) First-order rate constants for thioredoxin unfolding as a function of guanidine concentration for the temperatures shown. c) Temperature-dependence of the rate constants for thioredoxin unfolding at 2M guanidine and in the absence of guanidine (extrapolated in both cases from the linear dependencies shown in panel b). Representative time-scales are shown alongside the right y-axis.  
doi:10.1371/journal.pone.0023050.g002

for details), together with the temperature dependence of the unfolding rate constant based on experiment (figure 2c). Therefore, the general agreement found does support that the anomalous features of the cold denaturation of thioredoxin are mainly associated to kinetics (although, of course, contributions from other factors, such as residual structure in the unfolded state and decreased cooperativity, cannot be ruled out). Clearly, in a laboratory time scale on the order of minutes-to-hours (corresponding to heating rates of few degrees per minute in a temperature-scanning experiment) kinetic trapping prevents thioredoxin cold transition from taking place at the equilibrium cold-denaturation temperature; rather the process occurs at about the enthalpy-inversion temperature with little heat effect. As elaborated below in some detail, we expect this kinetic effect on cold-denaturation to be a common feature of many proteins that, as is the case with thioredoxin [32–34], have been naturally selected to have significant kinetic stability.

Proteins often work *in vivo* under conditions that favor the occurrence of irreversible alteration processes. For instance, proteolysis, aggregation and incorrect interactions with other macromolecular components may potentially occur in crowded intracellular or harsh extracellular environments. As a result, a protein might not remain in the native, biological functional state in a physiological-relevant time scale, even if the protein is stable in the usual thermodynamic sense (i.e., even if the free energy of the native state is lower than that of the unfolded state under physiological conditions). It is emerging from a number of different experimental studies [35–41] that, to avoid this situation, many proteins have been naturally selected to have a high free-energy barrier for unfolding under physiological conditions. Such high barrier “separates” the biologically functional state from inactive states and unfolded (and partially unfolded) states (i.e., the states



**Figure 3. Theoretical simulation of temperature-scanning profiles for protein heat-cold denaturation including slow folding-unfolding kinetics.** a) Folding-unfolding energetics used in the simulation (see Materials and Methods for details). Profiles of unfolding free energy and unfolding enthalpy versus temperature are shown. b) Plots of fraction of native protein versus temperature for several values of the temperature scanning-rate. c) Simulated scanning calorimetry profiles at the indicated scan rates. The equilibrium profile (formally corresponding to the limit of zero scan rate) is labeled EQ. Vertical dashed lines crossing the three panels indicate the values of the equilibrium heat- and cold-denaturation temperatures at the enthalpy inversion temperature.  
doi:10.1371/journal.pone.0023050.g003

prone to undergo irreversible alteration processes) and confers kinetic stability to the protein. Certainly, the high free-energy barrier associated to kinetic stability will lead to slow folding-unfolding kinetics at physiological temperatures, although, because of the expected increase of rate with temperature, slow kinetics will not be necessarily observed at the higher heat-denaturation temperatures. Thansthyretin, a protein involved in several misfolding diseases, provides a particularly clear illustration of the point. *In vitro* heat denaturation of wild type transthyretin occurs at about 100°C and can be adequately described in terms of a two-state mechanism involving equilibrium between the native tetrameric protein and the monomeric unfolded state [42]. Despite of this, recent successful approaches to treat transthyretin

amyloidoses (obviously at temperatures much lower than 100°C) use kinetic stabilizers, i.e., substances that have the capability of increasing the unfolding free-energy barrier, thus enhancing transthyretin kinetic stability [43,44].

It is also worth noting that a strong dissimilarity between the time-scales for the heat- and cold-denaturation processes is linked to a strong temperature dependence of the folding/unfolding rate constants and implies, therefore, high values for the corresponding activation energies. Activation energy values on the order of several hundreds of kJ/mol have been reported for the unfolding of many proteins and attributed in part to the asynchrony between water penetration and the break-up of internal interactions [45–47]. Indeed, such solvation/desolvation barriers have been shown to determine several relevant features of protein folding [48] and to be important for protein kinetic stability [41,45].

We conclude from all the above experiments, analyses and considerations, that the pattern of close-to-equilibrium heat denaturation (in the absence of irreversible alterations, such as aggregation) and kinetically-controlled cold-denaturation may actually be a quite common one. We may ask then why huge kinetic effects do not appear to have been noticed in many previous studies on cold denaturation. The answer is likely related to the fact that cold denaturation is often achieved *in vitro* with the help of destabilizing conditions that may increase, not only the equilibrium cold denaturation temperature, but also the rate of unfolding. Actually, the attenuation of the kinetic effect is clearly seen in our experimental data, as the presence of guanidine leads to higher rates of thioredoxin unfolding (compare in figure 2c the rates at 2M guanidine with the rates extrapolated at zero denaturant concentration). However, the rates at 2M are still low and the kinetic effect in thioredoxin is not fully eliminated by the destabilizing agent. That is, the kinetic effect we observe on thioredoxin cold-denaturation is not caused by the chemical denaturant but, rather, it occurs despite the presence of a chemical denaturant. The high kinetic stability of thioredoxin [32–34], together with the comparatively low guanidine concentration used, may have contributed to this result. It may be reasonable expected that, in many other cases, the destabilizing conditions used to bring cold-denaturation to an experimentally convenient temperature range have efficiently masked the kinetic effect in a laboratory time-scale.

Overall, kinetics emerges as an important factor linked to differential features of protein cold denaturation. Kinetics should be taken into account when analyzing *in vitro* cold-denaturation experiments, in particular in the absence of destabilizing conditions (using, for instance, water supercooling) or when designing conditions for the low-temperature storage of protein pharmaceuticals. Furthermore, we speculate that kinetic stabilization against cold-denaturation may play an important role in cold-adaptation of psychrophilic organisms.

## Materials and Methods

### Protein solutions

Wild type *E. Coli* thioredoxin was expressed and purified as previously described [32]. Stock protein solutions in aqueous buffer (50 mM Hepes pH 7) were prepared by exhaustive dialysis against the buffer. Stock solutions of 7M guanidinium chloride in Hepes buffer were prepared by mixing two guanidine solutions in the acidic and basic forms of the buffer used, so that the desired pH-meter reading was obtained. Thioredoxin solutions at 2M guanidine chloride were prepared by mixing appropriate volumes of stock solutions of the protein and 7 M guanidine chloride. Guanidinium chloride concentrations were determined from refraction index measurements using an Abbe 320 refractometer.

### Diferencial scanning calorimetry (DSC)

DSC experiments were carried out in a capillary VP-DSC microcalorimeter (Microcal, Northhampton, USA) with a 0.133 mL cell volumen. Experiments at different scan rates (0.25, 0.75, 1.5 and 4.2 K/min) were performed. In all cases, the protein solution was kept inside the calorimetric cell for 180 min at 2°C before the start of the temperature up-scanning experiments. However, we checked that several-days incubation at 2°C leads to essentially the same results. Note that, although all scanning experiments started at 2°C, the lowest experimental data point depends on the scan rate used (see Figure 2a). This is due to the fact that, during the instrumental equilibration period (about 4 min after the start of the temperature up-scanning), heat capacity values cannot be recorded. The “lost” temperature interval is given by the value of the equilibration period times the scan rate used and, obviously increases with scan rate.

Reheating runs were performed in all cases and demonstrated high operational reversibility of the heat-capacity denaturation profile. The DSC thermograms shown in Figures 1a and 2a are derived from experiments performed at a protein concentration of 4.6 mg/mL, which guarantees an excellent signal-to-noise ratio and minimizes baseline uncertainty effects commonly associated with calorimetric work with water-cosolvent mixtures. Nevertheless, we did carry out additional DSC experiments at different protein concentrations and found no significant protein concentration effect on the calorimetric profiles within the 0.8–4.6 mg/mL range, thus ruling out the possibility of distortions caused by protein oligomerization.

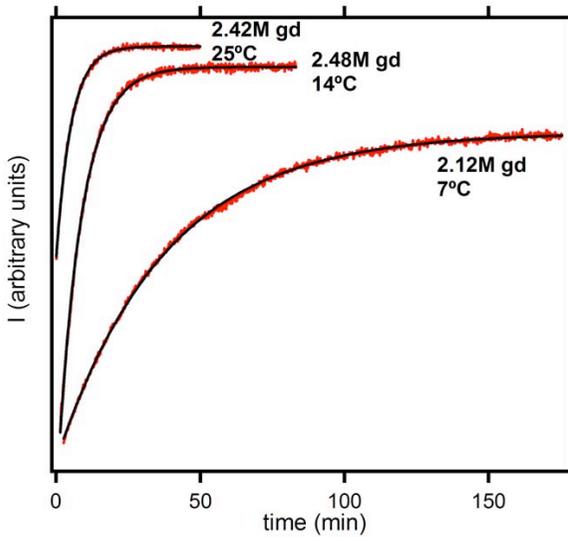
### Circular dichroism (CD) spectroscopy

CD experiments were carried out using a Jasco (Tokyo, Japan) J-715 spectropolarimeter equipped with a PTC-348WI. The change in CD signal within the temperature range of 2–98°C was monitored following a protocol identical to that used in the DSC experiments, i.e., keeping the protein solution in the measurement cell at 2°C for 180 min prior to the start of the upwards temperature-scan. A signal averaging time of 4 s and a bandwidth of 1 nm were used in all cases. For far-UV CD experiments (222 nm), a protein concentration of ~0.5 mg/mL and 1 mm pathlength cell were used. For near-UV CD experiments (280 nm) a protein concentration of ~0.9 mg/mL and 0.5 mm pathlength cell were used. In all cases, blanks in the absence of protein were recorded immediately prior to the sample measurements and subtracted prior to the mean molar ellipticity [ $\theta$ ] calculation.

### Unfolding kinetics studied by steady-state fluorescence measurements

Unfolding kinetics in 50 mM Hepes pH 7 in the presence of different guanidine concentrations were monitored by following the time dependence of the fluorescence emission at 350 nm (excitation at 276 nm) after suitable guanidine concentration jumps (100-fold dilution from zero guanidine concentration). Experiments were carried at different temperatures (7°C, 14°C, 25°C, 30°C and 37°C) and, for each temperature, at several guanidine concentrations. See Figure 4 for representative examples of the experimental fluorescence intensity versus time profiles. Values of the first-order rate constant ( $k_U$ ) were calculated from fittings of a single exponential to the experimental fluorescence intensity versus time data.

For each temperature, plots of logarithm of rate constant versus guanidine concentration (C) were linear (see Figure 2b), indicating that the experimentally studied guanidine concentration ranges correspond to the unfolding branches of the chevron plots. The



**Figure 4. Representative examples of fluorescence intensity versus time profiles for thioredoxin unfolding.** The numbers alongside the profiles stand for the temperature and the guanidine concentration used. Experimental data are shown in red and the black line represents the best fit of a single-exponential. doi:10.1371/journal.pone.0023050.g004

linear dependence can be written as,

$$\ln k_U = \ln k_U^0 - \frac{m_U}{RT} C \tag{1}$$

where  $m_U$  is the unfolding kinetic m-value, i.e., the minus derivative of the unfolding activation free-energy with respect to the concentration of chemical denaturant. Kinetic  $m_U$  values are commonly taken as a measure of exposure to solvent of the folding-unfolding transition state as compared with that of the native state. Experimental kinetic  $m_U$  values for thioredoxin unfolding (derived from the slopes of the plots in Figure 2b) do not significantly change with temperature within the studied range (7–37°C).  $k_U^0$  in equation 1 is the linear-extrapolation estimate of the unfolding rate constant for zero guanidine concentration. The values thus extrapolated (see values labeled “0M” on Figure 2c) must be viewed with due caution, as deviations from the linear extrapolation model have been reported at low guanidine concentrations and linked to the screening of charge-charge interactions [49]. Still, the extrapolated values are very useful as a qualitative indication of the time-scale of the unfolding process in the absence of chemical denaturants (see Figure 2c). Furthermore, deviations from the linear-extrapolation model have been detected at low guanidine concentration [49] and do not affect the extrapolation/interpolation of the linear dependencies to 2M guanidine (i.e., the values labeled “2M” in Figure 2c).

The value of  $\ln k_U$  at 2M guanidine increases with temperature in a clearly nonlinear fashion (see Figure 2c) and, indeed, Arrhenius plots (not shown) are clearly non-linear. Within the context of transition-state theory, this may be rationalized in terms of a significant value of the activation heat capacity change, although, a rugged energy landscape (implying super-Arrhenius dependence) could also contribute to the observed nonlinearity [50–52]. For the purpose of the computational simulations of

denaturation profiles (see further below) we have bypassed the issue of the molecular origin of the non-linear temperature dependence and used instead the following phenomenological equation to describe the  $k_U(T)$  dependence:

$$\ln k_U = \ln \{ \alpha + \beta \cdot e^{\gamma T} \} \tag{2}$$

where  $\alpha$ ,  $\beta$  and  $\gamma$  are fitting parameters. Fits of equation 2 to the experimental  $\ln k_U$  values are shown in Figure 2c.

**Computational simulation of heat-cold denaturation profiles including the effect of folding-unfolding kinetics**

The simulation procedure we have used is essentially that described by Plaza del Pino et al. [31]. Briefly, a temperature-scanning experiment can be seen as a series of very small, equally-spaced temperature jumps. For a two-macrostate (native-unfolded) model, the relaxation kinetics after the jump from  $T_i$  to  $T_{i+1}$  is given by,

$$X_N(T_{i+1}) = X_N^{EQ}(T_{i+1}) + \{ X_N(T_i) - X_N^{EQ}(T_{i+1}) \} \cdot \exp\left(-\frac{\Delta T}{v\tau}\right) \tag{3}$$

where  $X_N$  is the fraction of native protein,  $v$  is the scanning rate,  $\Delta T = T_{i+1} - T_i$  ( $= T_{i+2} - T_{i+1} = T_{i+3} - T_{i+2} = \dots$ ) and  $\tau$  is the folding-unfolding relaxation time,

$$\tau = \frac{1}{k_U + k_F} = \frac{1}{k_U \cdot (1 + 1/K)} \tag{4}$$

where we have used that the unfolding equilibrium constant equals the ratio of the unfolding to the folding rate constants ( $K = k_U/k_F$ ). The fraction of native protein at equilibrium in equation 4 is also related to  $K$ ,

$$X_N^{EQ} = \frac{1}{1 + K} \tag{5}$$

Accordingly,  $X_N$  versus temperature profiles can be computed from the repeated application of equation 3, provided that the temperature dependencies of  $K$  and  $k_U$  are known or assumed. These profiles can in turn be transformed into simulated DSC thermograms by using,

$$C_P^{EX} = -\Delta H \cdot \left(\frac{dX_N}{dT}\right) + \Delta C_P \cdot (1 - X_N) \tag{6}$$

where  $C_P^{EX}$  is an excess heat capacity defined with respect to the native heat capacity level, and  $\Delta C_P$  and  $\Delta H$  are, respectively, the unfolding changes in heat capacity and enthalpy.

Application of equations 3–6 to the simulation of denaturation profiles requires that the temperature dependencies of the rate constant  $k_U$ , the equilibrium constant  $K$ , the unfolding enthalpy ( $\Delta H$ ) and heat capacity ( $\Delta C_P$ ) are known (or reasonably estimated) as a function of temperature. The temperature dependence of the unfolding rate constant is known from experiment and adequately represented (Figure 2c) by equation 2. The temperature

dependencies of  $K$ ,  $\Delta H$  and  $\Delta C_p$  can be described on the basis of well-known, straightforward thermodynamics:

$$\Delta C_p = \Delta C_{p,R} + (d\Delta C_p/dT) \cdot (T - T_R) \quad (7)$$

$$\Delta H = \Delta H_R + \Delta C_{p,R} \cdot (T - T_R) + \frac{d\Delta C_p/dT}{2} (T - T_R)^2 \quad (8)$$

$$\Delta S = \Delta S_R + \{ \Delta C_{p,R} - (d\Delta C_p/dT) \cdot T_R \} \cdot \ln\left(\frac{T}{T_R}\right) + (d\Delta C_p/dT) \cdot (T - T_R) \quad (9)$$

$$\Delta G = \Delta H - T \cdot \Delta S \quad (10)$$

$$K = \exp\left(-\frac{\Delta G}{RT}\right) \quad (11)$$

In the above equations  $\Delta S$  and  $\Delta G$  are, respectively, the unfolding changes in entropy and free energy (the latter being immediately related with the value of the equilibrium constant through equation 11),  $\Delta H_R$ ,  $\Delta S_R$  and  $\Delta C_{p,R}$  are the unfolding changes in enthalpy, entropy and heat capacity at a given reference temperature ( $T_R$ ) and, finally, we take into account the possibility that the unfolding heat capacity change depends on temperature according to a constant slope  $d\Delta C_p/dT$ .

Simulations were started at a temperature of 2°C with the equilibrium distribution of protein states at that temperature. This is intended to mimic the experimental protocol, which involved a 180 min equilibration at 2°C before the start of the upward temperature-scan. In all the computer simulations shown

(Figure 3c) a very small temperature interval ( $\Delta T$  in equation 3) of 0.04 degrees was used. Nevertheless, we checked that using ten times-larger intervals led to the same results.

The energetic parameters used for the simulations shown in Figure 3c are:

heat denaturation temperature = 49°C,

unfolding enthalpy change at 25°C = -51.6 kJ/mol (note that 25°C is below the enthalpy inversion temperature and, consequently,  $\Delta H < 0$  at 25°C),

unfolding heat capacity change at 25°C = 7.2 kJ·K<sup>-1</sup>·mol<sup>-1</sup>,  
 $d\Delta C_p/dT = 0.1$  kJ·K<sup>-2</sup>·mol<sup>-1</sup>.

These values were arrived at as follows. Initial simulations were carried using the literature estimates of the unfolding heat capacity change, the heat-denaturation temperature and the unfolding enthalpy [53,54]. It was soon realized, however, that simulations using a temperature-independent unfolding heat capacity (i.e.  $d\Delta C_p/dT = 0$  in equations 7–9) could not reproduce the experimental denaturation profiles and that a positive  $d\Delta C_p/dT$  value was required to bring cold denaturation within the studied temperature range. It is interesting that  $d\Delta C_p/dT > 0$  implies that the  $\Delta C_p$  value at cold-denaturation temperatures is smaller than the value at heat-denaturation temperatures, and, therefore, may be interpreted as suggesting hydrophobic burial in the cold-denatured state. Finally, the energetic parameters were slightly modified (i.e., fine-tuned) to achieve a qualitative agreement of the predicted denaturation profiles with the experimental ones. Note, however, that the description of the folding-unfolding kinetics (based upon the experimental unfolding rate data of figure 2c, together with equations 2, 4 and 5) was not modified or fine-tuned in any way.

## Author Contributions

Conceived and designed the experiments: MLR-R BI-M JMS-R. Performed the experiments: BI-M AI-P MLR-R. Analyzed the data: BI-M MLR-R. Wrote the paper: JMS-R.

## References

- Privalov P (1990) Cold denaturation of proteins. *Crit Rev Biochem Mol Biol* 25: 281–305.
- Siddiqui KS, Cavicchioli R (2006) Cold-adapted enzymes. *Annu Rev Biochem* 75: 403–433.
- Todgham AE, Hoaglund EA, Hofmann GEJ (2007) Is cold the new hot? Elevated ubiquitin-conjugated protein levels in tissues of Antarctic fish as evidence for cold-denaturation of proteins in vivo. *J Comp Physiol B* 177: 857–866.
- Lazar KL, Patapoff TW, Sharma VK (2010) Cold denaturation of monoclonal antibodies. *MAbs* 2: 42–52.
- Pometum MS, Peterson RW, Babu CR, Wand AJ (2006) Cold denaturation of encapsulated ubiquitin. *J Am Chem Soc* 128: 10652–10653.
- Pastore A, Martin SR, Politou A, Kondapalli KC, Stemmler T, et al. (2007) Unbiased cold denaturation: low- and high-temperature unfolding of yeast frataxin under physiological conditions. *J Am Chem Soc* 129: 5374–5375.
- Privalov PL, Griko YV, Venyamov SY, Kutysenko VP (1986) Cold denaturation of myoglobin. *J Mol Biol* 190: 487–498.
- Griko YV, Privalov PL, Sturtevant JM, Venyamov SY (1988) Cold denaturation of staphylococcal nuclease. *Proc Natl Acad Sci U S A* 85: 3343–3347.
- Chen BL, Schellman JA (1989) Low-temperature unfolding of a mutant of phage T4 lysozyme. 1. Equilibrium studies. *Biochemistry* 28: 685–691.
- Antonino LC, Kautz RA, Nakano T, Fox RO, Fink AL (1991) Cold denaturation and 2H<sub>2</sub>O stabilization of a staphylococcal nuclease mutant. *Proc Natl Acad Sci U S A* 88: 7715–7718.
- Ibarra-Molero B, Makhadze GI, Sanchez-Ruiz JM (1999) Cold denaturation of ubiquitin. *Biochim Biophys Acta* 1429: 384–390.
- Yang WY, Grubele M (2005) Kinetic equivalence of the heat and cold structural transitions of lambda<sub>6</sub>-85. *Philo Transact A Math Phys Eng Sci* 363: 565–573.
- Molski MA, Goodman JL, Craig CJ, Meng H, Kumar K, et al. (2010) Beta-peptide bundles with fluororous cores. *J Am Chem Soc* 132: 3658–3659.
- Ascolece E, Graziano G (2008) On the cold denaturation of globular proteins. *Chem Phys Lett* 467: 150–153.
- Griko YuV, Venyamov SYu, Privalov PL (1989) Heat and cold denaturation of phosphoglycerate kinase (interaction of domains). *FEBS Lett* 244: 276–278.
- Griko YV, Privalov PL (1992) Calorimetric study of the heat and cold denaturation of beta-lactoglobulin. *Biochemistry* 31: 8810–8815.
- Freire E, Murphy KP, Sanchez-Ruiz JM, Galisteo ML, Privalov PL (1992) The molecular basis of cooperativity in protein folding. Thermodynamic dissection of interdomain interactions in phosphoglycerate kinase. *Biochemistry* 31: 250–256.
- Griko YuV, Kutysenko VP (1994) Differences in the processes of beta-lactoglobulin cold and heat denaturations. *Biophys J* 67: 356–363.
- Wong KB, Freund SM, Fersht AR (1996) Cold denaturation of barstar: 1H, 15N and 13C NMR assignment and characterisation of residual structure. *J Mol Biol* 259: 805–818.
- Huang GS, Oas TG (1996) Heat and cold denatured states of monomeric lambda repressor are thermodynamically and conformationally equivalent. *Biochemistry* 35: 6173–6180.
- Richardson JM, Lemaire SD, Jacquot JP, Makhadze GI (2000) Difference in the mechanisms of the cold and heat induced unfolding of thioredoxin h from *Chlamydomonas reinhardtii*: spectroscopic and calorimetric studies. *Biochemistry* 39: 11154–11162.
- Babu CR, Hilsner VJ, Wand AJ (2004) Direct access to the cooperative substructure of proteins and the protein ensemble via cold denaturation. *Nature Struct Mol Biol* 11: 352–357.
- Dyer RB, Maness SJ, Franzen S, Fesinmeyer RM, Olsen KA, et al. (2005) Hairpin folding dynamics: the cold-denatured state is predisposed for rapid refolding. *Biochemistry* 44: 10406–10415.
- Whiten ST, Kurtz AJ, Pometum MS, Wand AJ, Hilsner VJ (2006) Revealing the nature of the native state ensemble through cold denaturation. *Biochemistry* 45: 10163–10174.
- Li Y, Shan B, Raleigh DP (2007) The cold denatured state is compact but expands at low temperatures: hydrodynamic properties of the cold denatured state of the C-terminal domain of L9. *J Mol Biol* 368: 256–262.
- Andover M, Esposito V, Martorell G, Pastore A, Temussi PA (2010) Understanding cold denaturation: the case study of Yfh1. *J Am Chem Soc* 132: 16240–16246.

27. Shan B, McClendon S, Rospigliosi C, Eliezer D, Raleigh DP (2010) The cold denatured state of the C-terminal domain of protein L9 is compact and contains both native and non-native structure. *J Am Chem Soc* 132: 4669–4677.
28. Griko YuV, Privalov PL (1994) Thermodynamic puzzle of apomyoglobin unfolding. *J Mol Biol* 235: 1318–1325.
29. Saqdi M, de Alba E, Perez-Jimenez R, Sanchez-Ruiz JM, Muñoz V (2009) A designed protein as experimental model of primordial folding. *Proc Natl Acad Sci U S A* 106: 4127–4132.
30. Freire E, van Osdol WW, Mayorga OL, Sanchez-Ruiz JM (1990) Calorimetrically determined dynamics of complex unfolding transitions in proteins. *Annu Rev Biophys Biophys Chem* 19: 159–188.
31. Plaza del Pino IM, Pace CN, Freire E (1992) Temperature and guanidine hydrochloride dependence of the structural stability of ribonuclease T1. *Biochemistry* 31: 11196–11202.
32. Godoy-Ruiz R, Ariza F, Rodriguez-Larrea D, Perez-Jimenez R, Ibarra-Molero B, 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: 966–978.
33. Tur-Arlandis G, Rodriguez-Larrea D, Ibarra-Molero B, Sanchez-Ruiz JM (2010) Proteolytic scanning calorimetry: a novel methodology that probes the fundamental features of protein kinetic stability. *Biophys J* 98: L12–14.
34. Rodriguez-Larrea D, Perez-Jimenez R, Sanchez-Romero I, Delgado-Delgado A, Fernandez JM, Sanchez-Ruiz JM (2010) Role of conservative mutations in protein multi-property adaptation. *Biochem J* 429: 243–249.
35. Plaza del Pino IM, Ibarra-Molero B, Sanchez-Ruiz JM (2000) Lower kinetic limit to protein thermal stability: a proposal regarding protein stability in vivo and its relation with misfolding diseases. *Proteins* 40: 58–70.
36. Persikov AV, Brodsky B (2002) Unstable molecules form stable tissues. *Proc Natl Acad Sci U S A* 99: 1101–1103.
37. Jaswal SS, Sohl JL, Davis JH, Agard DA (2002) Energetic landscape of alpha-helical protease optimizes longevity through kinetic stability. *Nature* 415: 343–346.
38. Mehta R, Gantz DL, Gursky O (2003) Human plasma high-density lipoproteins are stabilized by kinetic factors. *J Mol Biol* 328: 183–192.
39. Xia K, Manning M, Hesham H, Lin Q, Byströff C, Colón W (2007) Identifying the subproteome of kinetically stable proteins via diagonal 2D SDS/PAGE. *Proc Natl Acad Sci U S A* 104: 17329–17334.
40. Halskau O, Perez-Jimenez R, Ibarra-Molero B, Underhaug J, Muñoz V, et al. (2008) Large-scale modulation of thermodynamic protein folding barriers linked to electrostatics. *Proc Natl Acad Sci U S A* 105: 8625–8630.
41. Sanchez-Ruiz JM (2010) Protein kinetic stability. *Biophys Chem* 148: 1–15.
42. Shnyrov VL, Villar E, Zhadan GG, Sanchez-Ruiz JM, Quintas A, et al. (2000) Comparative calorimetric study of non-amyloidogenic and amyloidogenic variants of the homotetrameric protein transthyretin. *Biophys Chem* 88: 61–67.
43. Hammarström P, Wiseman RL, Powers ET, Kelly JW (2003) Prevention of transthyretin amyloid disease by changing protein misfolding energetics. *Science* 299: 713–716.
44. Connolly S, Choi S, Johnson SM, Kelly JW, Wilson IA (2010) Structure-based design of kinetic stabilizers that ameliorate the transthyretin amyloidoses. *Curr Opin Struct Biol* 20: 54–62.
45. Rodriguez-Larrea D, Minning S, Borchert TV, Sanchez-Ruiz JM (2006) Role of solvation barriers in protein kinetic stability. *J Mol Biol* 360: 715–724.
46. Rodriguez-Larrea D, Ibarra-Molero B, Tur-Arlandis G, Sanchez-Ruiz JM (2006) Energetic and structural consequences of desolvation/solvation barriers to protein folding/unfolding assessed from experimental unfolding rates. *Biophys J* 91: L48–50.
47. Costas M, Rodriguez-Larrea D, De María L, Borchert TV, Gómez-Puyou A, Sanchez-Ruiz JM (2009) Between-species variation in the kinetic stability of TIM proteins linked to solvation-barrier free energies. *J Mol Biol* 385: 924–937.
48. Chan HS, Zhang Z, Wallin S, Liu Z (2011) Cooperativity, local-nonlocal coupling, and nonnative interactions: principles of protein folding from coarse-grained models. *Annu Rev Phys Chem* 62: 301–326.
49. Ibarra-Molero B, Loladze VV, Makhatazde GI, Sanchez-Ruiz JM (1999) Thermal versus guanidine-induced unfolding of ubiquitin. An analysis in terms of the contributions from charge-charge interactions to protein stability. *Biochemistry* 38: 8138–8149.
50. Akmal A, Muñoz V (2004) The nature of the free energy barriers to two-state folding. *Proteins* 57: 142–152.
51. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG (1995) Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* 21: 167–195.
52. Scalley ML, Baker D (1997) Protein folding kinetics exhibit an Arrhenius temperature dependence when corrected for the temperature dependence of protein stability. *Proc Natl Acad Sci U S A* 94: 10636–10640.
53. Georgescu RE, Garcia-Mira MM, Tasayco ML, Sanchez-Ruiz JM (2001) 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.
54. Perez-Jimenez R, Godoy-Ruiz R, Ibarra-Molero B, Sanchez-Ruiz JM (2004) The efficiency of different salts to screen charge interactions in proteins: a Hofmeister effect? *Biophys J* 86: 2414–2429.



### **3.2.3. Rediseño del centro activo de la tiorredoxina de LPBCA mediante la simulación mutacional de una carga enterrada**

La estabilidad marginal de las proteínas limita muchos objetivos de ingeniería de proteínas, como por ejemplo, el diseño de nuevos sitios activos. Esto hace necesario la optimización previa de la proteína bajo su estabilidad, con objeto de obtener un armazón molecular lo suficientemente estable capaz de alojar las mutaciones necesarias para conformar el nuevo sitio activo. En este sentido, las proteínas ancestrales, pueden ser muy apropiadas como punto de partida para objetivos de ingeniería de proteínas, ya que al ser tan estables, permiten acumular mutaciones desestabilizantes sin que se comprometa su viabilidad. En este trabajo, nos planteamos la posibilidad de partir de una enzima ancestral para desarrollar un proyecto de ingeniería de proteínas, una línea de investigación que hasta lo que sabemos, no ha sido explorada aun. En concreto, hemos investigado la posibilidad de rediseñar el centro activo de una variante de la tiorredoxina ancestral de LPBCA.

Frecuentemente, el desarrollo de la función de una proteína está relacionado con la presencia de residuos ionizables en su interior hidrofóbico, que son los responsables de modular propiedades especiales en la proteína, vinculadas a su función biológica (por ejemplo, la capacidad de modificar el valor del pK, la alta nucleofilia o la capacidad para mejorar la unión a determinados grupos químicos del sustrato [30]). El problema es que introducir residuos ionizables enterrados suelen ocasionar cambios muy drásticos en la proteína. Por ello, nos planteamos explorar la posibilidad de rediseñar el centro activo de la tiorredoxina, de manera que el efecto de un grupo ionizable enterrado, esencial para la catálisis enzimática, es llevado a cabo por un grupo expuesto al disolvente. Para abordar este objetivo nos planteamos el estudio de una biblioteca combinatorial sobre una variante de la tiorredoxina de LPBCA que ha sido previamente optimizada, en nuestro laboratorio, desde el punto de vista de su estabilidad. Esta variante

contiene las mutaciones V55F, Q69K y D104E en la tiorredoxina de LPBCA, la selección de dichas mutaciones se realizó como se describe en el capítulo 3.2.1.1.1. donde ha sido referida como variante triple. Su  $T_m$  es de 128°C, lo que nos permite disponer de un amplio margen de estabilidad para poder elegir las mutaciones que presumiblemente llevan a cabo el efecto en la catálisis enzimática del grupo ionizable enterrado, sin que se comprometa su viabilidad.

Por último, para el estudio de la biblioteca combinatorial planteamos el uso de una técnica de regresión por mínimos cuadrados parciales como herramienta de predicción del total de la biblioteca a partir del estudio experimentales de unas pocas variantes.

### **3.2.3.1. Diseño de la biblioteca combinatorial**

Las tiorredoxinas contienen dos cisteínas en el sitio activo siguiendo un motivo CXXC. La química de la reacción disulfuro oxidorreductasa catalizada por las tiorredoxinas contiene dos etapas:

- 1) La cisteína N-terminal de la tiorredoxina (numerada como Cys32 en tiorredoxina de *E. coli*) en su forma reducida ataca nucleofílicamente al disulfuro del sustrato. Se forma, con lo cual, un intermedio que contiene un enlace disulfuro mixto entre la Cys32 de la tiorredoxina y la cisteína del sustrato.
- 2) La cisteína C-terminal de la tiorredoxina (Cys35 en *E. coli*) ataca al enlace disulfuro del intermedio produciendo la reducción de la cisteína del sustrato y la oxidación de la cisteína N-terminal de la tiorredoxina (fig 3.2.20).

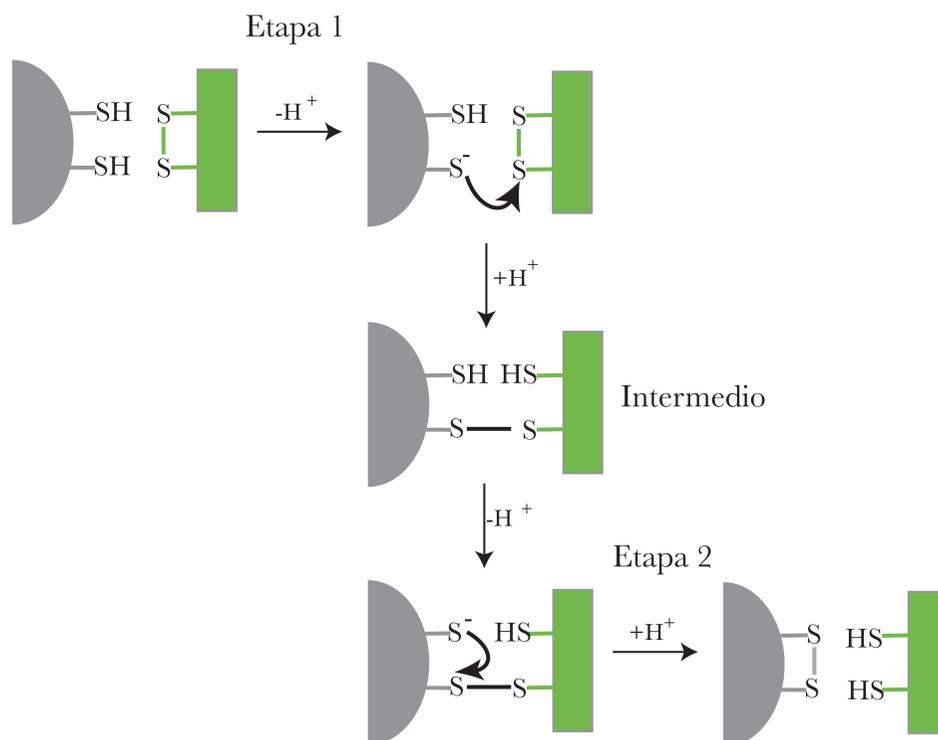
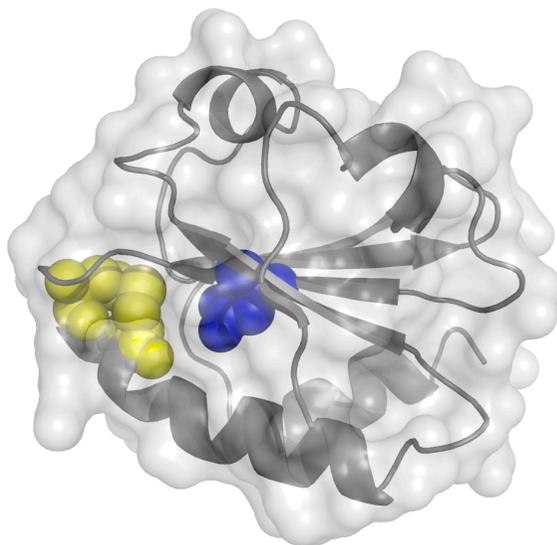


fig 3.2.20. Esquema del mecanismo de reducción de los puentes disulfuros catalizada por la tiorredoxina.

Se ha observado la existencia de un aspártico (Asp26) muy conservado tanto en las tiorredoxinas, como en el resto de familias de proteínas que comparten el motivo CXXC como sitio activo (fig 3.2.21). Además, se ha comprobado experimentalmente que dicho Asp26, que se encuentra en el núcleo hidrofóbico de la proteína, tiene un papel primordial en las reacciones redox que se producen en la actuación de la tiorredoxina como catalizador [31, 32]. Si bien el papel del Asp26 en la actividad reductasa ha sido descrito en la tiorredoxina de *E. coli*, la conservación de este residuo a lo largo de la evolución nos sugiere un papel similar en tiorredoxinas ancestrales. Bajo este supuesto, para nuestro propósito tomamos como proteína de partida una variante de la tiorredoxina de LPBCA de 128°C de  $T_m$ , a la que nos referiremos como variante triple y cuyo diseño se ha descrito en el apartado 3.2.1.1.1. Se eligió esta proteína porque con frecuencia la estabilidad

es un factor limitante en objetivos de ingeniería de proteínas, sin embargo, una proteína tan estable permite acumular mutaciones desestabilizantes sin que se comprometa su viabilidad.



*fig 3.2.21. Molécula de tioredoxina de LPBCA (código de pdb: 2yj7) resuelta en colaboración con José A. Gavira perteneciente al Laboratorio de Estudios Cristalográficos de Granada (CSIC). En azul se ha marcado el aspártico en posición 26 y en amarillo el puente disulfuro entre las cisteínas 32 y 35 responsables de la actividad de la proteína.*

En este trabajo, nos planteamos explorar la posibilidad de simular el efecto del Asp26 en la actividad reductasa de la variante triple, con mutaciones en residuos que estuvieran expuestos al solvente. Para ello, nos propusimos primero, mutar el aspártico en la posición 26 y luego, sobre la proteína resultante, construir una biblioteca combinatorial con todas las posibles combinaciones de las mutaciones seleccionadas para que presumiblemente lleven a cabo el efecto en la actividad reductasa del Asp26.

Para eliminar el aspártico en posición 26 se realizó, sobre la variante triple, la mutación D26I ya que estaba descrita en la bibliografía por ser estabilizante [33].

Obtuvimos así, la proteína de partida (*background*) para la construcción de la biblioteca combinatorial.

Para elegir las mutaciones que conformaron la biblioteca combinatorial, se siguieron varios pasos:

- 1) Se seleccionaron aquellos residuos periféricos al centro activo.
- 2) De los seleccionados según el primer criterio, se eligieron los residuos más expuestos al solvente.
- 3) Se excluyeron los residuos directamente implicados en el sitio activo.

Para seleccionar los residuos periféricos al centro activo, se tomó como referencia el azufre de la Cys32. Se calcularon las distancias entre dicho átomo y el átomo de cada residuo más cercano a la posición 32. Se descartaron las posiciones cuya distancia medida resultó mayor a 11 Å.

Para seleccionar los residuos más expuestos, se calcularon las accesibilidades de la cadena lateral de los residuos de la tiorredoxina de LPBCA. El cálculo de área superficial accesible se realizó como se explica en el apartado 3.2.1.2. La accesibilidad de los residuos se calculó como el cociente entre el ASA de la cadena lateral en la estructura nativa y la del tripéptido Gly-X-Gly. Los cálculos se realizaron con el programa ASA-ENG.exe y SELECTGA.exe desarrollado en nuestro grupo de investigación. Se seleccionaron aquellos residuos con una accesibilidad mayor a 0.3.

Para excluir los residuos directamente implicados en el sitio activo se eliminaron aquellos ubicados a menos de 7 Å de la Cys32 y también se descartaron los residuos descritos en bibliografía como fundamentales para la actividad de la tiorredoxina como son Trp31, Asp61, Pro76 y Lys57 [34].

En la tabla 3.2.2 se muestran las posiciones elegidas finalmente a mutar, según los criterios anteriormente descritos.

<b>Residuo</b>	<b>Posición</b>	$d_{Cys32}(\text{Å})$	<b>Accesibilidad</b>
<b>Prolina</b>	30	7.0	0.58
<b>Metionina</b>	37	9.4	0.73
<b>Alanina</b>	39	9.5	0.33
<b>Arginina</b>	73	10.1	0.93
<b>Serina</b>	74	7.5	0.58
<b>Valina</b>	91	10.2	0.56
<b>Valina</b>	93	8.4	0.33

tabla 3.2.2. Residuos seleccionados por el análisis anteriormente expuesto para simular el efecto en la actividad reductasa del aspártico enterrado en la posición 26. En la primera columna se muestran los residuos seleccionados, en la segunda, su posición, en la tercera se muestra la distancia de cada residuo al azufre de la Cys32 y en la cuarta columna su accesibilidad al solvente.

Como para el desarrollo de la catálisis de la tiorredoxina se produce el ataque nucleofílico del sustrato, cabe pensar que los residuos ionizables, al ser capaces de modificar la nucleofilidad, pudiesen tener un papel primordial en el desarrollo de la actividad reductasa. Así, para la construcción de la biblioteca según el método de “Gene Assembly Mutagenesis” (descrito en la sección 2.3 de metodología), se consideraron para cada posición tres posibilidades: no mutar, mutar a lisina y mutar a glutamato, excepto para la posición 73 cuyas tres posibilidades fueron: no mutar, mutar a glutamato y mutar a metionina. Se eligió la mutación a metionina por ser una mutación frecuente en el alineamiento de secuencias de *E. coli* con la base de datos Swiss-Prot de Mayo del 2009. La biblioteca, por tanto, contiene 2187 variantes que incluyen todas las posibles combinaciones de las 3 mutaciones realizadas en las 7 posiciones seleccionadas. En fig 3.2.22, se muestran las posiciones consideradas para construir la biblioteca combinatorial, en la estructura tridimensional de la tiorredoxina de LPBCA.

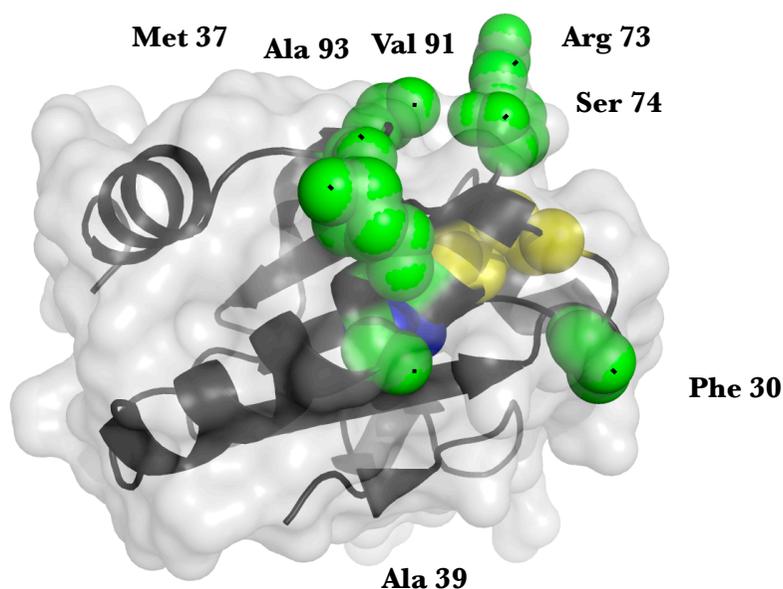


fig 3.2.22. Molécula de tiorredoxina de LPBCA (2yj7). Las siete posiciones expuestas al disolvente seleccionadas para mutar se muestran en verde, en azul se muestra el Asp26 que es el aminoácido ionizable enterrado cuyo efecto (en la actividad reductasa) se quiere simular y el puente disulfuro formado entre la Cys32 y la Cys35 se muestra en amarillo.

### 3.2.3.2. Estudio experimental de la actividad reductasa y estabilidad de las variantes de la biblioteca combinatorial

Para el estudio experimental de la biblioteca, se seleccionaron al azar 39 variantes y se expresaron y purificaron tal y como se describe en las secciones 2.4.1, 2.4.3, y 2.4.4.1 de metodología. Todas ellas se estudiaron conforme su actividad reductasa y su estabilidad.

La actividad reductasa se midió a 37°C con el ensayo turbidimétrico en el que la tiorredoxina cataliza la reducción de la insulina [35] (descrito en la sección 2.10. de metodología). En fig 3.2.23 se muestran los resultados del ensayo de actividad del *background* de la biblioteca (variante triple con isoleucina en posición 26), de la variante triple con aspártico en posición 26, de la tiorredoxina de *E. coli* y de algunas de las variantes de la biblioteca que han sido estudiadas. Como se puede observar, la variante triple tiene aproximadamente un 150% de la actividad de la

tioredoxina de *E. coli*, disminuyendo esta considerablemente al eliminar el Asp26. Parte de la actividad reductasa debida al efecto del Asp26 se recupera, en algunas de las variantes estudiadas de la biblioteca.

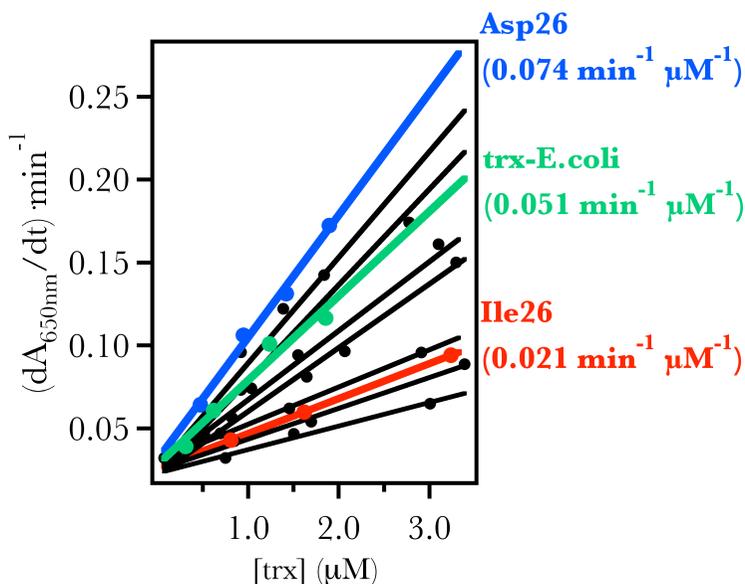


fig 3.2.23. Velocidad máxima de agregación de insulina en función de la concentración de variantes de tiorredoxina. En rojo se muestra el background de la biblioteca (variante triple con isoleucina en posición 26), en azul la variante triple con aspártico en posición 26 y en negro se muestran algunas de las variantes de la biblioteca estudiadas experimentalmente. Los valores de actividad que se muestran para el background, para la variante triple con el Asp26 y para la tiorredoxina de *E. coli*, se corresponden con la pendiente del ajuste lineal de los datos y será el dato que utilizaremos como indicativo de la actividad reductasa.

En fig 3.2.24 se muestra una comparativa de las actividades de las variantes de la biblioteca estudiadas experimentalmente con respecto a la actividad del *background*. Muchas de las variantes presentaron mayor actividad catalítica que la proteína de partida de la biblioteca, llegando en algunos casos a triplicar su actividad y, por lo tanto, recuperándose, al menos en parte, el efecto del Asp26 en la actividad reductasa.

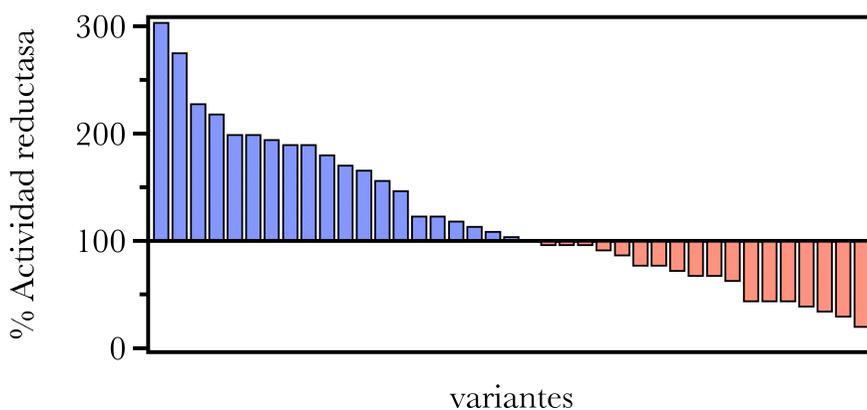


fig 3.2.24. Actividad reductasa de las variantes de la biblioteca estudiadas experimentalmente expresada en tanto por ciento con respecto al background, considerando la actividad del background de la biblioteca como el 100%. Las variantes más activas que el background se muestran en azul y las menos activas en rojo.

Para el estudio de la estabilidad, se realizaron experimentos de DSC de 20 a 130°C a 240K/h. Las muestras se prepararon en 50 mM HEPES a pH 7, a una concentración de proteína de 0.5 mg/mL. La estabilidad se midió tomando como valor de  $T_m$  la temperatura a la cual la capacidad calorífica es máxima. En algunos casos, la estabilidad de la proteína no se pudo determinar de forma directa ya que el pico de la transición calorimétrica no se observaba en el rango de temperaturas en el que se realizó el experimento. En esos casos, se estimó su estabilidad ajustando la traza al modelo de dos estados reversible fijando la entalpía a un valor razonable según el estudio descrito en la sección 3.2.1.1 (660 kJ mol<sup>-1</sup>). Los valores de  $T_m$  obtenidos de esta manera son una aproximación y deben tomarse con cautela. En fig 3.2.25 se muestran las trazas calorimétricas de la variante triple con aspártico en la posición 26, del *background* de la biblioteca (variante triple con isoleucina en la posición 26) y de la tiorredoxina de *E. coli*. Los valores de capacidad calorífica de la tiorredoxina de *E. coli* y de la variante triple con Asp26 son datos experimentales, mientras que para el *background* se muestra la

simulación de los datos experimentales al modelo de equilibrio de dos estados fijando la entalpía en  $660 \text{ kJ} \cdot \text{mol}^{-1}$ .

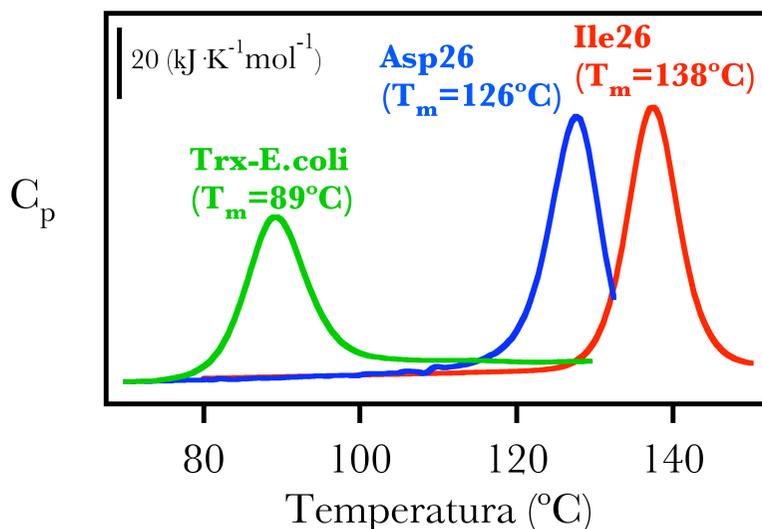


fig 3.2.25. Perfil calorimétrico experimental de la tiorredoxina de *E. coli* y de la variante triple con aspártico en la posición 26. Para el background de la biblioteca (variante triple con isoleucina en la posición 26) se muestra una extrapolación de los datos experimentales usando el modelo de equilibrio de dos estados reversible. Los experimentos se realizaron desde 20 a  $130^{\circ}\text{C}$ , a una velocidad de barrido de  $240\text{K/h}$  con una concentración de proteína de  $0.5 \text{ mg/mL}$  en  $50 \text{ mM HEPES}$ , a  $\text{pH } 7$ . Los valores de capacidad calorífica de la variante triple con Asp26 y de la tiorredoxina de *E. coli* son datos experimentales. Para el background (Ile26), en el rango de temperaturas en el que se registró el experimento no se observaba el pico de la transición, por lo que se ajustaron los datos experimentales al modelo de equilibrio de dos estados fijando la entalpía a un valor razonable según el estudio descrito en la sección 3.2.1.1. ( $660 \text{ kJ} \cdot \text{mol}^{-1}$ ). La mutación D26I, está descrita en la bibliografía como responsable del aumento de  $11^{\circ}\text{C}$  en la  $T_m$  de la tiorredoxina de *E. coli* [33] y de hecho, el mismo efecto parece tener sobre la variante triple de la tiorredoxina de LPBCA.

En fig 3.2.26 se muestran las diferencias de estabildades entre cada una de las variantes estudiadas y el *background*, medida como el incremento de  $T_m$ . Como se puede observar, la mayoría de las variantes son menos estables que el *background*, sin embargo, al partir de una proteína tan estable (con una  $T_m=138^{\circ}\text{C}$ ), todas las variantes estudiadas, a pesar de albergar mutaciones muy destabilizantes, aun

presentaban una estabilidad considerable (con valores de  $T_m$  superiores a los  $110^{\circ}\text{C}$  en todos los casos).

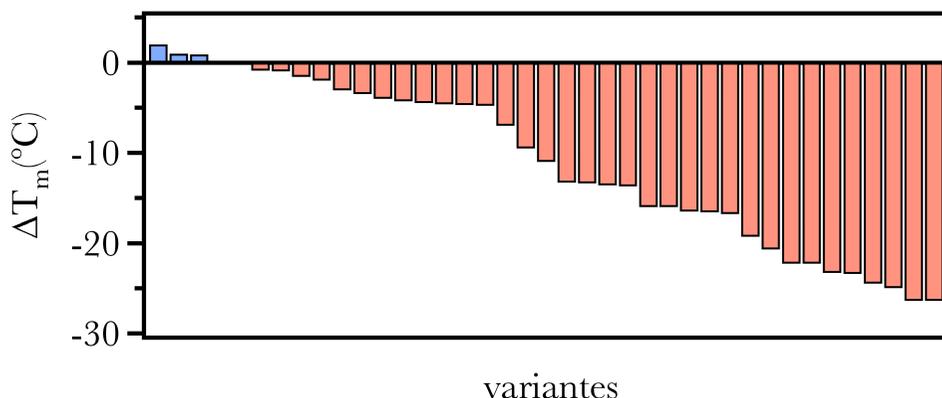


fig 3.2.26. Incremento de estabilidad, medida como la diferencia de  $T_m$  entre cada variante y el *background* de la biblioteca. En azul se representan las variables que resultaron más estables y en rojo aquellas que resultaron menos estables. La mayoría de las variantes contienen mutaciones desestabilizantes, aun así, todas las variantes estudiadas tenían una estabilidad considerable, con valores de  $T_m$  por encima de los  $100^{\circ}\text{C}$ .

A modo de resumen, en la tabla 3.2.3, se muestran las secuencias de las variantes estudiadas, su estabilidad y su actividad reductasa. La secuencia se indica en código de color, de manera que si el residuo es neutro, el recuadro de la tabla se representa en blanco, si está cargado positivamente, en azul y cargado negativamente en rojo. La estabilidad se mide como la  $T_m$  y el dato de actividad, se corresponde con la pendiente que resulta al ajustar linealmente la velocidad máxima de agregación de la insulina frente a la concentración de proteína que cataliza la reducción de la insulina. En fig 3.2.27 se representa, para la variante triple con el Asp26, el *background* y las 39 variantes de la biblioteca estudiadas experimentalmente, la relación entre la actividad reductasa y la estabilidad. Tras el estudio experimental de una pequeña parte de la biblioteca (menos del 2%), se han encontrado algunas variantes en las que el efecto del Asp26 en la actividad reductasa es llevado a cabo por otros residuos expuestos al solvente.

Phe30	Met37	Ala39	Arg73	Ser74	Val91	Ala93	T <sub>m</sub> (°C)	Actividad (min <sup>-1</sup> μM <sup>-1</sup> )
							124.7	0.064
							128.5	0.058
							115.7	0.048
							115.7	0.046
							133.7	0.042
							122.0	0.042
							117.3	0.041
							121.5	0.040
							121.2	0.040
							134.0	0.038
							124.6	0.036
							113.5	0.035
							131.0	0.033
							111.6	0.031
							137.0	0.026
							118.7	0.026
							111.6	0.025
							134.9	0.024
							134.5	0.023
							121.4	0.022
							138.0	0.021
							114.6	0.020
							122.0	0.020
							124.4	0.020
							127.0	0.019
							137.8	0.018
							124.3	0.016
							113.0	0.016
							136.4	0.015
							133.2	0.014
							133.3	0.014
							133.4	0.013
							133.5	0.009
							140.0	0.009
							139.0	0.009
							138.9	0.008
							114.7	0.007
							137.1	0.006
							136.0	0.004

tabla 3.2.3. Secuencias de las variantes de la biblioteca seleccionadas al azar y analizadas en término de su actividad y su estabilidad. Los residuos de las siete posiciones seleccionadas mantienen el código de color: en blanco si el residuo es neutro, en azul si está cargado positivamente y en rojo si está cargado negativamente. En la columna 8 se muestra el valor de la T<sub>m</sub> (°C) y en la última columna se muestra el valor de la actividad reductasa (min<sup>-1</sup> · μM<sup>-1</sup>). El valor de actividad reductasa se obtiene de la pendiente que resulta al ajustar linealmente la velocidad máxima de agregación de la insulina frente a la concentración de proteína.

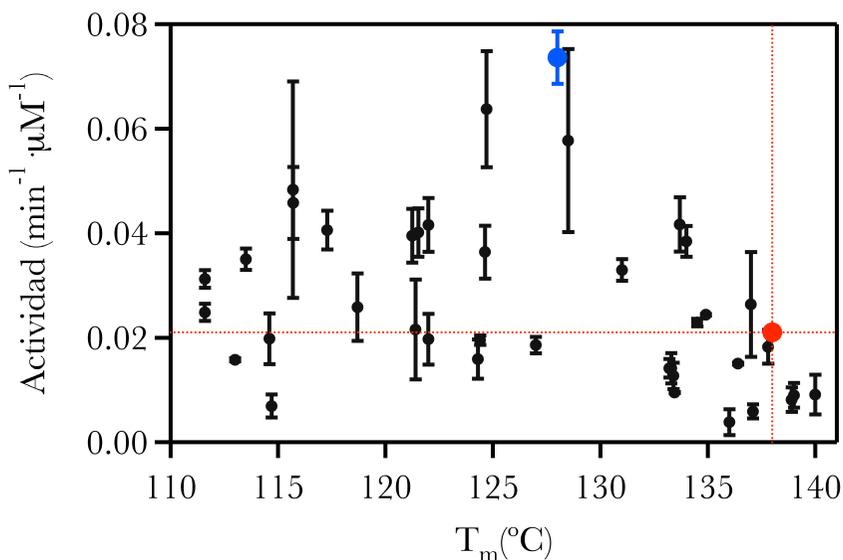


fig 3.2.27. Actividad reductasa de las 39 variantes de la bibliotecas elegidas al azar para su estudio frente a su temperatura de desnaturalización. En rojo se muestra el background de la biblioteca y en azul la variante triple con aspártico en la posición 26.

### 3.2.3.3. Reconstrucción del espacio mutacional completo de la biblioteca combinatorial mediante la regresión multivariante por mínimos cuadrados parciales

La regresión por mínimos cuadrados parciales, PLS-R, es un método de análisis multivariante utilizado para determinar el efecto que tienen un grupo de predictores (variables independientes o variables  $x$ ) sobre una o varias variables respuesta (variables dependientes o variables  $y$ ). Llamaremos  $Y$  a la matriz formada por el conjunto de  $M$  variables dependientes de  $N$  muestras, y  $X$  a la matriz de  $K$  variables independientes de las  $N$  muestras. El objeto del PLS-R es predecir la matriz  $Y$  a partir de la matriz  $X$ . Esta técnica de regresión es particularmente adecuada en el caso en el número de variables independientes sea muy grande en comparación con el número de variables dependientes, e incluso cuando superen en número a las muestras ( $K > N$ ). También es resaltable el hecho

de que sea capaz de trabajar con datos con considerable ruido, colinearidad e incluso con una elevada proporción de datos incompletos. En el modelo PLS-R las matrices  $X$  e  $Y$  pueden descomponerse en una serie de variables latentes (pocas y ortogonales entre sí) que se estiman como combinaciones lineales de las variables originales. Además, la descomposición de las matrices  $X$  e  $Y$  no es independiente, sino que es concurrente y está interrelacionada. El PLS-R, utiliza las variables latentes para ajustar la ecuación que nos permita predecir las variables dependientes a partir de las variables independientes.

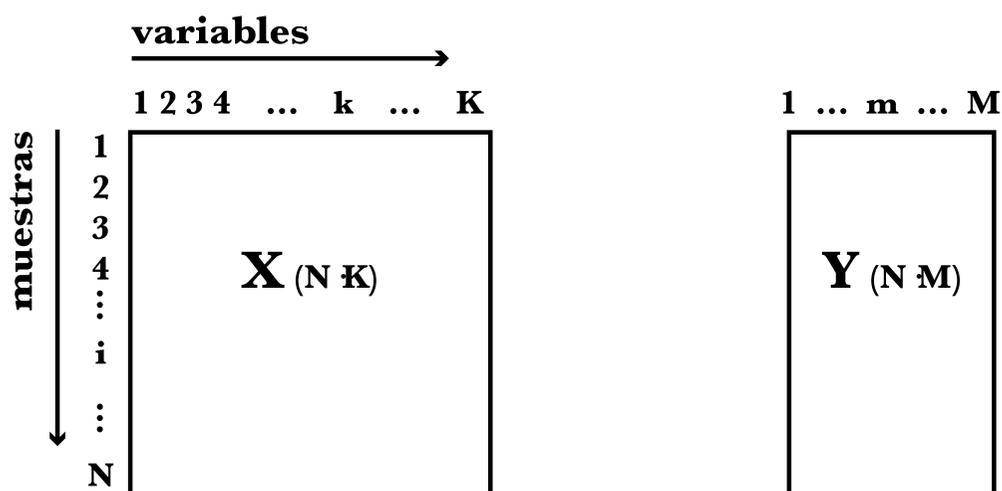


fig 3.2.28. El PLS-R trabajará con un conjunto de  $N$  observaciones, o muestras estudiadas con  $K$  variables independientes  $x$ , y  $M$  variables dependientes  $y$ . Formando así dos matrices de dimensiones  $(N \cdot K)$  y  $(N \cdot M)$ .

Como se ha descrito en el apartado anterior, el estudio experimental de una pequeña parte de la biblioteca (menos del 2%), ya ha mostrado variantes en las que el efecto del Asp26 en la actividad reductasa es llevado a cabo, al menos en parte, por otros residuos expuestos al solvente. Sin embargo, la mayoría de las variantes de la biblioteca no han sido estudiadas. Para el estudio de la biblioteca

completa usamos el PLS-R, un procedimiento novedoso con un gran potencial en el campo de la ingeniería de proteínas [36-39]. En este trabajo, se trató de explicar el comportamiento de una propiedad de la proteína (la actividad reductasa, que sería la variable dependiente) relacionándola con un gran número de variables independientes que contenían información de la secuencia de las proteínas. Para ello, el PLS-R lleva a cabo el ajuste de los datos experimentales basado en la ecuación ec.3.2.4, seguido de la reconstrucción de toda la biblioteca utilizando los valores de los parámetros de ajuste.

$$\ln A = \sum_i \delta_i p_i + \sum_i \sum_{i \neq j} \delta_{ij} p_{ij} \quad \text{ec.3.2.4}$$

El significado de los símbolos en la ecuación ec.3.2.4 es la siguiente:

- A es la variable dependiente, en nuestro caso la actividad reductasa.
- $\delta_i$  es una variable independiente que puede tomar tres valores según el tipo de residuo que se encuentre en la posición i: 100 en el caso de residuo neutro, 010 en el caso de residuo cargado positivamente o 001 en el caso de residuo cargado negativamente.
- $p_i$ , es una medida del efecto de la mutación en la posición i en la actividad A.
- $\delta_{ij}$ , es una variable independiente que toma diferentes valores según estén ocupadas las posiciones i y j. Los posibles valores de  $\delta_{ij}$  según el valor de  $\delta_i$  y de  $\delta_j$  se muestran en la tabla 3.2.4.
- $p_{ij}$  es una medida del efecto del acoplamiento entre las mutaciones en las posiciones i y j en la actividad.

<b>Posición i</b>	$\delta_i$	<b>Posición j</b>	$\delta_j$	$\delta_{ij}$
neutro	100	neutro	100	0010000000000000
neutro	100	positivo	010	0001000000000000
neutro	100	negativo	001	0000100000000000
positivo	010	neutro	100	0000001000000000
positivo	010	positivo	010	0000000100000000
positivo	010	negativo	001	0000000010000000
negativo	001	neutro	100	0000000001000000
negativo	001	positivo	010	0000000000100000
negativo	001	negativo	001	0000000000010000

tabla 3.2.4. Posible valores de  $\delta_{ij}$ , según el tipo de residuo que se encuentre en la posición i y j.

La ec.3.2.4 representa un modelo que incluye los efectos de las mutaciones en la actividad ( $p_i$ ) y además tiene en cuenta si los efectos de las mutaciones son aditivos ( $p_{ij} = 0$ ) o están correlacionados ( $p_{ij} \neq 0$ ). Se trata de ajustar 231 parámetros (21 valores de  $p_i$  y 210 valores de  $p_{ij}$ ) con sólo 39 valores experimentales para ajustar (el valor de la actividad reductasa de las variantes de la biblioteca estudiadas). El PLS-R opera simplificando el número de variables dependientes, de manera que a partir del set de variables originales construye el set de variables latentes. Las variables latentes se construyen como combinación lineal de las variables originales cumpliéndose:

- que sean ortogonales entre sí, es decir que cada una de las variables latentes explica una cierta proporción de variables originales diferente al que explica otra variable latente. La primera de las variables latentes es la que contiene la mayor información sobre las variables originales.

- que expliquen la mayor varianza del set de valores independientes originales, es decir que expliquen la dispersión en los valores de las variables independientes.
- que maximicen la covarianza de las variables dependientes, es decir que expliquen la correlación entre las variables dependientes e independientes.

Para obtener las variables latentes el PLS-R utiliza por tanto información del set de variables independientes originales, y también utiliza información del set de variables dependientes, de manera que dichas variables latentes, tengan mayor poder predictor que las que se obtienen con otros métodos de reducción de variables que solo utilizan información de la matriz  $X$ . La regresión por mínimos cuadrados parciales se calculó utilizando el algoritmo NIPALS (*Nonlinear Iterative Partial Least Squares*) [40] que es capaz de tolerar cierta cantidad de datos incompletos, lo hace sustituyendo iterativamente los datos vacíos por predicciones.

Utilizando las variables latentes, el PLS-R ajusta los datos experimentales de las 39 variables estudiadas a la ec.3.2.4. Con los datos del ajuste se calculó la actividad esperada para todas las variantes de la biblioteca (fig 3.2.29).

La evaluación de la incertidumbre asociada a la reconstrucción de la biblioteca se realizó utilizando un método de remuestreo (*bootstrapping*) antes del ajuste con el PLS-R. Partiendo del set inicial de variantes estudiadas (las 39 variantes estudiadas experimentalmente), con el programa BOOT.exe, desarrollado en nuestro grupo de investigación, se generaron 10 conjuntos de muestras de forma aleatoria con reemplazamiento. La reconstrucción de los valores de actividad para la biblioteca completa de las 10 replicas del set de datos iniciales se muestra en fig 3.2.29. En el primer panel se muestran los valores de actividad de la biblioteca completa para cada una de las 2187 secuencias y en el segundo panel se muestran los valores de actividad *versus* el número de mutaciones de la secuencia. La predicción reveló como con pocas mutaciones, se puede alcanzar un valor de

actividad comparable al de la variante triple con aspártico en la posición 26. Así que decidimos elegir aquellas variantes cuya actividad predicha fuera máxima, introduciendo el menor número de mutaciones: A39K, P31K, A39E, A39K/S74K, A39K/K73M y A39K/P30K (marcadas en rojo en fig 3.2.29).

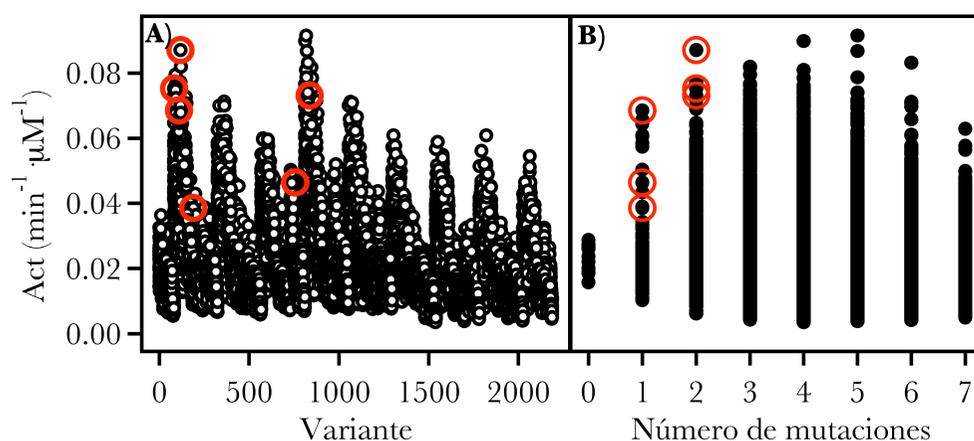


fig 3.2.29. Reconstrucción de los valores de actividad para la biblioteca completa a partir de las 10 réplicas del set de datos iniciales. A) Los datos de actividad se muestran para cada una de las 2187 variantes. B) Los valores de actividad se representan frente al número de mutaciones de la secuencia. Las variantes de la predicción seleccionadas para estudiarlas experimentalmente se muestran marcadas en rojo.

Se construyeron las variantes seleccionadas A39K, A39E, A39K/S74K, A39K/K73M y A39K/P30K mediante mutagénesis dirigida (descrito en la sección 2.1 de metodología), se expresaron y purificaron (descrito en sección 2.4.1, 2.4.3, y 2.4.4.1 de metodología) y se estudiaron conforme su actividad y su estabilidad. La variante P31K ya se había estudiado experimentalmente en el set inicial de las 39 variantes seleccionadas al azar. En la tabla 3.2.5, se muestra el valor predicho de actividad reductasa y el obtenido experimentalmente. Los valores de actividad obtenidos experimentalmente fueron menores que los predichos, sin embargo, aportaron información cualitativa útil para guiar el estudio de la biblioteca en la dirección de las variantes que, con pocas mutaciones, tienen una actividad reductasa comparable con la variante triple con Asp26.

Secuencia	Actividad predicha ( $\text{min}^{-1} \mu\text{M}^{-1}$ )	Actividad medida ( $\text{min}^{-1} \mu\text{M}^{-1}$ )
<b>A39K</b>	0.069	0.064
<b>P31K</b>	0.039	0.020
<b>A39E</b>	0.039	0.012
<b>A39K/S74K</b>	0.087	0.038
<b>A39K/K73M</b>	0.075	0.044
<b>A39K/P30K</b>	0.073	0.048

tabla 3.2.5. Variantes seleccionadas tras la reconstrucción completa de la biblioteca. En la segunda columna se muestra el valor predicho para la actividad con el PLS y en la tercera columna se muestra el valor medido experimentalmente.

En fig 3.2.30 se muestra, para las 44 variantes estudiadas experimentalmente (39 seleccionadas inicialmente al azar y 5 estudiadas tras el análisis por PLS), la relación entre la actividad reductasa y la estabilidad.

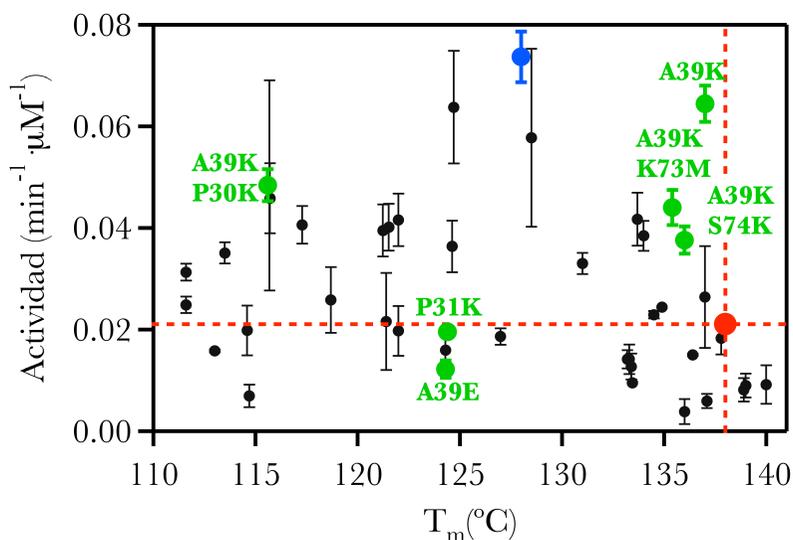


fig 3.2.30. Actividad del total de las variantes de la bibliotecas estudiadas experimentalmente frente a su temperatura de desnaturalización. En rojo se muestra el background de la biblioteca, en azul la variante triple con aspártico en la posición 26 y en verde las variantes propuestas por el PLS como aquellas cuya actividad es máxima introduciendo solamente una o dos mutaciones con respecto al background de la biblioteca.

Tras la realización de este trabajo, cabe resaltar varios aspectos. En primer lugar, en cuanto la metodología, la aplicación de la técnica de regresión multivariante, la llamada regresión por mínimos cuadrados parciales, nos ha permitido con el estudio experimental de unas pocas variantes, modelar la relación entre fenotipo y genotipo, y de esta forma inferir el comportamiento de toda la biblioteca. Así hemos podido acceder a las variantes con las características deseadas, que en nuestro caso fueron aquellas en las que el efecto del Asp26 en la actividad reductasa es llevado a cabo por pocos residuos (uno o dos) expuestos al solvente. En segundo lugar, partir de una variante ancestral para el diseño de la biblioteca, nos permitió explorar el efecto en la actividad reductasa de combinaciones de mutaciones muy desestabilizantes (en algunos casos con 30°C de disminución de la  $T_m$ ), esto fue posible por la hiper-estabilidad de la proteína de partida que al tener una  $T_m$  de 138°C, nos permite disponer de un amplio margen de estabilidad para elegir las mutaciones que presumiblemente llevan a cabo el efecto en la catálisis enzimática del grupo ionizable enterrado, sin que se comprometa su viabilidad.

Por otro lado, en cuanto a los objetivos específicos abordados, en este trabajo hemos demostrado la posibilidad de combinar residuos ionizables en la superficie expuesta al solvente de la tiorredoxina, para llevar a cabo el efecto en la actividad reductasa de un residuo ionizable enterrado, primordial para el desarrollo de la catálisis enzimática. En concreto, hemos demostrado como un solo residuo que se encuentra expuesto al solvente (Lys39), puede realizar el efecto en la actividad reductasa de un residuo ionizable enterrado (Asp26). Además, si bien la mayoría de las variantes estudiadas eran menos estable que el *background*, la variante con la mutación A39K, la más activa de las estudiadas, tiene un valor de  $T_m=137^\circ\text{C}$ . De manera que hemos obtenido una variante de tiorredoxina más activa que la de *E. coli* y con un valor de  $T_m$  48°C mayor.

### 3.2.4. Bibliografía

1. Pauling L, Z.E., *Chemical paleogenetics. Molecular "restoration studies" of extinct forms of life.* Acta Chemica Scandinava, 1963(17): p. S9-S16.
2. Jermann, T.M., et al., *Reconstructing the evolutionary history of the artiodactyl ribonuclease superfamily.* Nature, 1995. **374**(6517): p. 57-9.
3. Thornton, J.W., E. Need, and D. Crews, *Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling.* Science, 2003. **301**(5640): p. 1714-7.
4. Thomson, J.M., et al., *Resurrecting ancestral alcohol dehydrogenases from yeast.* Nature genetics, 2005. **37**(6): p. 630-5.
5. Chang, B.S., et al., *Recreating a functional ancestral archosaur visual pigment.* Molecular biology and evolution, 2002. **19**(9): p. 1483-9.
6. Perez-Jimenez, R., et al., *Single-molecule paleoenzymology probes the chemistry of resurrected enzymes.* Nature structural & molecular biology, 2011. **18**(5): p. 592-6.
7. Gaucher, E.A., et al., *Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins.* Nature, 2003. **425**(6955): p. 285-8.
8. Gaucher, E.A., S. Govindarajan, and O.K. Ganesh, *Palaeotemperature trend for Precambrian life inferred from resurrected proteins.* Nature, 2008. **451**(7179): p. 704-7.
9. Boussau, B., et al., *Parallel adaptations to high temperatures in the Archaean eon.* Nature, 2008. **456**(7224): p. 942-5.
10. Miyazaki, J., et al., *Ancestral residues stabilizing 3-isopropylmalate dehydrogenase of an extreme thermophile: experimental evidence supporting the thermophilic common ancestor hypothesis.* Journal of biochemistry, 2001. **129**(5): p. 777-82.

11. Shimizu, H., et al., *Extremely thermophilic translation system in the common ancestor commonote: ancestral mutants of Glycyl-tRNA synthetase from the extreme thermophile Thermus thermophilus*. Journal of molecular biology, 2007. **369**(4): p. 1060-9.
12. Perez-Jimenez, R., et al., *Diversity of chemical mechanisms in thioredoxin catalysis revealed by single-molecule force spectroscopy*. Nature structural & molecular biology, 2009. **16**(8): p. 890-6.
13. Bershtein, S., et al., *Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein*. Nature, 2006. **444**(7121): p. 929-32.
14. Zeldovich, K.B., P. Chen, and E.I. Shakhnovich, *Protein stability imposes limits on organism complexity and speed of molecular evolution*. Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(41): p. 16152-7.
15. Tokuriki, N. and D.S. Tawfik, *Chaperonin overexpression promotes genetic variation and enzyme evolution*. Nature, 2009. **459**(7247): p. 668-73.
16. Tokuriki, N. and D.S. Tawfik, *Stability effects of mutations and protein evolvability*. Current opinion in structural biology, 2009. **19**(5): p. 596-604.
17. Godoy-Ruiz, R., et al., *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*. Journal of molecular biology, 2006. **362**(5): p. 966-78.
18. Plaza del Pino, I.M., B. Ibarra-Molero, and J.M. Sanchez-Ruiz, *Lower kinetic limit to protein thermal stability: a proposal regarding protein stability in vivo and its relation with misfolding diseases*. Proteins, 2000. **40**(1): p. 58-70.
19. Privalov, P.L., *Stability of proteins: small globular proteins*. Advances in protein chemistry, 1979. **33**: p. 167-241.

20. Georgescu, R.E., et al., *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.* European journal of biochemistry / FEBS, 2001. **268**(5): p. 1477-85.
21. Pey, A.L., et al., *Modulation of buried ionizable groups in proteins with engineered surface charge.* Journal of the American Chemical Society, 2010. **132**(4): p. 1218-9.
22. Matthews, B.W., H. Nicholson, and W.J. Becktel, *Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding.* Proceedings of the National Academy of Sciences of the United States of America, 1987. **84**(19): p. 6663-7.
23. Nemethy, G., S.J. Leach, and H.A. Scheraga, *Influence of Amino Acid Side Chains on Free Energy of Helix-Coil Transitions.* Journal of Physical Chemistry, 1966. **70**(4): p. 998-&.
24. Shrake, A. and J.A. Rupley, *Environment and exposure to solvent of protein atoms. Lysozyme and insulin.* Journal of molecular biology, 1973. **79**(2): p. 351-71.
25. Chothia, C., *Structural invariants in protein folding.* Nature, 1975. **254**(5498): p. 304-8.
26. Ibarra-Molero, B., et al., *Thermal versus guanidine-induced unfolding of ubiquitin. An analysis in terms of the contributions from charge-charge interactions to protein stability.* Biochemistry, 1999. **38**(25): p. 8138-49.
27. Costas, M., et al., *Between-species variation in the kinetic stability of TIM proteins linked to solvation-barrier free energies.* Journal of molecular biology, 2009. **385**(3): p. 924-37.
28. Griko, Y.V. and P.L. Privalov, *Thermodynamic puzzle of apomyoglobin unfolding.* Journal of molecular biology, 1994. **235**(4): p. 1318-25.

29. Sadqi, M., et al., *A designed protein as experimental model of primordial folding*. Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(11): p. 4127-32.
30. Garcia-Seisdedos, H., B. Ibarra-Molero, and J.M. Sanchez-Ruiz, *How many ionizable groups can sit on a protein hydrophobic core?* Proteins, 2012. **80**(1): p. 1-7.
31. Dyson, H.J., et al., *Effects of buried charged groups on cysteine thiol ionization and reactivity in Escherichia coli thioredoxin: structural and functional characterization of mutants of Asp 26 and Lys 57*. Biochemistry, 1997. **36**(9): p. 2622-36.
32. LeMaster, D.M., P.A. Springer, and C.J. Unkefer, *The role of the buried aspartate of Escherichia coli thioredoxin in the activation of the mixed disulfide intermediate*. The Journal of biological chemistry, 1997. **272**(48): p. 29998-30001.
33. Bolon, D.N. and S.L. Mayo, *Polar residues in the protein core of Escherichia coli thioredoxin are important for fold specificity*. Biochemistry, 2001. **40**(34): p. 10047-53.
34. Eklund, H., F.K. Gleason, and A. Holmgren, *Structural and functional relations among thioredoxins of different species*. Proteins, 1991. **11**(1): p. 13-28.
35. Holmgren, A., *Thioredoxin catalyzes the reduction of insulin disulfides by dithiothreitol and dihydrolipoamide*. The Journal of biological chemistry, 1979. **254**(19): p. 9627-32.
36. Fox, R., et al., *Optimizing the search algorithm for protein engineering by directed evolution*. Protein engineering, 2003. **16**(8): p. 589-97.
37. Fox, R., *Directed molecular evolution by machine learning and the influence of nonlinear interactions*. Journal of theoretical biology, 2005. **234**(2): p. 187-99.

38. Fox, R.J., et al., *Improving catalytic function by ProSAR-driven enzyme evolution*. Nature biotechnology, 2007. **25**(3): p. 338-44.
39. Garcia-Seisdedos, H., B. Ibarra-Molero, and J.M. Sanchez-Ruiz, *Probing the mutational interplay between primary and promiscuous protein functions: a computational-experimental approach*. PLoS computational biology, 2012. **8**(6): p. e1002558.
40. Wold, H., *Estimation of Principal Component and Related Models by Iterative Least Squares*. *Multivariate Analysis*, ed. I.P.R. Krishnaiah1966, New York: Academic Press.



## **4. Resumen y conclusiones**



En la presente Tesis Doctoral hemos abordado diferentes trabajos de relevancia en ingeniería de proteínas.

En un primer bloque tomamos como objeto de estudio la Proteína Morfogenética de Hueso 2 (BMP-2 de sus siglas en inglés) que, por su capacidad para inducir regeneración ósea, es una proteína de gran interés en el campo de la biomedicina. Por ello, hemos explorado la posibilidad de utilizar la información evolutiva contenida en los alineamientos de secuencias, para la obtención de variantes de BMP-2 con su función biológica mejorada. Para nuestro propósito, hemos planteado un modelo simple basado en la estadística de la selección natural purificadora obteniendo diferentes variantes de BMP-2. Para testar su actividad hemos llevado a cabo ensayos celulares basados en la detección de la actividad fosfatasa alcalina asociado al fenotipo osteoblástico inducido. Aunque los resultados obtenidos son prometedores, hemos encontrado dificultades en su interpretación asociadas a un cierto nivel de irreproducibilidad inherente al procedimiento de purificación usado. Por ello, este trabajo se encuentra actualmente en desarrollo, ya que se está trabajando en un nuevo sistema de producción de proteína, necesario para el estudio riguroso de las variantes obtenidas.

En un segundo bloque, hemos estudiado tiorredoxinas ancestrales, con el objetivo general de explorar el potencial de proteínas precámbricas resucitadas en ingeniería de proteínas. En primer lugar, hemos abordado los factores que determinan la muy alta estabilidad de estas proteínas ancestrales, tomando como objeto de estudio específico la tiorredoxina perteneciente al último ancestro común de cianobacterias y los grupos de deinococcus y thermus (trx-LPBCA) por ser la más estable de las resucitadas. Así, para analizar los factores que determinan su inusual estabilidad, hemos realizado un estudio mutacional detallado con objeto de caracterizar su estabilidad termodinámica y cinética. Este estudio nos guió con éxito en la búsqueda de mutaciones estabilizantes y además nos permitió concluir

que la naturaleza de la estabilización de la *trx-LPBCA* con respecto a la tiorredoxina moderna de *E. coli* es fundamentalmente entrópica, reflejando posiblemente la menor entropía conformacional en el estado desplegado de la proteína ancestral asociada a la eliminación de residuos de glicina y a la introducción de residuos de prolinas. Como resultados de estos estudios, sugerimos que la correlación estabilidad/evolución no refleja, al menos para la tiorredoxina, la selección natural de la estabilidad termodinámica por sí misma, sino de otro factor que está asociado a ésta. Proponemos, apoyándonos en el estudio cinético comparativo realizado entre la *trx-LPBCA* y su análoga actual, la tiorredoxina de *E. coli*, que dicho factor es la estabilidad cinética.

En el transcurso de este trabajo, obtuvimos resultados experimentales que parecían indicar una desnaturalización fría no cooperativa en la *trx-LPBCA* y que sugerían la posibilidad de que las tiorredoxinas ancestrales resucitadas nos pudieran permitir abordar el problema de la evolución de la cooperatividad. Esta posibilidad sin embargo no se confirmó, ya que hemos encontrado el mismo patrón en la desnaturalización fría en la tiorredoxina de *E. coli* y, además, lo hemos podido explicar en base a la lentitud de la cinética de desplegamiento a baja temperatura. Proponemos por tanto, apoyándonos en los resultados obtenidos, que la estabilización cinética a baja temperatura podría jugar un papel primordial en la adaptación al frío de organismos psicrófilos.

Por último, hemos realizado lo que posiblemente sea el primer trabajo de ingeniería de proteínas que toma como punto de partida una enzima ancestral resucitada. En concreto, hemos investigado la posibilidad de rediseñar el centro activo de la *trx-LPBCA*, mediante el estudio de una biblioteca combinatorial de variantes proteicas, cuyas mutaciones fueron propuestas por criterios estructurales. En cuanto a la metodología utilizada, cabe resaltar la aplicación de una técnica de regresión multivariante, la llamada regresión por mínimos cuadrados parciales, que nos permite, con el estudio de unas pocas variantes modelar la relación entre

genotipo y fenotipo, y de esta manera inferir el comportamiento de toda un biblioteca y por ende, acceder a las mejores variantes. En cuanto al objetivo específico abordado, hemos demostrado como el efecto de un residuo enterrado, esencial para la catálisis enzimática, puede sustituirse por el efecto de un residuo expuesto al solvente. Creemos que este resultado puede tener implicaciones para la ingeniería de sitios activos *de novo*, ya que mientras que el diseño de grupos enterrados implicados en catálisis es extremadamente difícil, el cribado de una biblioteca basada en mutaciones expuestas no debe ofrecer grandes dificultades, en particular si la biblioteca se ha construido sobre una variante hiper-estable, como es el caso de proteínas precámbricas resucitadas.





